An Automated Approach to Assess Relative Galectin-Glycan Affinity Following Glycan Microarray Analysis

Alex D. Ho[1†], Shang-Chuen Wu[1†], Nourine A. Kamili[1], Anna V. Blenda[2],
Richard D. Cummings[3], Sean R. Stowell *[1] and Connie M. Arthur *[1]

[1]Joint Program in Transfusion Medicine, Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, [2]Department of Biomedical Sciences, University of South Carolina School of Medicine Greenville, Greenville, SC, United States, [3]Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

Numerous studies have highlighted the utility of glycan microarray analysis for the elucidation of protein-glycan interactions. However, most current glycan microarray studies analyze glycan binding protein (GBP)-glycan interactions at a single protein concentration. While this approach provides useful information related to a GBP's overall binding capabilities, extrapolation of true glycan binding preferences using this method fails to account for printing variations or other factors that may confound relative binding. To overcome this limitation, we examined glycan array binding of three galectins over a range of concentrations to allow for a more complete assessment of binding preferences. This approach produced a richer data set than single concentration analysis and provided more accurate identification of true glycan binding preferences. However, while this approach can be highly informative, currently available data analysis approaches make it impractical to perform binding isotherms for each glycan present on currently available platforms following GBP evaluation. To overcome this limitation, we developed a method to directly optimize the efficiency of assessing association constants following multi-GBP concentration glycan array analysis. To this end, we developed programs that automatically analyze raw array data (kdMining) to generate output graphics (kaPlotting) following array analysis at multiple doses. These automatic programing methods reduced processing time from 32.8 h to 1.67 min. Taken together, these results demonstrate an effective approach to glycan array analysis that provides improved detail and efficiency when compared to previous methods.

Keywords: galectin, microarray, dissociation constant, R studio, python

INTRODUCTION

Glycan binding proteins (GBPs) play crucial roles in various biological activities (Elola et al., 2007; Yang et al., 2008; Arthur et al., 2015a; Bochner and Zimmermann, 2015; Schnaar, 2015). These diverse functions are commonly dictated by specific interactions with glycan ligands (Rabinovich and Toscano, 2009). As a result, clear understanding of glycan binding specificity is key to elucidation of

---

Abbreviations: Gal-3, galectin-3; Gal-3C, galectin-3C; Gal-7, galectin-7; Gal-9, galectin-9; Gal-9N, galectin-9N; Gal-9C, galectin-9C.

biological function ((Raman et al., 2005; Sato et al., 2009; Vasta, 2012). Given the often complex nature of glycan ligand synthesis, early studies often utilized relatively simple sugar substrates to probe GBP-glycan interactions (Cummings et al., 2009). However, as numerous studies suggest that subtle shifts of cell surface glycan modification are key to regulation of cell sensitivity to GBP activity, development of glycan microarrays representing extensive libraries of complex carbohydrate structures have provided key insight into GBP function (Blixt et al., 2004; Stowell et al., 2008a; McQuillan et al., 2019). In this way, glycan array analysis has enabled immense progress within the field of glycobiology by allowing assessment of GBP binding towards a number of complex glycan substrates simultaneously (Blixt et al., 2004; Stowell et al., 2008a; Song et al., 2009; Arthur et al., 2014; Arthur et al., 2015b; Gao et al., 2019). However, while glycan array analysis provides a wealth of information regarding GBP interactions with a wide variety of glycan substrates, glycan binding is often probed at only a single high concentration of GBP. This reflects challenging nature of microarray development and analysis, which remains time and resource demanding, often limiting analyses to a single concentration of a given GBP on an array. While this represents a reasonable method for the screening of glycan binding preferences for a given GBP, potential variations in the density of glycans printed on a glycan microarray may cause subtle inaccuracies in binding data. Often, this inaccuracy comes from the varied glycan density of different glycans on a printed array slide. While relatively minor, these differences, which may suggest varied printing efficiency across glycans, can result in less accurate glycan binding affinity comparisons. For this reason, we sought to establish whether binding analysis across multiple concentrations of GBP may enable relative binding affinity analyses that could serve to overcome inherent variability in the glycan printing process.

Assessment of GBP glycan array binding over a series of concentrations allows for estimation of maximal binding values (B_{max}) and a relative dissociation constant (K_D) (Perkins et al., 2012; Harding, 2021; Acipreste Hudson et al., 2022). Among GBPs analyzed for carbohydrate binding specificity, galectins represent some of the most studied. Galectins can be subdivided into three subgroups based on the quaternary structure of the protein (Verkerke et al., 2022). Each of the galectins analyzed in this study represent an example of each subgroup described including chimera type (hGal-3), prototype (hGal-7) and tandem repeat type (hGal-9) (Arthur et al., 2015a). Glycan binding was compared using a rank order approach at a single concentration compared to the relative affinity observed over a range of concentrations. We also included the C-terminal carbohydrate recognition domain (CRD) of hGal-3 (hGal-3C) and both the N and C-terminal CRDs of hGal-9 (hGal-9N and hGal-9C, respectively) for direct comparison to their full-length counterparts using this approach. To facilitate this work and improve the efficiency and productivity of GBP binding interrogation by glycan microarray analyses, while also reducing the possibility for human errors, we developed an automatic program, kdMining, to streamline the glycan-protein calculation and its saturation

determination. This program was used to calculate B_{max}, K_D, K_A, and % saturation (when saturation toward a given glycan did not occur) for each GBP analyzed accurately and efficiently. Furthermore, we developed a complementary program, kaPlotting, which processes some calculated outputs of kdMining, including relative K_A (for saturated binding) and % max (for unsaturated binding) into nested pie charts to enable more efficient glycan-protein interaction analyses and multi-GBP comparisons. Together these tools provide a useful method for analyzing GBP binding preference.

MATERIALS AND METHODS

Sample Collection and Processing

Human galectin was expressed and stored as outlined previously (Wu et al., 2021a; Paul et al., 2022). Six previously published human galectin datasets, human Gal-3, Gal-3C, Gal-7, Gal-9, Gal-9N, and Gal-9C (Wu et al., 2021b PMID:35754739; Blenda et al., 2022), were used to evaluate the performance of kdMining and kaPlotting. Each purified galectin was directly labeled with Alexa Fluor™ 488 NHS Ester (Invitrogen) per manufacturer instructions. Labeled protein was tested for glycan binding on a Consortium for Functional Glycomics (CFG) glycan microarray slide as described previously at the indicated concentrations and scanned using GenePix 4000B microarray scanner (Stowell et al., 2007; Stowell et al., 2008; Stowell et al., 2009; Stowell et al., 2009). hGal-3 and hGal-3C were analyzed on CFG V3.0. Glycans common to both array versions were renumbered into a merged glycan list for analysis, while glycans present on only one of the two array versions were not considered in the current analysis. Imagene software (GenePix Pro 7) was used to generate integrated spot intensities that were then converted to a spreadsheet file (GPR, GPRS) containing the raw RFU data representing binding at each printed glycan location as outlined previously (Stowell et al., 2007; Stowell et al., 2008b; Stowell et al., 2009a; Stowell et al., 2009b; Wu et al., 2021c; Arthur et al., 2022; Stowell et al., 2022). RFU values for selected GBP-glycan pairs were then plotted manually using Graphpad Prism V9.0. Alternatively, GPR spreadsheets served as the input file for the kdMining program.

Programming Method

kdMining was written using Python 3, a general-purpose and high-level programming language (Python Software Foundation, Wilmington, DE). Each input file had over 500 glycan samples along with their relative fluorescence units (RFU). The program first scanned all input spreadsheet files (GPR; GenePix Pro, San Jose, CA) in the designated directory once to ensure all input glycan sample files were present and in the correct format. The program can process and execute an infinite number of files (with multiple concentrations) given that the information does not exceed the capacity of the computer memory. All input files are stored within a dictionary, a data structure consists of key-value pairs in Python, while each key represents a single protein concentration that contains RFU values in the form of data frame. The program iterated through each protein
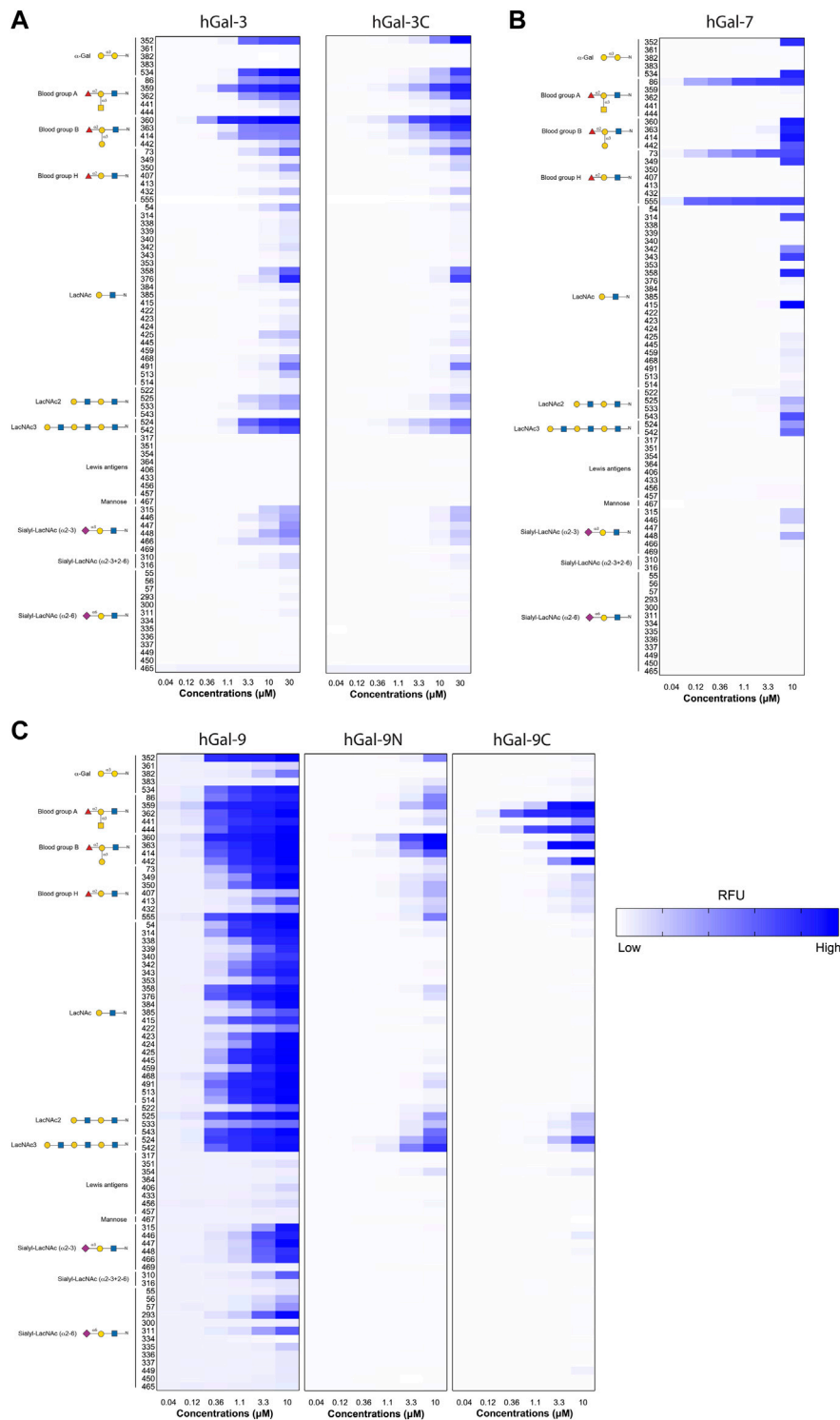
FIGURE 1 | Examination of GBP-glycan binding affinity over multiple concentrations. **(A)** Heatmaps for the RFU values of Gal-3 and Gal-3C over a broad range of concentrations starting from 0.04 to 30 μM for antigens on N-Glycans. **(B)** Heatmap for the RFU values of Gal-7 over a range of concentrations starting from 0.04 to 10 μM for antigens on N-Glycans **(C)** Heatmaps for the RFU values of Gal-9, Gal-9N, and Gal-9C over a range of concentrations starting from 0.04 to 30 μM for antigens on N-Glycans. Detailed Symbol Nomenclature for Glycans (SNFG) structures and examples of glycans examined are shown on the left. The higher RFU value is shown with darker blue and lower RFU value shown in lighter blue.

concentration to calculate the average RFU and standard deviation (STDEV) from the 6 replicate spots for each sample. The program then combined all input files from the dictionary into a single data frame to determine percentage saturation (% max), average RFU values, dissociation constants (K_D) and association constants (K_A) for each glycan sample (Islam et al., 2012). The K_D was calculated by fitting a sample's RFU values across all concentrations, while the K_A was calculated by taking the inverse of K_D. For interactions that did not result in saturation, a % max value was provided by taking the absolute difference of a particular sample's highest concentration RFU and the sample's highest concentration RFU overall and calculating the percent difference. At the end of program execution, a spreadsheet file (Excel; Microsoft, Redmond, WA) of the samples' calculated B_{max}, K_D, K_A, and % max and their corresponding graphs were available.

kaPlotting was written using R, a programming language for statistical computing and graphics (R Foundation for Statistical Computing, Vienna, Austria). The program accepted the spreadsheet output from kdMining in the designated directory and processed one spreadsheet per execution. The program first parsed through the spreadsheet to ensure all needed columns, such as glycan chart number, glycan linkage and type, K_A, and % max, are present in the file, then mapped all information onto the nested-circles. At the end of program execution, a total of three nested-circles were plotted based on the glycan structure: N-glycans, Lewis Antigen, and Sialic Acid are shown as examples. The program was designed such that each nested circle contains three layers of information: K_A, glycan linkage, and glycan structure, from inner to outer, respectively. Since K_A and % max were numeric variables, they were presented with different color gradients from white to red and white to blue, respectively. The latest kdMining, kaPlotting, instructions, and a demo video can be found and downloaded from Mendeley Data (DOI: 10.17632/hrbtct6ryd.2).
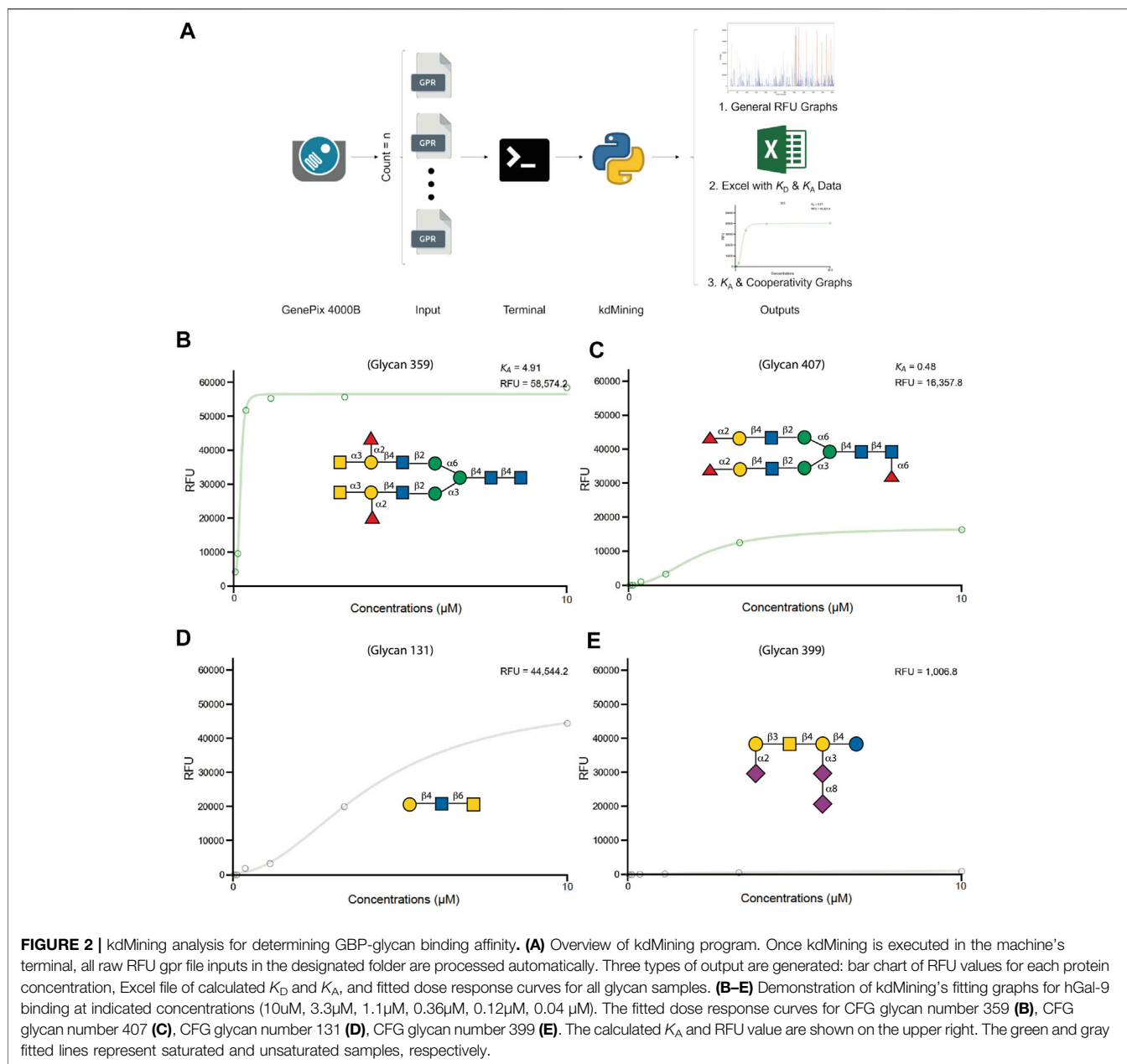
RESULTS

To directly evaluate the relevance of multi-concentration screening of GBP-glycan interaction, we assessed glycan binding of human Gal-3, Gal-3C, Gal-7, Gal-9, Gal-9N, and Gal-9C over a range of concentrations (indicated in figure legends) as done previously (Stowell et al., 2008). We next visualized glycan binding as a heat map plot of RFUs detected in selected glycan binding. To better assess glycan binding patterns, complex glycan structures were grouped by common terminal glycan modification. When viewed this way, clear binding patterns emerge for each galectin tested, with lactosamine and blood group antigens appearing as the dominant terminal modification recognized (**Figures 1A–C**). These general binding capabilities can be appreciated by considering binding at the highest concentrations alone. However, evaluation of glycan binding at multiple concentrations reveals greater preference for particular terminal modifications as well as the important contribution of underlying glycan structural variation on true binding preference. In particular, hGal-7 binding provides a strong example of this limitation of assessment at only high concentrations as various glycan ligands appear equally preferred

when considered at 10 µM alone, while dose dependent binding reveals exquisite specificity for particular presentations of the H antigen at lower concentrations (**Figure 1B**).

While assessment of glycan binding over multiple concentrations can offer significantly more insight than single concentration evaluation only, plotting data and evaluating binding patterns by manual calculation of average RFU for each dilution, followed by plotting the calculated RFU in a spreadsheet by density gradient (heatmap format) for visualization, represented large task. On average, to process and analyze one sample took 3.5 min, which translates to full dataset analyses of 555 glycans for all logistic patterns with curves requiring at least 1940 min (32.3 h). Additionally, even though this determination approach is commonly adopted, the possibility for human error and time demanded represented significant drawbacks to this method. To increase productivity and efficiency while reducing potential errors for K_A determination, we adopted a programming approach to expedite the process (**Figure 2**). To achieve this, we utilized an open-source programming language, Python 3, which offers the most substantial flexibility in time complexity and data structure. The output of the program includes general RFU graphs for each dilution, curve-fitting graphs for each sample, and calculated B_{max}, K_D, K_A, and % max binding for unsaturated glycan ligands in a spreadsheet file using the same approach as outlined previously (Blenda et al., 2022; Wu et al., 2021).
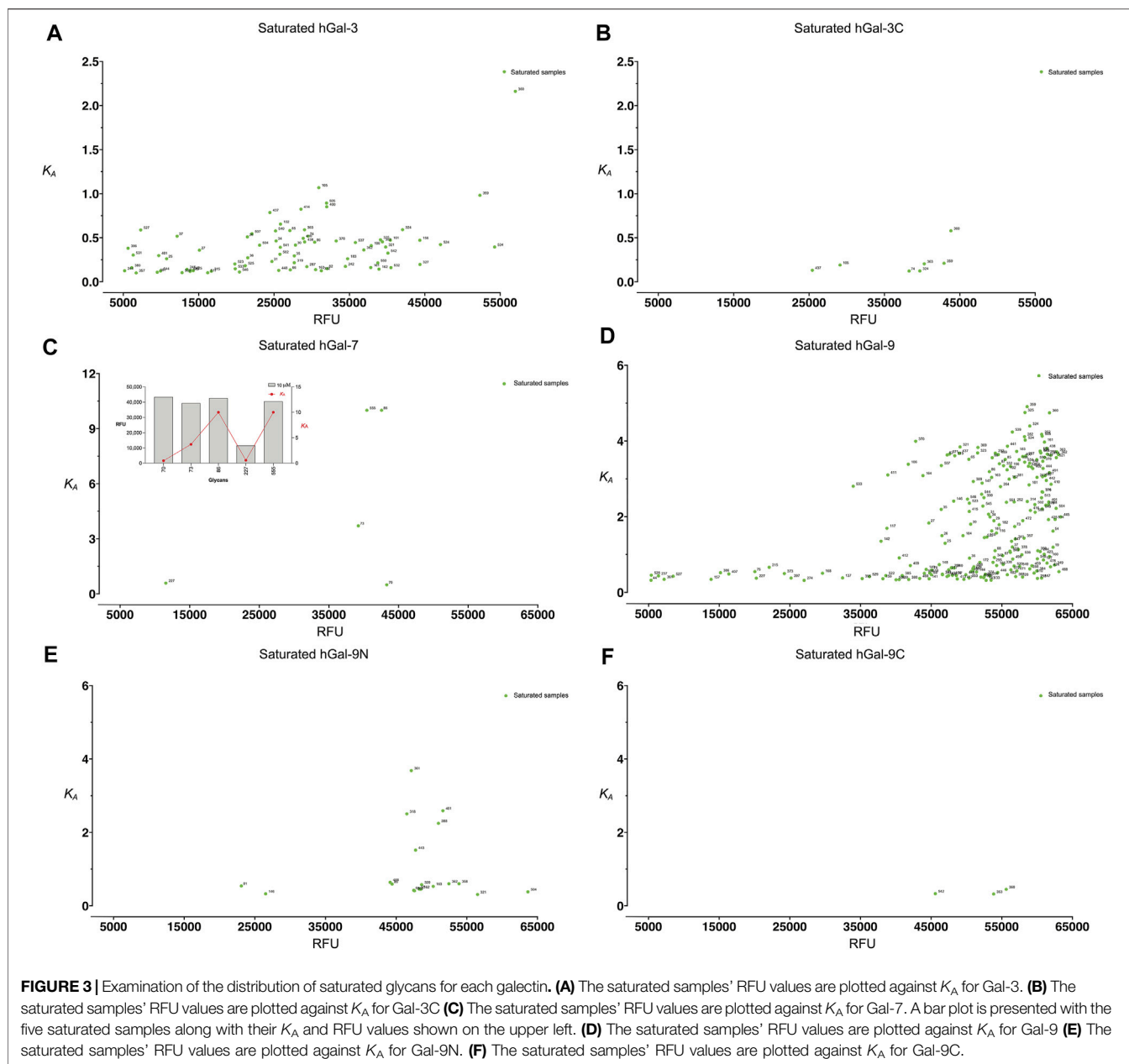
To demonstrate the programs' performance, we fed the raw RFU data (GPR files) of human Gal-3, Gal-3C, Gal-7, Gal-9, Gal-9N, and Gal-9C into kdMining and got their corresponding calculated average B_{max}, K_D, K_A, and % max (**Supplementary Table S1**) This program automatically produces individual RFU plots for each glycan represented on the array, with fitted curves and estimated K_A, as well as the B_{max} for each plot. Each sample's concentrations were plotted against their calculated average RFU values and fitted to a dose response curve. For demonstration purposes, we plotted average RFU values of Gal-9 and presented the fitted curves for several Gal-9 bound glycans. A green fitted line indicates that GBP binding over the multiple tested concentrations resulted in glycan binding saturation, while a gray fitted line indicates GBP binding that occurred over the multiple tested concentrations tested, but for which saturation was not achieved. To further investigate the relationship between RFU values obtained at the highest concentration tested as a rank order approach and K_A values obtained for the saturated samples, we plotted the RFU at the highest concentration tested against K_A for all galectins analyzed. Among all galectins, Gal-3 and Gal-9 binding resulted in saturation of many more glycans when compared to Gal-3C, Gal-7, Gal-9N, and Gal-9C (**Figure 3**). Finally, to reassess the effectiveness of kdMining in identifying accurate preferential GBP binding, we next plotted K_A values against maximum RFU for the top 60 bound glycans. Glycans were listed in order from highest to lowest maximum RFU using the rank order approach and corresponding K_A values were also shown as calculated by kdMining. Parallel analysis of K_A and maximum RFU revealed a poor match overall between K_A and maximum RFU (**Figure 4**). This analysis was also completed with glycans ranked by K_A, showing similar results (**Supplementary Figure S1**). Importantly, preferential binding indicated by K_A calculations within our

FIGURE 2 | kdMining analysis for determining GBP-glycan binding affinity. **(A)** Overview of kdMining program. Once kdMining is executed in the machine's terminal, all raw RFU gpr file inputs in the designated folder are processed automatically. Three types of output are generated: bar chart of RFU values for each protein concentration, Excel file of calculated K_D and K_A, and fitted dose response curves for all glycan samples. **(B–E)** Demonstration of kdMining's fitting graphs for hGal-9 binding at indicated concentrations (10uM, 3.3μM, 1.1μM, 0.36μM, 0.12μM, 0.04 μM). The fitted dose response curves for CFG glycan number 359 **(B)**, CFG glycan number 407 **(C)**, CFG glycan number 131 **(D)**, CFG glycan number 399 **(E)**. The calculated K_A and RFU value are shown on the upper right. The green and gray fitted lines represent saturated and unsaturated samples, respectively.

program matched the highest binders identified by manual analysis of binding at multiple concentrations (**Figure 1**).

Although kdMining was able to determine useful information regarding glycan-protein interactions, having the calculated values alone made it difficult to visualize the differences in the level of saturation and binding affinity when comparin multiple galectins. To address this limitation, we used an open-source programming language, R, which is known for its aesthetic graphing and plotting and strong statistical analysis to create a second programming step, kaPlotting, for data presentation (**Figure 5**). This program directly intakes the spreadsheet output from kdMining, and parses through the calculated values once to ensure all information is present. After each program execution, a total of three three-layered nested pie charts were drawn. Each nested pie chart consisted of similar

families of glycan structures, and the layers showed the calculated K_A, glycan linkage, and glycan structure for all glycan structures, from inner to outer, respectively. For demonstration purposes, we selected three pie charts: each of which primarily focus on N-glycans, blood group antigens and related structures, and sialic acid containing glycans, that had the most bindings for human Gal-3, Gal-7, Gal-9, Gal-3C, Gal-9N, and Gal-9C (**Figures 5B–D**, **Supplementary Figure S2A–C**). Within the K_A layer, there are two sets of gradients: red and blue. Red gradients indicate the K_A value for saturated samples, while blue gradients represent % max for unsaturated samples, and white indicates that no significant binding was detected. Since the glycan linkage and glycan structure layers contained categorical variables, discrete color schemes were used. As shown in the nested pie charts,

FIGURE 3 | Examination of the distribution of saturated glycans for each galectin. **(A)** The saturated samples' RFU values are plotted against K_A for Gal-3. **(B)** The saturated samples' RFU values are plotted against K_A for Gal-3C **(C)** The saturated samples' RFU values are plotted against K_A for Gal-7. A bar plot is presented with the five saturated samples along with their K_A and RFU values shown on the upper left. **(D)** The saturated samples' RFU values are plotted against K_A for Gal-9 **(E)** The saturated samples' RFU values are plotted against K_A for Gal-9N. **(F)** The saturated samples' RFU values are plotted against K_A for Gal-9C.

Gal-3 bound to more glycan samples compared to Gal-3C, and its proportion of having saturated binding was also higher than human Gal-3C (**Figure 5B**, **Supplementary Figure S2A**). Gal-7 had much more unsaturated samples than the saturated ones (**Figure 5C**). Gal-9 bound to and saturated with a great number of glycans compared to Gal-9N and Gal-9C (**Figure 5D**, **Supplementary Figure S2B, C**).

While the information provided by kdMining and kaPlotting yielded accurate and efficient evaluation of GBP-glycan binding, we next wanted to evaluate the reduction in turnaround time needed to obtain K_A results from raw RFU values using KdMining. To accomplish this, we created different sizes of datasets and conducted multiple trials to test the analysis time between these two approaches

(**Figure 6**). A runtime table was included to demonstrate the time required for each approach to execute the given number of samples (**Figure 6A**). As shown in the graph, even though both approaches had a linear runtime, the manual approach would take roughly 600 times longer on average than our approach using kdMining (**Figure 6B**). To analyze one sample, the runtimes of the manual approach and kdMining are 3.50 and 0.01 min, respectively. Furthermore, when the number of samples increased to 555, the manual approach took more than 1939 min (32.3 h), while kdMining took less than 1.7 min to complete the same amount of work. Based on these results, kdMining appears to be an effective and efficient approach to analyzing K_A and binding affinity following glycan microarray analysis.
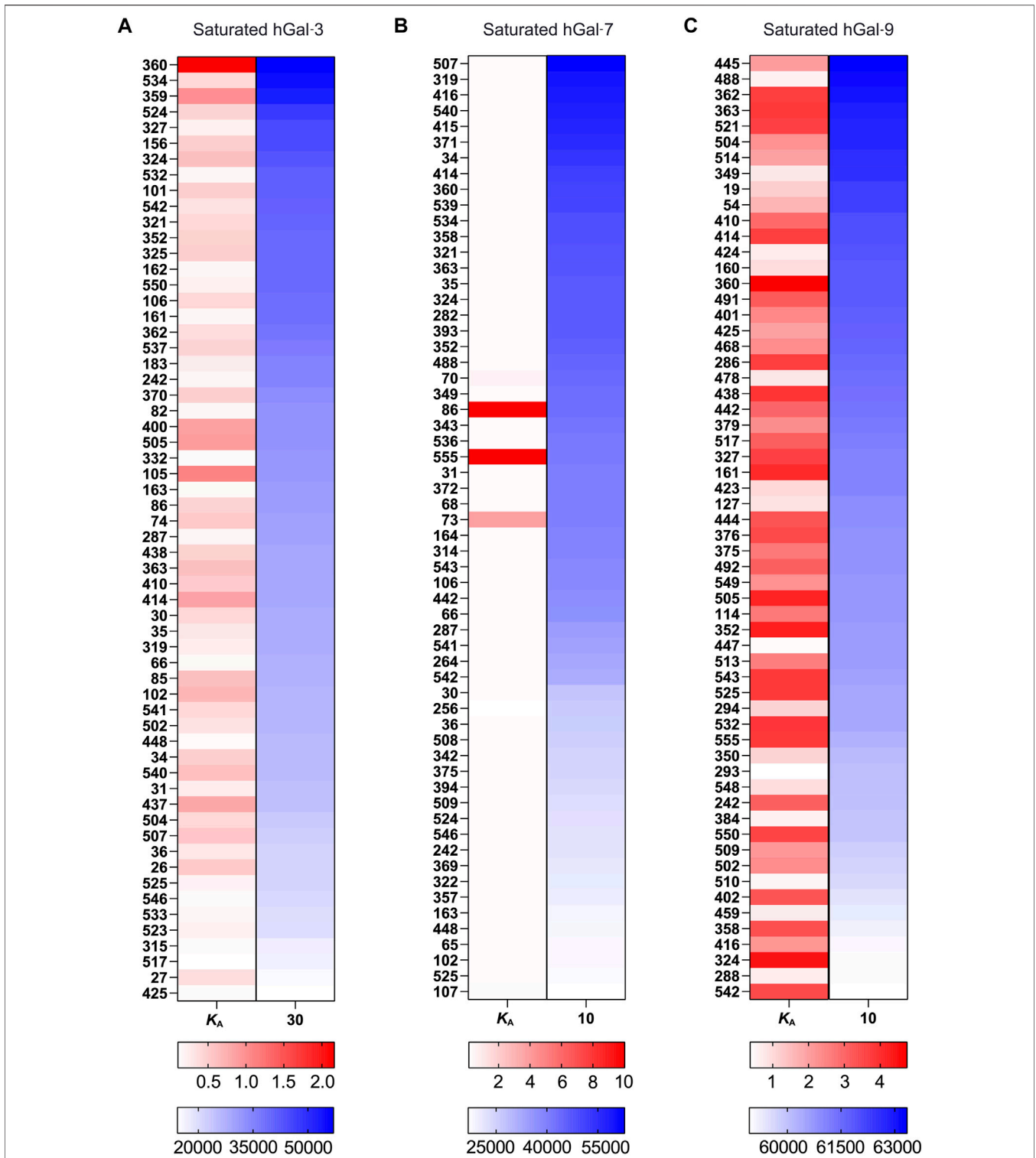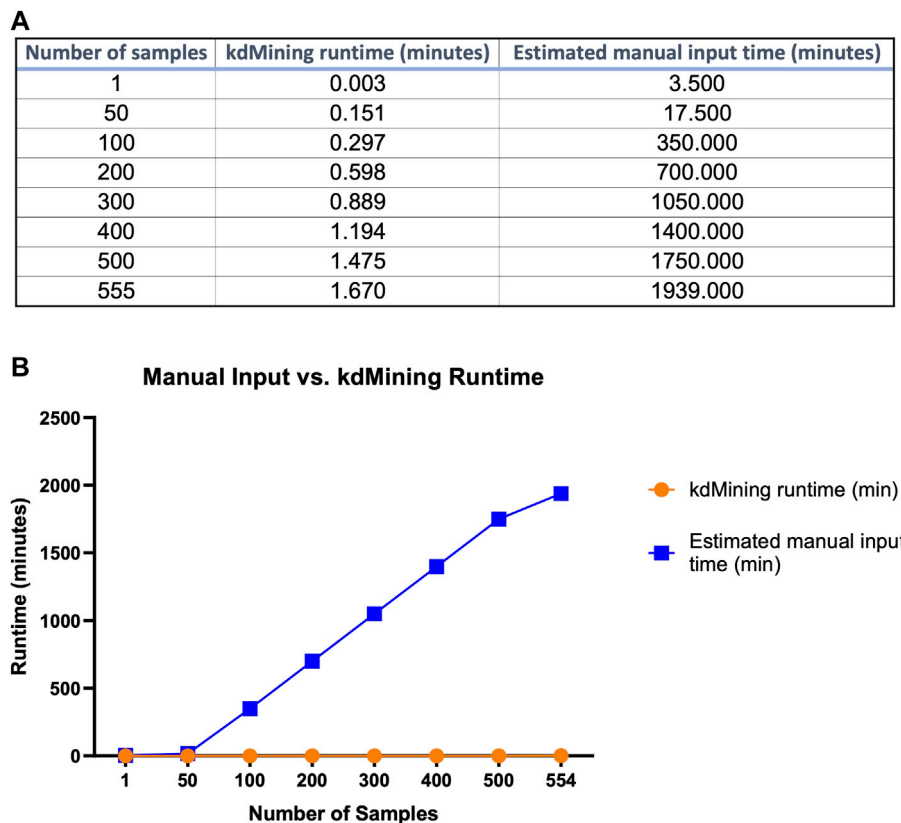
FIGURE 4 | Examination of K_A and RFU correlation for selected galectin **(A)** A correlation map with the top 60 samples' K_A and RFU values, sorted by RFU, for Gal-3. **(B)** A correlation map with the top 60 samples' K_A and RFU values, sorted by RFU, for Gal-7. **(C)** A correlation map with the top 60 samples' K_A and RFU values, sorted by RFU, for Gal-9. There are two sets of gradients: red and blue to indicate binding strength, for K_A and RFU, respectively. For both gradients, darker tint = higher calculated value, lighter tint = lower calculated value.
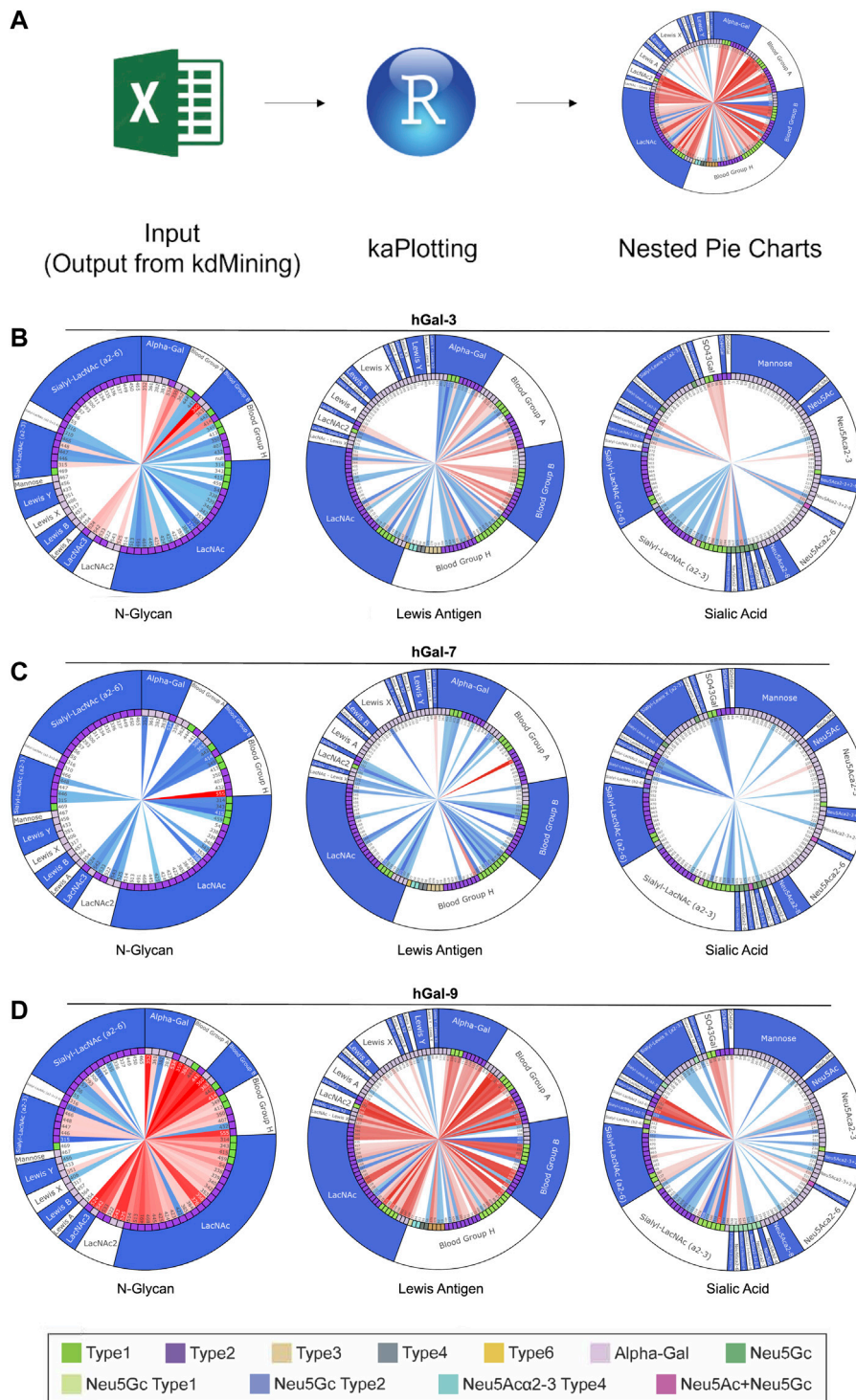
FIGURE 5 | Demonstration of kaPlotting's nested pie charts. **(A)** Overview of kaPlotting intakes and processing steps. A total of three three-layer-nested pie charts that consist of K_A, glycan linkage, and glycan structure are generated. **(B–D)** Nested pie charts shown for Gal-3 **(B)**, Gal-7 **(C)** and Gal-9 **(D)**. Charts are grouped by N-Glycan, Blood Group Antigen, and Sialic Acid, and plotted using kaPlotting. Each nested pie chart has three layers of information: K_A, glycan linkage, and glycan structure, from inner to outer, respectively. There are two sets of gradients: red and blue, that represent K_A and % max respectively on the inner circle. For both gradients, darker tint = higher calculated value, lighter tint = lower calculated value. Legend at the bottom shows the glycan linkage (second layer on the nested pie charts). Type 1: Type 1 LacNAc and type 2: Type 2 LacNAc.

A

Number of samples	kdMining runtime (minutes)	Estimated manual input time (minutes)
1	0.003	3.500
50	0.151	17.500
100	0.297	350.000
200	0.598	700.000
300	0.889	1050.000
400	1.194	1400.000
500	1.475	1750.000
555	1.670	1939.000

B

FIGURE 6 | Execution time comparison of kdMining and manual approach. **(A)** To process and analyze 555 samples, kdMining takes less than 1.67 min, whereas the manual approach takes more than 32.3 h. **(B)** Both kdMining and the manual approach have a linear runtime, but kdMining can process information at a much faster rate.

DISCUSSION

Given the post-translational nature of glycan modifications (Cummings, 2009), defining GBP interactions with distinct glycan determinants has been challenging (Johannes et al., 2018). A variety of approaches, including isothermal calorimetry, surface plasmon resonance, equilibrium dialysis, fluorescence polarization and frontal affinity chromatography, have been used to measure the actual affinity of a given GBP with a specific glycan ligand (Ahmad et al., 2004; Brewer, 2002; Morris et al., 2004; Sörme et al., 2004; Hirabayashi et al., 2002; Leppänen et al., 2005; Mehta-D'souza, 2015). While each of these methods certainly has advantages over the approach outlined in the present work, there are significant challenges to using these strategies in a high throughput manner. Glycan array technology appeared to solve some of the throughput limitations of prior approaches by increasing the number of glycan determinants that can be analyzed following incubation with a single GBP(Rillahan and Paulson, 2011; Gao et al., 2019). Glycan microarrays also allow analyses to be achieved with very small amounts of glycan material, expanding the use of source materials for more GBP analyses (Rillahan and Paulson, 2011).

Despite the success of microarray analysis, common strategies employed in array analysis certainly have limitations. Most glycan array analysis is performed at a single GBP concentration, followed by rank order analysis to determine the glycans that appear to possess the highest affinity interaction with the GBP examined (Jonas, 1975; Delso et al., 2016). This approach has simplified glycan binding analysis and in so doing, provided a wealth of information regarding the key glycan determinants bound by individual GBPs(Bochner et al., 2005; Stevens et al., 2006; Carlsson et al., 2007; Stowell et al., 2010; Stowell et al., 2014; Wesener et al., 2015; Noll et al., 2016). However, possible glycan printing irregularities between slide manufacturing attempts or even intrinsic differences in printing efficiencies between individual glycans themselves can result in variation in the amount of glycan printed. This raises the possibility that differences in glycan binding achieved following GBP incubation could in part be influenced by the glycan print density and not the relative affinity of the GBP for the individual glycan. This can result in variation in the apparent strength of binding that is not a consequence of intrinsic glycan preference. As batch-to-batch glycan array printing can differ with respect to glycan printing efficiency, it can be difficult to normalize results to reduce bias introduced by uneven glycan

printing; no positive control for glycan presentation and printing efficiency exists for every glycan. As a result, the use of single GBP concentration approaches, while capable of providing important insight into general binding capabilities for a given GBP, may miss more subtle GBP binding preferences. Such binding preferences may be important biologically, as relatively simple glycan modifications may alter GBP-glycan binding dynamics in ways that have real consequences on the overall outcome of GBP binding to a cell surface (Carlsson et al., 2007; Stowell et al., 2008c).

In an effort to overcome inherent limitations both in glycan microarray manufacturing and single concentration approaches to GBP glycan binding analysis, we and others have examined GBP binding over a range of concentrations (Stowell et al., 2008a). Using this approach, the B_{max} not only provides a more accurate reflection of the relative amount of a particular glycan that may be present, but also can be used to extrapolate a binding isotherm and K_D value to provide a relative affinity constant for GBP interactions with a given glycan (Garcia et al., 2014; Alhazmi, 2018; Huang et al., 2021). In doing so, the relative affinity of a GBP for a given glycan can be determined regardless of unintended variations in the amount of glycan printed. In this way, relative K_D assessment allows analysis approaches to more effectively control for glycan print variation while also providing a more accurate measure of the relative binding affinity of a GBP for a glycan ligand. The utility of this approach can be appreciated by the relative lack of correlation between K_A values and the maximum RFU obtained for each galectin analyzed. Thus, using this approach, more accurate assessments of binding affinity can be obtained. Importantly, using this approach, predictions regarding glycan binding specificity have accurately predicted actual glycan binding to cell surface glycans, providing important insight into previously unrecognized activities of a given galectin (Stowell et al., 2008a; Blenda et al., 2022).

Despite the possible utility of relative affinity measurements following glycan microarray analysis, the overall process required to analyze data in this manner is resource intense. Current programing packages for array analysis are designed to integrate spot intensity into an RFU value, which can be exported into an excel file for easy initial evaluation of glycan binding. This approach has been used for over a decade to assess glycan binding and can provide a rank order analysis of glycans bound by the GBP analyzed. However, for B_{max} and K_A value extrapolation of array data obtained over a range of GBP concentrations, there is no readily available automated approach capable of directly interfacing with raw microarray output data to generate these values. As a result, B_{max} and K_A value determinations are limited to either manual calculations based on array analyses, which often requires custom configuration of an excel platform, or the employment of commercially available software such as Prism or Sigmaplot. These later programs are user-friendly and can readily assist in B_{max} and K_A calculations. However, the manual input of data and limitations surrounding the number of glycans analyzed within an analysis file can make this approach cumbersome and vulnerable to common errors introduced by human data entry, especially when considering possible GBP interactions with

over 500 individual glycans are analyzed. To facilitate this process, we generated kdMining, which allows K_D values to be directly extrapolated from raw microarray output data within a fraction of the time required for manual input. Using this approach, the relatively binding affinity toward individual glycans can be readily assessed within a fraction of the time required for manual data entry.

In addition to challenges assessing the actual binding affinity toward a given glycan, the sheer volume of data obtained following glycan microarray analysis can make data analysis difficult to manage and interpret. Rank order analysis does provide the highest binding glycans, but it fails to allow equally accessible analysis of those glycans to which a given GBP fails to bind. Negative binding outcomes are often just as critical as positive interactions as these data can provide critical information regarding the impact of key glycan binding determinants required for glycan binding (Carlsson et al., 2007). To facilitate K_A interpretation and therefore overall GBP binding specificity, we developed kaPlotting, which bins glycan binding data accordingly to common glycan determinants in order to facilitate interpretation of possible differences in GBP binding due to subtle variation in glycan presentation. By configuring the data in this manner, distinct GBPs can be directly compared and visually assessed. However, when glycan binding saturation does not occur, but interactions are present, these binding events are acknowledged through a distinct color code using the percent maximal binding similar to what was previously used to perform rank order analysis of glycan binding. By separating K_A and %max binding in this manner, higher affinity interactions can be observed in red, while variation in lower affinity binding can still be perceived within shades of blue. By combining this approach with recently developed analytical tools for glycan array analyses, overlapping and distinct glycan binding preferences between unique GBPs begin to become apparent (Mehta and Cummings, 2019).

In summary, the glycan binding profiles established following glycan binding analysis over a range of concentrations on existing glycan microarray technology has been shown to be a useful strategy in assessing GBP binding preference. However, limitations in the time required for analysis have prevented this approach from being commonly employed. Use of the kdMining and kaPlotting tools may aid in this analysis, allowing relative affinities to be obtained following GBP analysis towards hundreds of glycan determinants. Additionally, the developed programs could likewise be applied to a variety of microarray formats such as protein microarrays and DNA microarrays. These microarrays can be manufactured with the same microarray printer as the glycan microarray and analysis is often achieved using a similar approach. It should be noted that despite the possible advantages of coupling K_A calculations with glycan microarray analysis, this approach simply provides a *relative* K_A value. Actual K_A values will require confirmatory studies using more refined approaches, as outlined previously (Brewer, 2002; Hirabayashi et al., 2002; Ahmad et al., 2004; Morris et al., 2004; Sörme et al., 2004; Leppänen et al., 2005), to detail the binding affinity of a GBP toward individual glycans. Furthermore, there are clearly caveats to this approach and the impact of antigen printing densities and additional consideration could certainly influence the overall apparent binding affinity observed for a given GBP as stated

previously (Zhang et al., 2010; Godula and Bertozzi, 2012; Tsouka et al., 2021). While the purpose of this study was to provide an automated approach to analyzing relative GBP binding affinities towards immobilized glycans in an array format, this is still a screening approach and exhaustive analysis of GBP glycan binding preferences often requires the utilization of multiple modalities, including surface plasmon resonance, isothermal calorimetry, fluorescent polarization, and other approaches (Brewer, 2002; Stowell et al., 2008a; Kumar et al., 2019), to appreciate the full binding characteristics. As these approaches often require higher concentrations of glycans or glycan modifications, binding isotherms generated following array analysis can provide an important triaging approach to determine which glycan determinants warrant additional consideration using more refined approaches. Equally important, array data and more refined approaches designed to define the glycan binding specificity of a given GBP require critical confirmatory experiments to determine whether binding affinities and overall specificity observed translate to glycan binding on a cell surface. By using combined approaches of array analysis, additional biochemical studies and ultimately biological approaches, a picture of the binding specificity of a GBP can emerge that provides important insight into its overall function.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

REFERENCES

Acipreste Hudson, E., Campos de Paula, H. M., Coelho, Y. L., Glanzmann, N., da Silva, A. D., Mendes da Silva, L. H., et al. (2022). The Kinetics of Formation of resveratrol-β-cyclodextrin-NH2 and Resveratrol analog-β-cyclodextrin-NH2 Supramolecular Complexes. *Food Chem.* 366, 130612. doi:10.1016/j.foodchem.2021.130612

Ahmad, N., Gabius, H. J., Sabesan, S., Oscarson, S., and Brewer, C. F. (2004). Thermodynamic Binding Studies of Bivalent Oligosaccharides to Galectin-1, Galectin-3, and the Carbohydrate Recognition Domain of Galectin-3. *Glycobiology* 14 (9), 817–825. doi:10.1093/glycob/cwh095

Alhazmi, H. A. (2018). Measurement of Interaction Behavior of Six Biologically Important Noble Metal Ions with the Iron(III) Binding Protein, Apo-Transferrin, Using Mobility-Shift Affinity Electrophoresis. *Pharmazie* 73 (3), 143–149. doi:10.1691/ph.2018.7179

Arthur, C. M., Baruffi, M. D., Cummings, R. D., and Stowell, S. R. (2015). Evolving Mechanistic Insights into Galectin Functions. *Methods Mol. Biol.* 1207, 1–35. doi:10.1007/978-1-4939-1396-1_1

Arthur, C. M., Cummings, R. D., and Stowell, S. R. (2014). Using Glycan Microarrays to Understand Immunity. *Curr. Opin. Chem. Biol.* 18, 55–61. doi:10.1016/j.cbpa.2013.12.017

Arthur, C. M., Rodrigues, L. C., Baruffi, M. D., Sullivan, H. C., Heimburg-Molinaro, J., Smith, D. F., et al. (2015). Examining Galectin Binding Specificity Using Glycan Microarrays. *Methods Mol. Biol.* 1207, 115–131. doi:10.1007/978-1-4939-1396-1_8

Arthur, C. M., Zerra, P. E., Shin, S., Wang, J., Song, X., Doering, C. B., et al. (2022). Nonhuman Glycans Can Regulate Anti-factor VIII Antibody Formation in Mice. *Blood* 139 (9), 1312–1317. doi:10.1182/blood.2020009210

AUTHOR CONTRIBUTIONS

SS, AH and S-CW designed key experiments, which were conducted by AH and S-CW. SS and CA provided critical oversight in the design and execution of these studies, which was additionally commented on by NK, AB and RC. AH, S-CW, CA and SS wrote the manuscript, which was additionally commented on and edited by the remaining authors.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.893185/full#supplementary-material

Blenda, A. V., Kamili, N. A., Wu, S.-C., Abel, W. F., Ayona, D., Gerner-Smidt, C., et al. (2022). Galectin-9 Recognizes and Exhibits Antimicrobial Activity toward Microbes Expressing Blood Group-like Antigens. *J. Biol. Chem.* 298, 101704. doi:10.1016/j.jbc.2022.101704

Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., et al. (2004). Printed Covalent Glycan Array for Ligand Profiling of Diverse Glycan Binding Proteins. *Proc. Natl. Acad. Sci. U.S.A.* 101 (49), 17033–17038. doi:10.1073/pnas.0407902101

Bochner, B. S., Alvarez, R. A., Mehta, P., Bovin, N. V., Blixt, O., White, J. R., et al. (2005). Glycan Array Screening Reveals a Candidate Ligand for Siglec-8*. *J. Biol. Chem.* 280 (6), 4307–4312. doi:10.1074/jbc.M412378200

Bochner, B. S., and Zimmermann, N. (2015). Role of Siglecs and Related Glycan-Binding Proteins in Immune Responses and Immunoregulation. *J. Allergy Clin. Immunol.* 135 (3), 598–608. doi:10.1016/j.jaci.2014.11.031

Brewer, C. F. (2002). Thermodynamic Binding Studies of Galectin-1, -3 and -7. *Glycoconj J.* 19 (7-9), 459–465. doi:10.1023/B:GLYC.0000014075.62724.d0

Carlsson, S., Öberg, C. T., Carlsson, M. C., Sundin, A., Nilsson, U. J., Smith, D., et al. (2007). Affinity of Galectin-8 and its Carbohydrate Recognition Domains for Ligands in Solution and at the Cell Surface. *Glycobiology* 17 (6), 663–676. doi:10.1093/glycob/cwm026

Cummings, R. D., and Esko, J. D., (2009). "Principles of Glycan Recognition," in *Essentials of Glycobiology*. Editor A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, and M. E. Etzler (New York: Cold Spring Harbor)

Cummings, R. D. (2009). The Repertoire of Glycan Determinants in the Human Glycome. *Mol. Biosyst.* 5 (10), 1087–1104. doi:10.1039/b907931a

Delso, I., Valero-González, J., Marca, E., Tejero, T., Hurtado-Guerrero, R., and Merino, P. (2016). Rational Design of Glycomimetic Compounds Targeting theSaccharomyces cerevisiaeTransglycosylase Gas2. *Chem. Biol. Drug Des.* 87 (2), 163–170. doi:10.1111/cbdd.12650

Elola, M. T., Wolfenstein-Todel, C., Troncoso, M. F., Vasta, G. R., and Rabinovich, G. A. (2007). Galectins: Matricellular Glycan-Binding Proteins Linking Cell Adhesion, Migration, and Survival. *Cell. Mol. Life Sci.* 64 (13), 1679–1700. doi:10.1007/s00018-007-7044-8

Gao, C., Wei, M., McKitrick, T. R., McQuillan, A. M., Heimburg-Molinaro, J., and Cummings, R. D. (2019). Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front. Chem.* 7, 833. doi:10.3389/fchem.2019.00833

Garcia, J.-M., Lai, J. C. C., Haselhorst, T., Choy, K. T., Yen, H.-L., Peiris, J. S. M., et al. (2014). Investigation of the Binding and Cleavage Characteristics of N1 Neuraminidases from Avian, Seasonal, and Pandemic Influenza Viruses Using Saturation Transfer Difference Nuclear Magnetic Resonance. *Influenza Other Respi Viruses* 8 (2), 235–242. doi:10.1111/irv.12184

Godula, K., and Bertozzi, C. R. (2012). Density Variant Glycan Microarray for Evaluating Cross-Linking of Mucin-like Glycoconjugates by Lectins. *J. Am. Chem. Soc.* 134 (38), 15732–15742. doi:10.1021/ja302193u

Harding, S. E. (2021). Analytical Ultracentrifugation as a Matrix-free Probe for the Study of Kinase Related Cellular and Bacterial Membrane Proteins and Glycans. *Molecules* 26 (19), 6080. doi:10.3390/molecules26196080

Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., et al. (2002). Oligosaccharide Specificity of Galectins: a Search by Frontal Affinity Chromatography. *Biochim. Biophys. Acta* 1572 (2-3), 232–254. doi:10.1016/s0304-4165(02)00311-2

Huang, L., Zhang, Z., and McMacken, R. (2021). Interaction of the Escherichia coli HU Protein with Various Topological Forms of DNA. *Biomolecules* 11 (11), 1724. doi:10.3390/biom11111724

Islam, Z., Gilbert, M., Mohammad, Q. D., Klaij, K., Li, J., van Rijs, W., et al. (2012). Guillain-Barré Syndrome-Related Campylobacter Jejuni in Bangladesh: Ganglioside Mimicry and Cross-Reactive Antibodies. *PLoS One* 7 (8), e43976. doi:10.1371/journal.pone.0043976

Johannes, L., Jacob, R., and Leffler, H. (2018). Galectins at a Glance. *J. Cell Sci.* 131 (9), jcs208884. doi:10.1242/jcs.208884

Jonas, A. (1975). Self-association of the Major Protein Component of Bovine Serum High Density Lipoprotein. *Biochimica Biophysica Acta (BBA) - Protein Struct.* 393 (2), 471–482. doi:10.1016/0005-2795(75)90075-6

Kumar, R., Ignjatović, M. M., Peterson, K., Olsson, M., Leffler, H., Ryde, U., et al. (2019). Structure and Energetics of Ligand-Fluorine Interactions with Galectin-3 Backbone and Side-Chain Amides: Insight into Solvation Effects and Multipolar Interactions, *ChemMedChem* 14, 1528–1536. doi:10.1002/cmdc.201900293

Leppänen, A., Stowell, S., Blixt, O., and Cummings, R. D. (2005). Dimeric Galectin-1 Binds with High Affinity to α2,3-Sialylated and Non-sialylated Terminal N-Acetyllactosamine Units on Surface-Bound Extended Glycans. *J. Biol. Chem.* 280 (7), 5549–5562. doi:10.1074/jbc.M412019200

McQuillan, A. M., Byrd-Leotis, L., Heimburg-Molinaro, J., and Cummings, R. D. (2019). Natural and Synthetic Sialylated Glycan Microarrays and Their Applications. *Front. Mol. Biosci.* 6, 88. doi:10.3389/fmolb.2019.00088

Mehta, A. Y., and Cummings, R. D. (2019). GLAD: GLycan Array Dashboard, a Visual Analytics Tool for Glycan Microarrays. *Bioinformatics* 35 (18), 3536–3537. doi:10.1093/bioinformatics/btz075

Mehta-D'souza, P. (2015). Evaluation of Galectin Binding by Surface Plasmon Resonance. *Methods Mol. Biol.* 1207, 105–114. doi:10.1007/978-1-4939-1396-1_7

Morris, S., Ahmad, N., Andre, S., Kaltner, H., Gabius, H. J., Brenowitz, M., et al. (2004). Quaternary Solution Structures of Galectins-1, 3, and-7. *Glycobiology* 14 (3), 293–300. doi:10.1093/glycob/cwh029

Noll, A. J., Yu, Y., Lasanajak, Y., Duska-McEwen, G., Buck, R. H., Smith, D. F., et al. (2016). Human DC-SIGN Binds Specific Human Milk Glycans. *Biochem. J.* 473 (10), 1343–1353. doi:10.1042/BCJ20160046

Paul, A., Wu, S.-C., Patel, K. R., Ho, A. D., Allen, J. W. L., Verkerke, H., et al. (2022). Purification of from Different Species Using Distinct Methods. *Methods Mol. Biol.* 2442, 55–74. doi:10.1007/978-1-0716-2055-7_3

Perkins, S. J., Nan, R., Li, K., Khan, S., and Miller, A. (2012). Complement Factor H-Ligand Interactions: Self-Association, Multivalency and Dissociation Constants. *Immunobiology* 217 (2), 281–297. doi:10.1016/j.imbio.2011.10.003

Rabinovich, G. A., and Toscano, M. A. (2009). Turning 'sweet' on Immunity: Galectin-Glycan Interactions in Immune Tolerance and Inflammation. *Nat. Rev. Immunol.* 9 (5), 338–352. doi:10.1038/nri2536

Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C., and Sasisekharan, R. (2005). Glycomics: an Integrated Systems Approach to Structure-Function Relationships of Glycans. *Nat. Methods* 2 (11), 817–824. doi:10.1038/nmeth807

Rillahan, C. D., and Paulson, J. C. (2011). Glycan Microarrays for Decoding the Glycome. *Annu. Rev. Biochem.* 80, 797–823. doi:10.1146/annurev-biochem-061809-152236

Sato, S., St-Pierre, C., Bhaumik, P., and Nieminen, J. (2009). Galectins in Innate Immunity: Dual Functions of Host Soluble β-galactoside-binding Lectins as Damage-Associated Molecular Patterns (DAMPs) and as Receptors for Pathogen-Associated Molecular Patterns (PAMPs). *Immunol. Rev.* 230 (1), 172–187. doi:10.1111/j.1600-065X.2009.00790.x

Schnaar, R. L. (2015). Glycans and Glycan-Binding Proteins in Immune Regulation: A Concise Introduction to Glycobiology for the Allergist. *J. Allergy Clin. Immunol.* 135 (3), 609–615. doi:10.1016/j.jaci.2014.10.057

Song, X., Xia, B., Stowell, S. R., Lasanajak, Y., Smith, D. F., and Cummings, R. D. (2009). Novel Fluorescent Glycan Microarray Strategy Reveals Ligands for Galectins. *Chem. Biol.* 16 (1), 36–47. doi:10.1016/j.chembiol.2008.11.004

Sörme, P., Kahl-Knutsson, B., Huflejt, M., Nilsson, U. J., and Leffler, H. (2004). Fluorescence Polarization as an Analytical Tool to Evaluate Galectin-Ligand Interactions. *Anal. Biochem.* 334 (1), 36–47. doi:10.1016/j.ab.2004.06.042

Stevens, J., Blixt, O., Paulson, J. C., and Wilson, I. A. (2006). Glycan Microarray Technologies: Tools to Survey Host Specificity of Influenza Viruses. *Nat. Rev. Microbiol.* 4 (11), 857–864. doi:10.1038/nrmicro1530

Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J.-P., Heimburg-Molinaro, J., et al. (2010). Innate Immune Lectins Kill Bacteria Expressing Blood Group Antigen. *Nat. Med.* 16 (3), 295–301. doi:10.1038/nm.2103

Stowell, S. R., Arthur, C. M., McBride, R., Berger, O., Razi, N., Heimburg-Molinaro, J., et al. (2014). Microbial Glycan Microarrays Define Key Features of Host-Microbial Interactions. *Nat. Chem. Biol.* 10 (6), 470–476. doi:10.1038/nchembio.1525

Stowell, S. R., Arthur, C. M., Mehta, P., Slanina, K. A., Blixt, O., Leffler, H., et al. (2008). Galectin-1, -2, and -3 Exhibit Differential Recognition of Sialylated Glycans and Blood Group Antigens. *J. Biol. Chem.* 283 (15), 10109–10123. doi:10.1074/jbc.M709545200

Stowell, S. R., Qian, Y., Karmakar, S., Koyama, N. S., Dias-Baruffi, M., Leffler, H., et al. (2008). Differential Roles of Galectin-1 and Galectin-3 in Regulating Leukocyte Viability and Cytokine Secretion. *J. Immunol.* 180 (5), 3091–3102. doi:10.4049/jimmunol.180.5.3091

Stowell, S. R., Arthur, C. M., Slanina, K. A., Horton, J. R., Smith, D. F., and Cummings, R. D. (2008). Dimeric Galectin-8 Induces Phosphatidylserine Exposure in Leukocytes through Polylactosamine Recognition by the C-Terminal Domain. *J. Biol. Chem.* 283 (29), 20547–20559. doi:10.1074/jbc.M802495200

Stowell, S. R., Karmakar, S., Arthur, C. M., Ju, T., Rodrigues, L. C., Riul, T. B., et al. (2009). Galectin-1 Induces Reversible Phosphatidylserine Exposure at the Plasma Membrane. *MBoC* 20 (5), 1408–1418. doi:10.1091/mbc.E08-07-0786

Stowell, S. R., Cho, M., Feasley, C. L., Arthur, C. M., Song, X., Colucci, J. K., et al. (2009). Ligand Reduces Galectin-1 Sensitivity to Oxidative Inactivation by Enhancing Dimer Formation. *J. Biol. Chem.* 284 (8), 4989–4999. doi:10.1074/jbc.M808925200

Stowell, S. R., Karmakar, S., Stowell, C. J., Dias-Baruffi, M., McEver, R. P., and Cummings, R. D. (2007). Human Galectin-1, -2, and -4 Induce Surface Exposure of Phosphatidylserine in Activated Human Neutrophils but Not in Activated T Cells. *Blood* 109 (1), 219–227. doi:10.1182/blood-2006-03-007153

Stowell, S. R., Rodrigues, L. C., Dias-Baruffi, M., Cummings, R. D., and Arthur, C. M. (2022). Examining Galectin Binding Specificity Using. *Methods Mol. Biol.* 2442, 151–168. doi:10.1007/978-1-0716-2055-7_9

Tsouka, A., Hoetzel, K., Mende, M., Heidepriem, J., Paris, G., Eickelmann, S., et al. (2021). Probing Multivalent Carbohydrate-Protein Interactions with On-Chip Synthesized Glycopeptides Using Different Functionalized Surfaces. *Front. Chem.* 9, 766932. doi:10.3389/fchem.2021.766932

Vasta, G. R. (2012). Galectins as Pattern Recognition Receptors: Structure, Function, and Evolution. *Adv. Exp. Med. Biol.* 946, 21–36. doi:10.1007/978-1-4614-0106-3_2

Verkerke, H., Dias-Baruffi, M., Cummings, R. D., Arthur, C. M., and Stowell, S. R. (2022). Galectins: An Ancient Family of Carbohydrate Binding with Modern Functions. *Methods Mol. Biol.* 2442, 1–40. doi:10.1007/978-1-0716-2055-7_1

Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., et al. (2015). Recognition of Microbial Glycans by Human Intelectin-1. *Nat. Struct. Mol. Biol.* 22 (8), 603–610. doi:10.1038/nsmb.3053

Wu, S. C., Paul, A., Ho, A., Patel, K. R., Allen, J. W. L., Verkerke, H., et al. (2021a). Generation and Use of Recombinant Galectins. *Curr. Protoc.* 1 (3), e63. doi:10.1002/cpz1.63

Wu, S.-C., Ho, A. D., Kamili, N. A., Wang, J., Murdock, K. L., Cummings, R. D., et al. (2021b). Full-Length Galectin-3 Is Required for High Affinity Microbial Interactions and Antimicrobial Activity. *Front. Microbiol.* 12, 731026. doi:10.3389/fmicb.2021.731026

Wu, S.-C., Arthur, C. M., Wang, J., Verkerke, H., Josephson, C. D., Kalman, D., et al. (2021c). The SARS-CoV-2 Receptor-Binding Domain Preferentially Recognizes Blood Group A. *Blood Adv.* 5 (5), 1305–1309. doi:10.1182/bloodadvances.2020003259

Yang, R.-Y., Rabinovich, G. A., and Liu, F.-T. (2008). Galectins: Structure, Function and Therapeutic Potential. *Expert Rev. Mol. Med.* 10, e17. doi:10.1017/S1462399408000719

Zhang, Y., Li, Q., Rodriguez, L. G., and Gildersleeve, J. C. (2010). An Array-Based Method to Identify Multivalent Inhibitors. *J. Am. Chem. Soc.* 132 (28), 9653–9662. doi:10.1021/ja100608w