AD_____

Award Number: W81XWH-12-1-0279

TITLE: Synthetic Lectins: New Tools for Detection and Management of Prostate Cancer

PRINCIPAL INVESTIGATOR: Paul Thompson

CONTRACTING ORGANIZATION: The Scripps Research Institute
La Jolla, CA 92037-1000

REPORT DATE: August 2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
August 2013

2. REPORT TYPE
Annual

3. DATES COVERED
19 July 2012 - 18 July 2013

4. TITLE AND SUBTITLE
Synthetic Lectins: New Tools for Detection and Management of Prostate Cancer

5a. CONTRACT NUMBER
W81XWH-12-1-0279

5b. GRANT NUMBER
W81XWH-12-1-0279

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Paul R. Thompson

Email: pthompso@scripps.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Scripps Research Institute

La Jolla, CA 92037-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Among US men, prostate cancer is the most common cancer (besides non-malignant skin cancer), afflicting over 200,000 each year, and is the second leading cause of cancer-related death, over 30,000 per year. Thus, our long term goal is to develop synthetic lectin (SL) arrays for the detection and diagnosis of prostate cancer. We are pursuing this goal because healthy and diseased cells produce different biomarkers, which provide unique signatures by which these cells can be distinguished. Taking advantage of the fact that aberrant protein glycosylation is a hallmark of cancer; we propose to develop a novel sensor platform that can be used to detect Cancer Associated Glycans/Glycoproteins for the diagnosis of prostate cancer. The basis of this diagnostic is the differential display of boronic acids on peptides and peptoids. Boronic acids are used because they form covalent yet reversible bonds with specific structural motifs (i.e., diols) present on all Cancer Associated Glycans/Glycoproteins. The covalent interaction increases the affinity of the SL for the target Cancer Associated Glycans/Glycoproteins, while the peptide/peptoid backbone and preorganization of the boronic acids define the selectivity of binding. Building on preliminary data, which demonstrated the ability to identify synthetic lectins, assemble them into an array, and discriminate between normal, cancerous and metastatic colon cancer cell lines, we will: **(1)** generate synthetic lectins that recognize specific Cancer Associated Glycans/Glycoproteins; **(2)** probe the biochemical and biophysical basis for the glycan-SL interactions to enhance binding affinities and selectivities; and **(3)** create multi-component sensor arrays to differentiate cell and tissue types to diagnose and monitor prostate cancer. The development of these synthetic lectins is highly significant because they can be used to generate an array based diagnostic that has the potential to revolutionize the early diagnosis of prostate cancer.

15. SUBJECT TERMS
Lectins, prostate cancer, glycans, glycosylation.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	27	

Table of Contents

Introduction.

The overall goal of this proposal is to develop synthetic lectins (SLs) that bind to prostate cancer associated glycans and glycoproteins (CAGs). These studies are being pursued to develop this methodology into a robust system that can diagnose and monitor the stage of prostate cancer. Related to the proposed system, aberrant glycosylation is a hallmark of cancer and, as such, the differential display of boronic acid moieties on peptides and peptoids will allow for monitoring the changes (over- or neoexpression of CAGs) associated with oncogenesis and metastasis, thereby providing a new paradigm for the development of a prostate cancer diagnostic. AIM 1 describes a library based approach for the discovery of SLs targeting CAGs. AIM 2 describes biochemical and biophysical approaches to identify the factors that are required for the selective recognition of CAGs. It is expected that the results of these studies will provide information that will allow us to improve the design of the libraries described in AIM 1, towards second and third generation libraries. In AIM 3, selective and cross-reactive SLs will be assembled into an SL-based array. The efficacy of this array will be evaluated using both prostate cancer derived CAGs and actual cell lines.

Body.

Significant progress has been made in the prior funding period. Tasks to be completed/initiated during the first year include:

Task 1. Use a library-based approach to identify synthetic lectins that bind to prostate cancer associated glycans/glycoproteins (CAGs). Note that this aim will continue over the life of the grant to continuously identify more selective and useful SLs. (Months 1-36)

***Initiating PI*:**

Task 1 a): Synthesize bead based peptoid libraries that incorporate phenylboronic acid moieties. (Months 1-4)

Peptoid libraries were constructed using 9 amine building blocks (diversity = 9^5; 5.9 x 10^4 members) using the scheme depicted in Figure 1A. Briefly, bromoacetic acid was coupled using DIC to Tentagel $-NH_2$ beads already coated with our MRBB linker sequence. The beads were split and the 9 different amines were added to equal amounts of beads and reacted in DMF. The beads were then washed, re-pooled and treated with bromoacetic acid and DIC to couple the second diversity element. The Dde protecting group was selectively removed using hydrazine to uncover the primary amine to be conjugated to phenylboronic acid (PBA). PBA installation was verified using ARS and several beads were randomly selected for library quality evaluation.

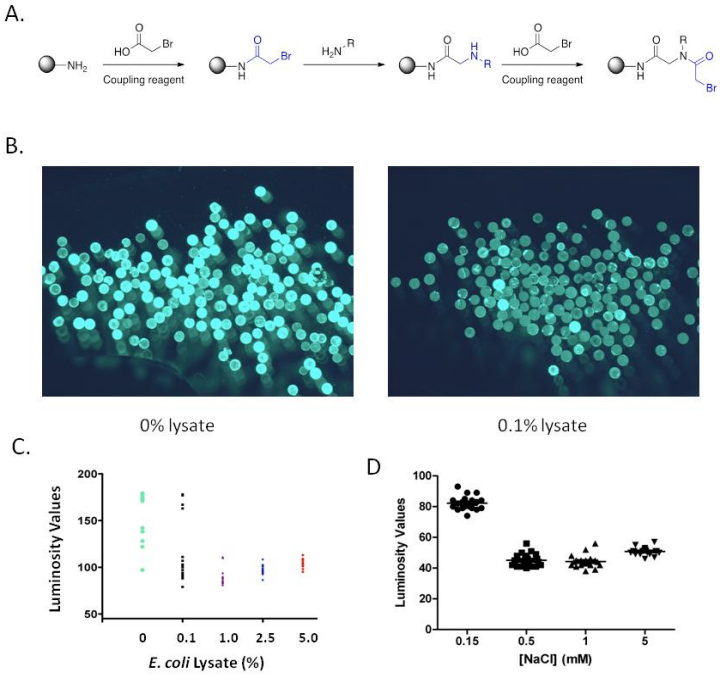


Figure 1. Optimization of screening conditions. **A**. Scheme for generating peptoid library. **B**. Images of phenylboronic acid (PBA) library with either 0% *E. coli* lysate (EL, left) or 0.1 % EL lysate (right). **C.** Bead quantification of increasing amounts (0 – 5 %) EL. **D.** Increasing amounts of NaCl decrease the binding of the library down to background.

With the synthesized libraries in hand, we turned our attention to identifying ideal screening conditions. Our goal was to identify stringent conditions so we could identify highly selective hits from our libraries. Based on previous studies,[1] we used *E. coli* lysates (EL) to both pre-block the beads and minimize non-specific interactions during analyte incubation. Figure 1B shows the drastic decrease in fluorescence when adding 0.1% EL to the screening buffer. Indeed, an EL gradient (Figure 1C) identified 0.1% EL as the optimal concentration since higher concentrations showed to strong of a decrease in fluorescence. We then optimized the salt concentrations (Figure 1D) and determined that 150 mM NaCl is ideal.

Task 1 b): Screen peptoid libraries with prostate cancer associated glycoproteins and complex glycans to identify highly selective and cross-reactive synthetic lectin (SL) hits. (Months 3-36)

To identify SLs that are specific for CAGs (Figure 2A), we designed a screening platform that used biotinylated complex carbohydrates conjugated to fluorescently labeled streptavidin (SA) (Figure 2B). Briefly, a series of biotinylated carbohydrates (i.e., sialyl Lewis X, sialyl Lewis A, Lewis X and Lewis A) were obtained from the Consortium of Functional Glycomics (CFG). Because of our previous success with

peptide library screening, we initially optimized our screening conditions using phenylboronic acid based peptide libraries instead of peptoid based ones incorporating either the phenylboronic acid or benzoboroxole moieties. For this assay, we preincubated the CFG glycans with FITC-streptavidin for 1 h in a 4:1 glycan-SA ratio then added this complex to our PBA-peptide library in screening buffer. Using this method, we identified 2 hits when screening with sLe^x as the target glycan. These hits were sequenced and had the following sequences: sLe^x1 = MRBB – LDRFRDL-Ac and sLe^x2 = MRBB – RDRWVDY-Ac. In addition to validating this screening modality for identifying both peptide and peptoid based libraries, further analyses demonstrate that these hits bind sialyl Lewis X better than Le^x or either of the Le^a derivatives (see below).

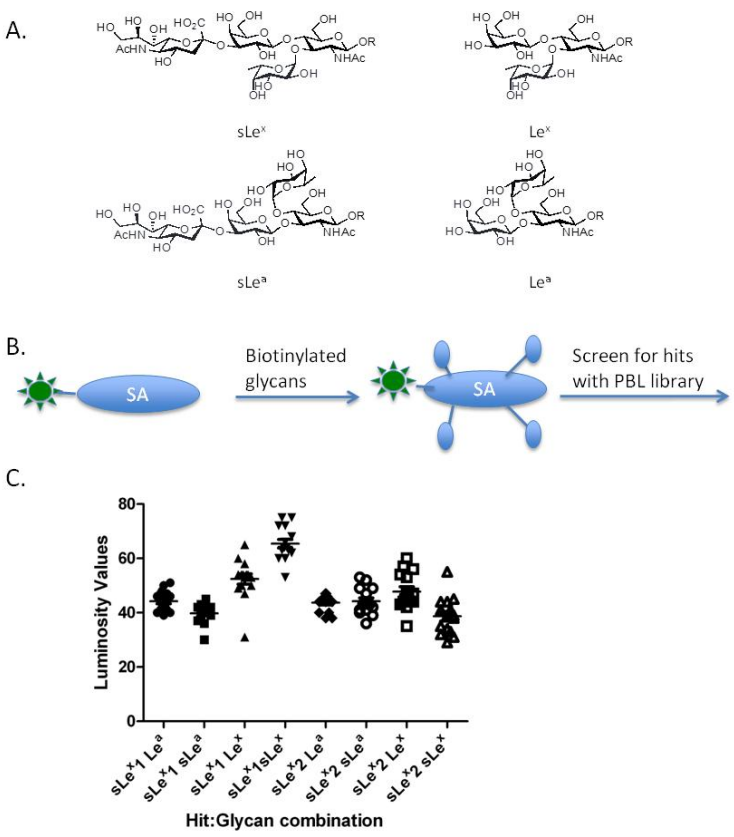


Figure 2. New screening targets. **A.** Structures of cancer associated glycans. **B.** Diagram of the glycan screening methods using FITC-Streptavidin (SA). **C.** Library screens of the different glycans from A. using the approach in B.

Task 1 c): Upon identifying ≥5 hits, we will sequence, resynthesize, and determine their selectivity of identified hits towards the target that they were selected against as well as the other prostate cancer associated glycoproteins and complex glycans. (Months 3-36)

We set out to validate our two PBA-peptide hits by first resynthesizing the two hits identified in (Task 1b), sLe^x1 and sLe^x2. We then screened these hits against Le^x, Le^a and sLe^a (Figure 2A) and determined that both of the hits bind sLe^x better than Le^x or either of the Le^a derivatives (Figure 2C). These results are encouraging and will be expanded as the number of hits increases after additional rounds of screening.

***Partnering PI*:**

Task 1 a): Synthesize bead based peptide libraries that incorporate phenylboronic acid moieties. (Months 1-4)

Two peptide-based fixed-position libraries were synthesized on Tentagel resin analogous to those previously described.[2] The effectiveness of the coupling was assessed using MALDI-MS in the past, here however, we ran into difficulties. From all of our efforts, our MS analysis consistently indicated incomplete deprotection of the iv-Dde protecting groups on the Dab side-chains (where boronic acids are attached). This appeared to be a significant portion of the product, composing up to 60%. Moreover, our MS analysis frequently suggested that we were getting incomplete coupling of the first Dab moiety. These were problems we had not encountered previously, yet appeared to be an issue when even re-synthesizing known SLs.

Consequently, we thoroughly evaluated the quality of the batches of TentaGel resin, hydrazine (used to deprotect the iv-Dde) and Fmoc-Dab(iv-Dde)-OH from the vendors. Note that we were using the same vendors as we had in the past. No apparent anomalies were detected in these reagents. Furthermore, upon a

detailed investigation of the literature, we identified much "controversy" and similar problems were indicated with respect to deprotecting the iv-Dde protecting group.

We thus opted to re-evaluate our synthetic approach and tried different side-chain amine protecting groups on Dab including alloc and MTT. From these studies, we determined that the deprotection of alloc was sensitive to water and oxygen, making it difficult to work with at times. Furthermore while the MTT group was easy to deprotect, amino acids with this group on the side-chain were often difficult to couple to the resin due to the size of the MTT group and increased steric interactions.

Interestingly, when we synthesized SL5 on a cleavable Rink Amide Resin using Fmoc-Dab(iv-Dde)-OH, we were able to confirm the presence of fully deprotected SL5 as the major product using MALDI-MS. Next, we more rigorously investigated the relative ratios of protected and deprotected SL5 from the TentaGel resin using LC-MS. Remarkably, using this method we observed only ~3% of the mono-and di-protected analogs combined. Still, by MALDI-MS we were seeing nearly 40% of the protected products from the same sample. After numerous control experiments, including investigating the ionization efficiencies for all of the possible products and using an Orbi-Trap MS-MS to confirm sequences, we were able to confirm the validity of the LC-MS analysis.

Ultimately, we accepted the fickle-nature of MALDI-MS and again felt confident in our synthetic protocols for library development. Conformation of the attachment of the boronic acids proceeded with less uncertainty, relying on a previously identified binding assay with alizarin red S (ARS). In the end, we were able to identify other orthogonal amine protecting groups (i.e. MTT on long side-chain amines) that will simplify syntheses related to studies on poly-valency as well as for incorporating other side-chain functionality such as biotin. Using the Orbi-Trap MS we were also able to obtain better sensitivity and enhanced sequencing efficiency as compared to MALDI-MS.

Task 1 b): Screen peptide libraries with prostate cancer associated glycoproteins and complex glycans to identify highly selective and cross-reactive synthetic lectin (SL) hits. (Months 1-36)

The screening methods previously used to identify SL1-SL5 were employed to screen portions of our library against prostate cancer associated glycoproteins. As we continue to improve these screening methods we will continue to improve the quality of the hits we identify. Initially, we screened the library with ovalbumin (OVA) and porcine stomach mucin (PSM) as these glycoproteins contain glycans of interest that have been associated with prostate cancer (PCa), namely mannose and N-acetyl glucosamine (GlcNAc) on OVA and GlcNAc and fucose on PSM. From these screens, four new SLs were isolated and sequenced (SL6-SL9 in Table 1).

Table 1. Sequences of identified SLs.

SL Hit	Sequence	Glycoprotein Screened	Glycoprotein Selectivity
SL1	Ac-RG**D*VTF**D***R-BBRM-resin	OVA	Cross reactive
SL2	Ac-RT**D***RFL**D***V-BBRM-resin	OVA	OVA
SL3	Ac-RS**D***VTT**D***R-BBRM-resin	OVA	OVA
SL4	Ac-RR**D***TQT**D***Q-BBRM-resin	PSM	OVA, PSM
SL5	Ac-RA**D***TRV**D***V-BBRM-resin	PSM	PSM
SL6	Ac-RT**D***NRN**D***F-BBRM-resin	PSM	OVA, BSM
SL7	Ac-RS**D***YFT**D***Q-BBRM-resin	PSM	OVA, PSM
SL8	Ac-RT**D***YGN**D***N-BBRM-resin	PSM	PSM
SL9	Ac-RT**D***YQV**D***A-BBRM-resin	PSM	OVA, PSM

Beyond simply identifying new SLs, we have learned a great deal about how we do our analysis, specifically in how we image our resin and extract color data. In all of our image acquisition and analysis we have always been conscientious of the quality of the image and how we extract luminosity data. Still, until recently all decisions had been made by the user, which can introduce user bias. Therefore, in order to limit introduction of external bias we wrote a bead finding and data extraction algorithm using MATLAB. Of particular interest to us was eliminating any inhomogeneity across the field of view, which could result from variation,

between users, in the illumination source settings, focus or hardware alignment. The simplest approach was to define a region of interest (ROI) that could be set and used to reduce any edge effects. From there we could simply have the software "find" the beads based on relative intensity changes. In addition, we created the option to reject any identified objects based on size (area or circumference), circularity and/or pixel saturation at any given percentile of the pixels for each bead. Remarkably, reprocessing existing images with this algorithm, using only the ROI and rejection based on size, improved classification accuracy, based on leave-one-out methods, from 97% to 99% for 5 cell lines.

Task 1 c): Upon identifying ≥5 hits, we will sequence, resynthesize, and determine the selectivity of identified hits towards the target that they were selected against as well as the other prostate cancer associated glycoproteins and complex glycans. (Months 3-36)

As described above, the four new hits listed in Table 1 were sequenced using MS-MS techniques and were resynthesized on TentaGel resin. To identify general selectivity trends, and for comparison with the original five SLs identified, each SL was bound with three glycoproteins (OVA, BSM, and PSM) as well as BSA, which was used as the control for nonspecific protein binding to the beads. Briefly, the library and the SLs were blocked with 1% BSA to minimize nonspecific binding, and then incubated with 0.1 mg/mL FITC-labeled analytes for 16 hours. After washing with PBS to remove unbound analyte, beads were imaged using a fluorescent microscope and color data extracted using the MATLAB algorithm described above. The library was used as a control, to reduce the differences between each glycoprotein in the extent of fluorescent labeling and degree of glycosylation. As such, the average raw intensity values for the library was subtracted from each replicate measure for each SL binding analyte. This normalized difference was then divided by the raw intensity of the library to afford a relative percent change for each SL binding each analyte. As shown in Figure 3, all of the SLs are cross-reactive to some degree. For example, while SL1 is considered completely cross-reactive, showing virtually no selectivity for any particular analyte, SL5 and SL6 display exquisite selectivity for PSM over BSM (~50-fold) and BSM over PSM (~60-fold), respectively, The remaining newly identified SLs show between 1.6 and 18-fold selectivity for one analyte over another.

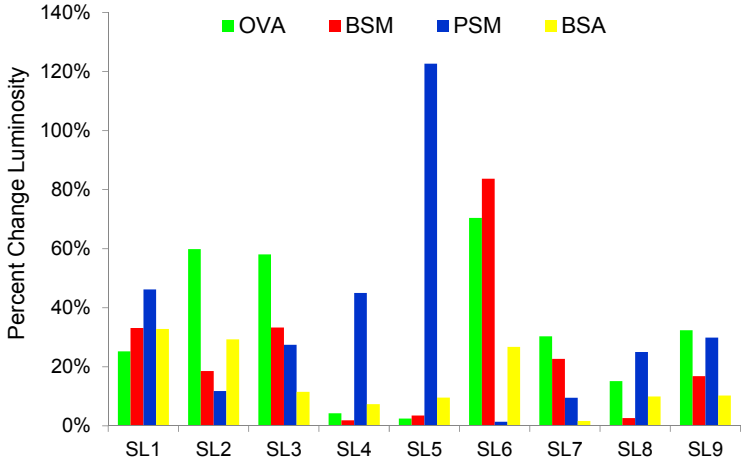


Figure 3. SL selectivity trends. Relative percent change in luminosity for SL1-SL9 binding ovalbumin (OVA), bovine submaxillary mucin (BSM), porcine stomach mucin (PSM) and bovine serum albumin (BSA).

Task 2. *Initiating PI*: Examine the biochemical/biophysical basis of the glycan•SL interaction. (Months 3-36)

Task 2 a): Upon identifying ≥5 hits (Task 1), we will develop a structure-activity relationship for highly selective SLs based on: 1) Alanine scanning 'mutagenesis'; 2) Varying the tether length; 3) Varying the boronic acid linkage and substitution patterns; and 4) Examining boronic acid substituent effects, to identify the factors that promote the selective recognition of a glycan by a particular SL. (Months 3-32)

While we have had previous success using 2-phenylboronic acid as our glycan targeting moiety, we also wanted to see if the recently described benzoboroxole would serve as a more suitable boronic acid. We first synthesized the carboxy-benzoboroxole (Figure 4A) and then coupled it to the same sidechain Dab amine on

SL5 as was used for the PBA derivative. Interestingly, benzoboroxole-SL5 showed increased affinity for PSM when compared to the original PBA derivative (Figure 4B). Due to the improved affinity, we built both a peptide library (diversity = 11^5; 1.6 x10^5 members) as well as a peptoid library (diversity = 9^5; 5.9 x10^4 members) incorporating the benzoboroxole moiety. While were able to successfully screen the peptide library and identify a hit ("Box1" - MRBB–VDARTDGR), sequencing the boroxole hits has been challenging due to the effect of the benzoboroxole moiety on ionization. As such, we are optimizing a variety of oxidations and cross couplings that we expect will efficiently remove the benzoboroxole functionality, and thereby facilitate the successful sequence of hits. Additional structure-activity relationships will be determined once we accumulate ≥5 hits.

Task 2 c): Feed information from the above studies back into the library design process to aid the generation and subsequent identification of highly selective SLs. (Months 9-32).

Based on our experience with the benzoboroxole, which improved the affinity of SL5 for PSM, we are focused on incorporating this moiety into libraries once we optimize library sequencing. The lessons learned from the Partnering PI's structure-activity-relationships are also being incorporated into the design process (see below).

Task 3. *Partnering PI*: Examine the biochemical/biophysical basis of the glycan•SL interaction and develop SL-based sensor arrays for the proposed prostate cancer diagnostic. (Months 1-36)

Task 3 a): Develop a structure-activity relationship for previously identified SLs (SL2 and SL5) based on: 1) Alanine scanning 'mutagenesis'; 2) Varying the tether length; 3) Varying the boronic acid linkage and substitution patterns; and 4) Examining boronic acid substituent effects to identify the factors that promote the selective recognition of a glycan by a particular SL. (Months 1-12)

In our analysis of how structure impacts binding affinity and selectivity of SLs for glycoproteins, we have identified some expected and some unexpected correlations. These studies primarily revolved around SL2 and SL5 because they represent opposite ends of the spectrum; in that SL2 displayed modest selectivity (~2-fold) with high background binding while SL5 exhibited high, nearly 50-fold selectivity, with low non-specific binding. In selecting these two SLs we wanted to learn more about what factors most significantly impact binding for highly selective and modestly selective SLs to better understand if the same factors are important for each in order to improve new SL development.

Using alanine scanning mutagenesis with SL2 for binding OVA (***Task 3 a-1***, Figure 5A) we see that charge on the peptide is important for binding affinity. Specifically, replacing R4 with alanine causes a 60% decrease in binding compared to native-SL2. Similarly, R1 and the arginine found in the C-terminal MRBB-sequence also reduce binding, though to a lesser extent (45% and 24% respectively). Likewise, binding affinity is reduced by more than 50% when the aminomethyl-phenyl boronic acids (D* = 3,7-Dab-PBA) are replaced with alanine or phenylalanine. However, when the Dab residues were left

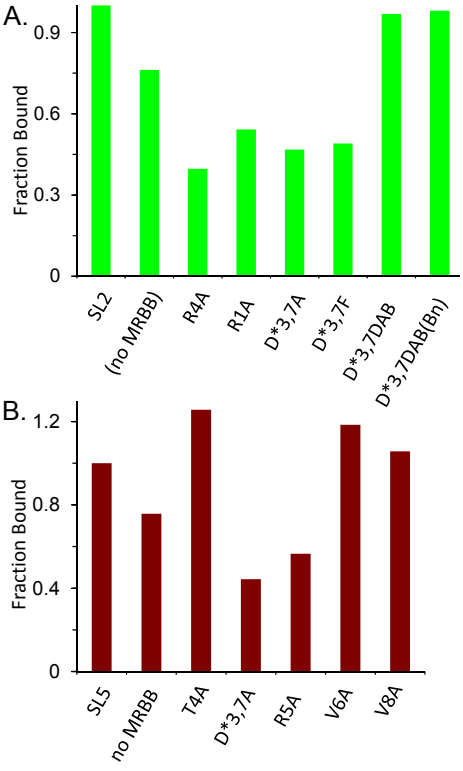


Figure 5. Alanine scanning mutagenesis. **A.** Relative binding for SL2-mutants with OVA. **B.** Relative binding for SL5-mutants with PSM.

unmodified or alkylated with benzaldehyde, thereby leaving the charged ammonium at neutral pH, binding affinity was only diminished 2-3%. Similar trends were observed in SL5 for binding with PSM (Figure 5B). For example, when R5 was replaced by alanine binding was decreased nearly 55% and replacing both D* with alanine resulted in a 65% binding decrease. Interestingly, when T4 was replaced by alanine PSM binding was enhanced 25%. Similarly, when V6 or V8 was replaced with alanine a 20% and 5% increase in PSM binding was observed, respectively.

The role that the boronic acids play in defining SL binding affinity and selectivity was also studied (***Task 3 a-3***). In general, there was no observed loss of affinity when regio-isomeric phenyl boronic acids (PBAs) were used in SL2 and/or SL5, yet the PBA is undoubtedly important for defining selectivity (Figure 6). As seen in Figure 6A, there is no appreciable change in the selectivity patterns whether the boronic acid is *ortho*-, *meta*- or *para*- to the linkage to the peptide. This observation was unexpected 1) because of expected conformational preferences for sugar binding based on positioning the boronic acid in a specific orientation to bind the sugar, and 2) because when the boronic acid is *ortho*- to the amino-methyl group enhanced diol binding is expected due to conformational and Lewis acidity trends. When the more sterically crowded and conformationally restricted 2-Ac-PBA is incorporated into SL2 the binding preference for OVA actually increases, though modestly (from 3-fold to ~6-fold). Most notably, however, is that when the PBA is replaced with a simple benzyl-group all selectivity is lost. SL5 showed similar trends (Figure 6B); with the orientation of the boronic acid having no significant influence on glycoprotein binding. Interestingly, in contrast to what was observed for SL2, the binding selectivity for SL5 decreased when the bulky 2-Ac-PBA was used. The final boronic acid modification, adding electron-donating (-OCH_3, -NR_2) and electron-withdrawing (-CF_3, -NO_2, -CN) substituents onto the PBA to alter the Lewis acidity of the boronic acid (***Task 3 a-4***), unquestionably showed no impact on analyte binding.

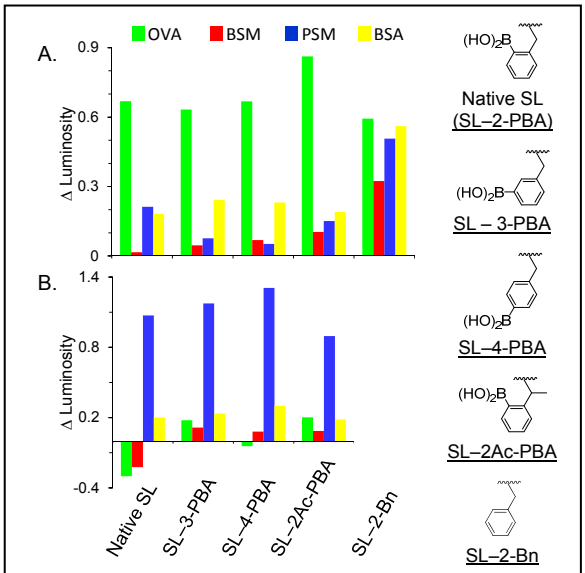


Figure 6. Impact on binding selectivity between SL and different test glycoproteins (OVS, BSM, PSM, BSA) when different regio-isomeric boronic acids are used for SL2 (**A.**) and SL5 (**B.**).

The length of the side-chain connecting the PBA to the peptide (i.e., the tether length, ***Task 3 a-2***) was also investigated. For this analysis, Dab and Lys were incorporated as the amino acid to which the boronic acid was attached in order to probe how degrees of freedom and thus preorganization can impact binding selectivity. Figure 7A and B show representative fluorescence images of portions of two libraries, derived independently from attachment of PBA to either DAB or LYS, after incubation with FITC-OVA. The Dab-based library displays decreased non-selective binding, as indicated by the decreased background fluorescence and increased library differentiation. Figure 7C is a binning chart, in which individual bead luminosities are plotted for each library. The greater spread in the data obtained for the Dab-containing library, versus the otherwise identical LYS-containing library, is an indication of greater differentiation and selectivity for binding the targeted glycoprotein.

As a final investigation of how structure can impact binding between SL and glycan, we looked at what impact the fluorescent label could have. SL1-SL5 are cationic, each containing a minimum of three arginine residues, and fluorescein is anionic at physiological pH. Based on what we learned about how charge impacts affinity in our alanine scanning mutagenesis studies, we wanted to determine how the dye charge was impacting binding affinity. We therefore labeled each of our glycoproteins with coumarin (as a neutral alternative) and rhodamine (as a cationic alternative) separately. If the charge on the dye significantly impacts the affinity of the SL for any given glycoprotein, we should see a decrease in the binding response as we move from fluorescein to coumarin, which is in fact what we observe (Figure 8). Still, rhodamine labeled glycoproteins would be expected to have a further reduced binding affinity due to the cationic dye, which is contrary to our results. We conclude from this that our microscope filter set is somehow inappropriate for the coumarin dye we are using, even though the wavelengths described seem relevant. Regardless, we are much more confident that labeling our targets is an appropriate method for identifying hits diagnostic.

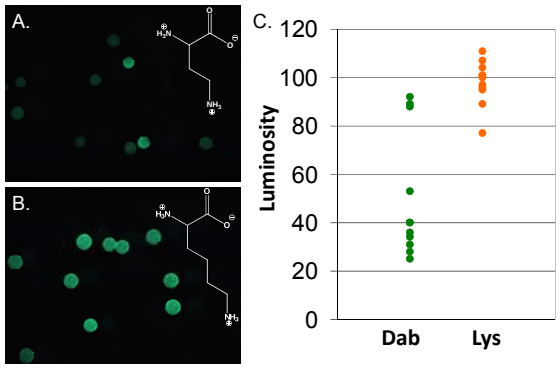


Figure 7. Representative fluorescence images of libraries incorporating Dab (**A.**) or Lys (**B.**) upon binding FITC-OVA. **C.** A binning chart of individual bead luminosities showing decreased background fluorescence and increased differentiation when Dab is used.

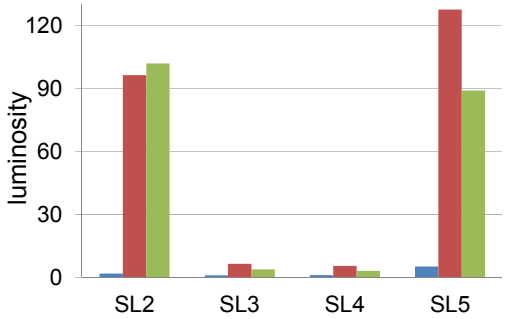


Figure 8. Relative binding response for SL2-SL5 binding with PSM, labeled with coumarin (blue), fluorescein (green) or rhodamine (red).

Task 3 b): Upon identifying ≥5 selective and cross-reactive SLs (Task 1), we will assemble them, and others identified in Task 2, into an array-based diagnostic format. (Months 1-36)

Task 3 c): Evaluate the ability of the array to discriminate complex glycans (i.e., TF antigen, Le^a, Le^x, sLe^a, sLe^x). Note that because the development of the arrays will be continually evolving, as we identify new and more selective SLs. Thus, the time frame for this task is the entire proposal period. (Months 1-36)

Task 3 d): Evaluate the ability of the array to discriminate prostate cancer cell lines (i.e. PC-3, LNCaP, and DU145), as well as RWPE-1, WPE1-NA22, WPE1-NB14, WPE1-NB11, and WPE1-NB26, which are referred to as the MNU cell lines, all available from the ATCC. Note that because the development of the arrays will be continually evolving, as we identify new and more selective SLs, the time frame for this task is the entire proposal period. (Months 1-36)

The vast majority of our work to date in developing and working with arrays has focused on how we analyze our array data. As described above, we have improved our data collection methods to obtain better consistency between replicate measurements as well as optimizing how intensity values are extracted.

In this regard, we have begun to evaluate our array response using color space intensities and not just luminosity. In particular we have focused on the popular "Red-Green-Blue" (RGB) color space to obtain more of a full spectral response from our array. In so doing we have improved our classification accuracy from 97% to 100% for a five cell line panel (including: HT-29, CT-26, CT-26-F1, CT-26-FL3, and 3T3/NIH) made up of 114 replicates we often use to evaluate our models.

To further validate our approach we have assessed the ability of our array to identify analytes which it has never seen before. Specifically, we used ten cell lines including a mix of mouse and human lines as well as colon (7 - 3T3, HT29, HCT116, CT26, CT26-F1, CT26-FL3, and Lovo), breast (2 - MCF7 and MCF10A) and prostate (1 - PC3) cell lines. To do this we create a statistical model based on 9 cell lines while leaving data from one cell line out and then attempt to classify this excluded line, in much the same way that a diagnostic test must determine the disease status for a patient that did not contribute to the calibration data set. As such, when classifying our samples as healthy, cancerous/non-metastatic or cancerous/metastatic we only obtained 56% overall classification accuracy (Figure 9, blue). However, if we simply look to diagnose the cancer and not stage it at the same time, thereby identifying our data as either healthy or cancerous, we improve our classification accuracy to just over 83% (Figure 9, green). Still, by ignoring the 3T3/NIH mouse fibroblast line, the most out of place cell line in this analysis, and looking at the remaining nine cell lines using this same approach, we can "diagnose" the presence of cancer 100% of the time, with a sample set of n = 434.

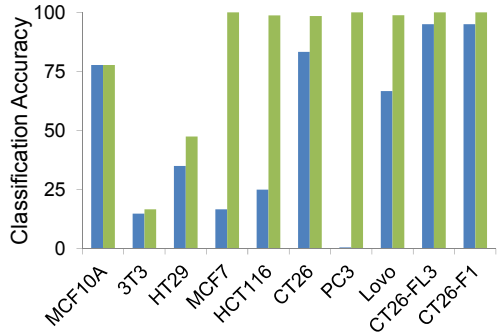


Figure 9. Individual classification accuracies (3-class groupings in blue, 2-class groupings in green) for each of 10 cell lines derived from a model lacking all input from that line.

Finally, we realize that using linear discriminant analysis (LDA) is not necessarily the best approach for analyzing our data. We also recognize that not all samples can be controlled as tightly as ours have been previously. As such we evaluated our complete data set derived from colon cancer cell lines, including variations in incubation time (1 h to 24 h), incubation temperature (4 ^{o}C, 25 ^{o}C and 37 ^{o}C), and sample dilution (20x, 50x and 100x). In total this afforded nearly 12,000 measurements. Using support vector machines we were able to obtain 93% classification accuracy and using regression tree analysis we improved the classification accuracy to 97%. Working closely with Prof. Edsel Pena in the Department of Statistics at the University of South Carolina we are continuing to explore our options, being cautious that the approach we take is appropriate for the type of analysis we are doing as well as verifying that we do not "over-train" our models and that we maintain statistical validity

Key Research Accomplishments

- **Synthesized peptoid libraries (PRT).** Peptoid based SL libraries (diversity = 9^5; 5.9 x10^4 members) were synthesized on Tentagel macro beads and their utility for identifying SL's targeting proof-of-concept glycoproteins assessed. The library was also used to further optimize our screening procedures. Screening with this library to identify selective SL's is ongoing. We are also moving toward the synthesis of β-amino acid containing libraries, which are intrinsically structured/pre-organized, we expect to further aid the identification of SL's with improved selectivity.
- **Synthesized peptide libraries (JJL and PRT).** Peptide based SL libraries (diversity = 11^5; 1.6 x10^5 members) were synthesized on Tentagel macro beads and also used to further optimize our screening procedures and identify several new selective SLs (see below).
- **Optimization of screening protocols (PRT).** The above libraries were used to identify optimized conditions for identifying SLs that selectively bind our proof-of-concept glycoproteins and CAGs. These conditions are: 10 mM HEPES, 150 mM NaCl, 0.1% E. coli lysate (stock conc. 8 mg/mL) and 0.05% TWEEN.
- **Developed a structure activity relationship (JJL).** Used SL2 and SL5 to develop a structure activity relationship. The key findings were that positive charge and the boronic acid are critical for affinity and selectivity. This information is being fed back into the library design process to aid in the generation and subsequent identification of highly selective SL's (see Tasks 2c and 3a).
- **Identified boroxole as a high affinity sugar binding motif (PRT).** The 2-formylphenyl boronic acid moiety was replaced with several different boronic acids to explore boronic acid substituent effects, and thereby identify the factors that promote the selective recognition of a glycan by a particular SL. The key findings were that the substitution pattern did not matter and that substituent effects (e.g. electron donating/withdrawing group) were minimal. Also, the boroxole moiety was identified as an alternative moiety with improved affinity.
- **Optimization of image capture and analysis (JJL).** A Matlab algorithm was successfully developed to automate data extraction from microscope images of our bead-based assays. The algorithm not only identifies each bead and extracts color space intensity values, but also allows for data rejection based on customizable threshold values for size, circularity and/or color space percentile high values (i.e., relating pixel saturation). Using this automated data collection system, additional statistical analyses have been performed on our colon cancer data sets, and using quadratic discriminant analysis and/or support vector machines, our classification accuracies improved from 97% to >99%.
- **Identified 4 additional SLs that bind proof-of-concept glycoproteins (JJL and PRT).** Screens of peptide libraries containing either 2-formylphenyl boronic acid or boroxole identified 4 additional SLs that bind proof-of-concept glycoproteins.
- **Identified SLs that selectively bind sialyl Lewis X over Lewis X, sialyl Lewis A, and Lewis A (PRT).** Screens of peptide libraries versus biotinylated-sialyl Lewis X identified two SLs (SLex1 and SLex2). Confirmation assays demonstrated that SLex2 selectively binds sialyl Lewis X over Lewis X, sialyl Lewis A, and Lewis A.
- **Used existing SL array to demonstrate the utility in diagnosing and staging prostate, breast, and colon cancer (JJL).** Using our SL array to classify various colon cancer cell lines according to metastatic potential, we achieved 97% classification accuracy as reported in our *Chem Sci* manuscript. Inclusion of additional colon, breast and prostate cancer cell lines (n = 10; 426 separate measurements), and grouping the different cell lines according to whether they are healthy, cancerous and cancerous/metastatic we achieve 84% classification accuracy. However, if we look at it from a diagnostic perspective, i.e. cancerous versus non-cancerous, the classification accuracy improves to 95%.

Reportable Outcomes

- Published a manuscript in *Chemical Sciences*[3] (see Appendices) detailing the utility of SL arrays to discriminate cancer cell lines based on metastatic potential, thereby setting the stage for further developing this approach for the diagnosis and staging of cancer.
- Kevin Bicker, who played a key role in developing the SL array, will begin his tenure track faculty position at Middle Tennessee State University in August 2013.
- Lavigne presented a seminar to the College of Pharmacy at the Medical University of South Carolina.
- Held joint lab meeting at The Scripps Research Institute, Scripps Florida, on July 25, 2013. Anna Veldkamp, Kathleen O'Connell, and Daniel Lewallen presented seminars on their SL studies.

Conclusions

Significant progress has been made in the first year of funding on this project to develop synthetic lectin (SL) arrays that bind to prostate cancer associated glycans and glycoproteins (CAGs) to detect glycosylation patterns associated with cancer. These studies are being pursued to develop this methodology into a robust system, thereby providing a new paradigm that can diagnose and stage prostate cancer. Moreover, these studies directly relate to the "Imaging," and "Biomarker" focus areas of the PCRP overarching challenges. In particular, the progress made towards creating a cross-reactive sensor platform will allow for more reliable diagnosis of prostate cancer and thus improve the likelihood of accurate detection and aid in managing prostate cancer, thereby decreasing many of the negative impacts associated with prostate cancer.

Thus far, peptide and peptoid libraries have been synthesized and screened against cancer associated analytes. Consequently, six new synthetic lectins have been identified targeting both glycans (2 new SLs) and glycoproteins (4 new SLs). In so doing, we have been able to improve our methods for binding SLs to CAGs to reduce background binding, thereby improving our signal to noise ratio. We have also been able to advance our approaches to 1) acquire assay images, 2) extract assay response values and 3) analyze the assay outcome. Ultimately, these improvements have allowed us to verify the validity of our approach while also improving the overall assay accuracy. As such, we have enlarged our data set to nearly 12,000 measurements while expanding the assay relevance and at the same time maintaining classification accuracies between 93-97%. These results reflect assay responses to a combination of prostate, colon and breast cancer cell lines.

In addition to enhancing the overall assay performance, we have also advanced our understanding of what factors are important for SLs to bind CAGs. Specifically, we have demonstrated that boroxoles are efficient replacements for the originally proposed boronic acids and can improve the binding affinity of SLs for certain CAGs. We have also begun to develop a detailed structure-activity-relationship that has to date indicated that charge on the SL is important for defining binding affinity with CAGs while the boronic acids significantly contribute to binding selectivity.

As this project progresses, we will continue to expand our understanding of the factors important for SLs to bind CAGs. Specifically, we will synthesize sequence and positional mutants of other SLs to better define the role of each residue in binding CAGS and therefore to be able to draw more detailed broad conclusions. At the same time we will continue to evaluate the benefits of using peptoids and boroxoles in our array development. Significantly, we are continually screening our libraries for new hits that better target prostate cancer and subsequently these hits are included into our array and used to better discriminate prostate cancer cell lines while simultaneously improving our signaling strategies, our data analysis and the overall utility of our approach.

Despite being located at two different sites, PRT at TSRI and JJL at USC, the project has continued to grow and evolve through constant email and phone contact, as well as organized weekly meetings and scheduled site visits. As revealed above, each PI has contributed to different aspects of this project; with both PIs having overlapping and supporting roles for the other. Clearly, this team works well together, providing their own expertise to result in a level of productivity that is greater than that achievable by each PI working independently. Certainly, this project would not exist without the input of both PIs.

References

(1) Liu, X., Dix, M., Speers, A. E., Bachovchin, D. A., Zuhl, A. M., Cravatt, B. F., and Kodadek, T. J. (2012) Rapid development of a potent photo-triggered inhibitor of the serine hydrolase RBBP9, *Chembiochem 13*, 2082-2093.

(2) Bicker, K. L., Sun, J., Lavigne, J. J., and Thompson, P. R. (2011) Boronic acid functionalized peptidyl synthetic lectins: combinatorial library design, peptide sequencing, and selective glycoprotein recognition, *ACS Comb Sci 13*, 232-243.

(3) Bicker, K. L., Sun, J., Harrell, M., Zhang, Y., Pena, M. M., Thompson, P. R., and Lavigne, J. J. (2012) Synthetic lectin arrays for the detection and discrimination of cancer associated glycans and cell lines, *Chemical Science 3*, 1147-1156.

Appendix A

Bicker, K. L.; Sun, J.; Harrell, M.; Zhang, Y.; Pena, M. M.; Thompson, P. R.; Lavigne, J. J. Synthetic Lectin Arrays for the Detection and Discrimination of Cancer Associated Glycans and Cell Lines. *Chem. Sci.* **2012**, *3*, 1147-1156. DOI: 10.1039/C2SC00790H.

Chemical Science

Dynamic Article Links

Cite this: *Chem. Sci.*, 2012, **3**, 1147

www.rsc.org/chemicalscience

EDGE ARTICLE

Synthetic lectin arrays for the detection and discrimination of cancer associated glycans and cell lines†

Kevin L. Bicker,[ab] Jing Sun,[a] Morgan Harrell,[a] Yu Zhang,[c] Maria M. Pena,[c] Paul R. Thompson*[b] and John J. Lavigne*[a]

Received 12th October 2011, Accepted 3rd January 2012
DOI: 10.1039/c2sc00790h

Aberrant glycosylation is a hallmark of various disease states, including cancer, and effective detection and discrimination between healthy and diseased cells is an important challenge for the diagnosis and treatment of many diseases. Here, we describe the use of boronic acid functionalized synthetic lectins (SLs) in an array format for the differentiation of structurally similar cancer associated glycans and cancer cell lines; discrimination is based on subtle variations in glycosylation patterns. We further demonstrate the utility of our SLs in recognizing glycoproteins with up to 50-fold selectivity, even in 95% human serum. Given their robust and selective nature, these SLs were able to effectively distinguish (a) five structurally similar glycans with 94% accuracy; (b) seven normal, cancerous and metastatic colon cancer cell lines, including three isogenic cell lines, with 92% accuracy; and (c) these same seven cell lines using a guided statistical analysis to improve our analysis to 97% accuracy. In total, these data suggest that an SL-based array will be useful for the diagnosis of cancer.

Introduction

The intracellular and extracellular biomarkers displayed by healthy and diseased cells provide unique signatures by which these cells can be distinguished. For example, in healthy cells, post-translational glycosylation of proteins plays a critical role in cell–cell interactions and in cell signaling.[1] However, aberrant protein glycosylation is a hallmark of numerous diseases including inflammation and cancer, thus providing a means for the detection and classification of healthy and diseased states. Related to cancer, distinguishing between healthy and cancer cells that possess either low or high metastatic potentials typically relies on detecting subtle variations in the types and levels of specific biomarkers (*e.g.*, DNA, RNA, and proteins) using high-affinity, target-selective sensors, *e.g.* antibodies. Regardless of the analyte, these approaches all require prior knowledge of the markers targeted and no specific biomarker or combination of biomarkers has been identified to sufficiently differentiate between healthy, cancerous/non-metastatic and cancerous/metastatic cell types. An alternative to this "lock-and-key" approach[2–6] would be to use cross-reactive recognition elements as part of a sensor array.

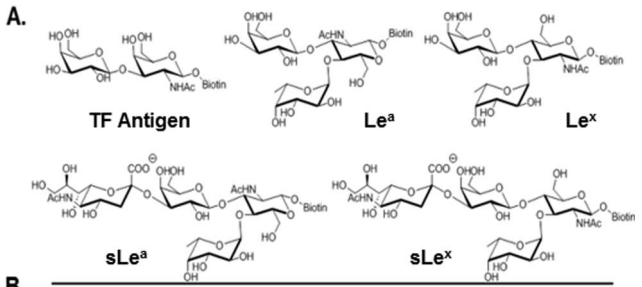
Cross-reactive sensor arrays incorporate multiple receptors with different affinities such that each component has a selective and unique interaction with the targeted analyte(s). As a result, the response from the entire array produces a fingerprint pattern characteristic of the analyte to which it is responding. That is to say that classification is not based on the response from a single receptor, but rather it is the composite response from the entire array that allows for identification and classification of the analyte. This practice has often been referred to as the "electronic nose" approach,[7–13] though, in this case, used for solution-based analysis.

While natural lectins (sugar binding proteins) display cross-reactivity, and lectin arrays can often offer an effective approach to cancer diagnostics, the methodology is often complex and the constituents are of inherently low stability and high cost.[14–17] Here we describe an alternate approach based on the covalent yet reversible binding between boronic acid functionalized synthetic lectins (SLs) and cancer associated glycans and glycoproteins. This design does not require previous knowledge of the biomarkers targeted; rather it is focused on identifying changes in glycosylation patterns, a factor that is known to play a significant role in oncogenesis and metastasis.

In cancerous cells, the expression of specific glycan structures can be increased, decreased, or even newly expressed. These changes often co-opt cellular signaling pathways to promote growth, division and metastasis.[1] For example, sialyl Lewis X (sLe^x) and sialyl Lewis A (sLe^a) (Fig. 1A) are overexpressed in

[a]*Department of Chemistry & Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC, USA 29208. E-mail: JLavigne@chem.sc.edu; Fax: +(803)-777-9521; Tel: +(803)-777-5264*
[b]*Department of Chemistry, The Scripps Research Institute, Scripps Florida, 120 Scripps Way, Jupiter, Florida, USA 33458. E-mail: PThompso@scripps.edu; Fax: +(561)-228-3050; Tel: +(561)-228-2860*
[c]*Department of Biological Sciences, University of South Carolina, 715 Sumter Street, Columbia, SC USA 29208*

† Electronic supplementary information (ESI) available: Complete methods including: labeling, membrane extraction and screening protocols, Supplementary Figures S1–S5 and LDA classification data. See DOI: 10.1039/c2sc00790h



B.

SL Hit	Sequence
SL1	Ac-RG**D*VTF**D***R-BBRM-resin
SL2	Ac-RT**D*RFL**D***V-BBRM-resin
SL3	Ac-RS**D*VTT**D***R-BBRM-resin
SL4	Ac-RR**D*TQT**D***Q-BBRM-resin
SL5	Ac-RA**D*TRV**D***V-BBRM-resin

Fig. 1 (A) The structures of biotinylated cancer associated glycans used in this study. (B) The sequences of the SLs used for validation studies and in the array assessments.

breast, colon and pancreatic cancers,[1] and the increased expression of sLe^x is known to enhance tumor metastasis.[18–21] Tests to detect specific aberrant glycosylation events are used for both initial disease diagnosis and monitoring disease progression yet suffer from limitations including a high number of false positives and a reliance on inherently unstable and costly antibodies or natural lectins.[14–17] For example, elevated levels of CEA (carcinoembryonic antigen), an aberrantly glycosylated glycoprotein, are associated with an increased risk of colon cancer relapse and metastasis.[15] However, the test for CEA is only effective in 4% and 25% of Stage I and II cancers, respectively, which is problematic for a cancer diagnostic because it is during these early stages when the disease is most effectively treated.[22]

The development and use of boronic acid functionalized synthetic lectins (SLs) for saccharide detection and cancer diagnosis is a rapidly growing field.[23–36] Boronic acids are incorporated into the SLs to enhance glycan binding *via* their ability to form covalent yet reversible bonds to the 1,2- and 1,3-diols present on many saccharides. These small molecule SLs generally show enhanced stability compared to antibodies and natural lectins, and it has been shown that incorporation of synthetic lectins into an array format allowed for the recognition and discrimination between simple monosaccharides and oligosaccharides in neutral aqueous media as well as real-world beverage samples, *i.e.* sweet tea with added Splenda.[37] Further advances using cross-reactive nanoparticle-conjugated polymer based arrays have shown utility in differentiating normal, cancerous and metastatic cell types.[38]

We previously described the design, synthesis and utility of boronic acid functionalized peptide-based SLs in binding to glycoproteins[36] and highlighted efforts in library design optimization and peptide sequencing.[35] SLs, that were both cross-reactive and up to 5-fold selective for a particular glycoprotein, were identified.

Herein, we report the identification and characterization of three additional SLs that bind to proof-of-concept glycoproteins with up to 50-fold selectivity, even in complex matrices (*i.e.*, human serum). Additionally, a four-component SL array was used to detect and differentiate five structurally similar cancer associated glycans (Fig. 1), as well as one 'healthy' and six cancer cell lines with high classification accuracy. By combining selective and cross-reactive SLs within the array, the selectivity of an individual SL need not be high as each sensor need only be incrementally different to create an array that maximizes variation in the array response to different analytes.[39,40] Further analyses using directed partitioning, based on similarities in metastatic potential, was used to enhance the classification accuracy. Our results demonstrate the utility of using SL arrays for the diagnosis of cancer. Furthermore, since the analyte for which each SL was selected is not found on any of the cancer-associated cells studied, our array displays inherent adaptability.[39,40] That is to say that this relatively small array was able to "learn" and accurately classify never before seen analytes.[39,40]

Results and discussion

Employing the same approach used to identify SL1 and SL2,[35,36] SL3, SL4 and SL5 (Fig. 1B) were identified by screening our bead-based fixed position library with fluorescein isothiocyanate (FITC)-tagged versions of ovalbumin (OVA) and porcine stomach mucin (PSM). These SLs were subsequently re-synthesized and their selectivity and cross-reactivity evaluated using OVA, PSM, BSM (bovine submaxillary mucin) and BSA (bovine serum albumin). OVA, PSM and BSM are all glycoproteins, and it is noteworthy that the two mucins contain the same type of glycans but to differing extents and displayed in different environments. BSA, which is not glycosylated, was used as a control for non-specific protein binding.

SL selectivity studies

To control for differences in the extent of labeling or glycosylation, the fluorescence intensity of a similarly sized set of the SL library was used as a reference. The fluorescence intensity of the library was subtracted from the fluorescence intensity of the re-synthesized SL incubated with the same FITC-tagged glycoprotein (Fig. S1, ESI†), providing a change in fluorescence intensity upon binding. A percent change in binding was obtained by dividing this difference by the fluorescence intensity

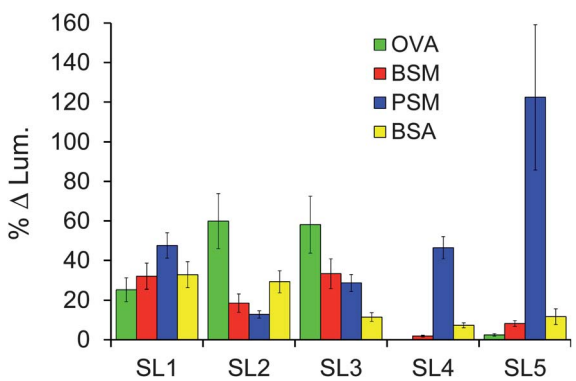


Fig. 2 Percent change in luminosity of each identified SL towards four different analytes (analyte identification indicated in the legends above). Error bars represent the standard error of the percent change relative to the control as this propagated uncertainty is based on the variance between replicate measurements for the sample and control reference.

of the library (Fig. 2). To compare the ability of each SL to differentially bind to glycoproteins, a selectivity factor was obtained by dividing the percent increase for each analyte by the percent increase of the weakest binder for that SL (Table 1). The library was chosen as the reference because it provides a control containing all of the potential cross-reactive elements that could interfere with our assessment of binding selectivity. Outliers from the control were removed using the studentized t-test at the second quartile to give an accurate average for standardization purposes.

The data for SL1 and SL2 have been previously described[35] and are included in Fig. 2 and Table 1 for comparison. Here we see that SL1 is completely cross-reactive, binding with no more than 2-fold selectivity for any one analyte. In contrast, SL2 shows modest selectivity for binding OVA. The 3- and nearly 5-fold selectivity SL2 shows over BSM and PSM, respectively, demonstrated the ability of this approach to distinguish between similar analytes. However the 2-fold selectivity of SL2 for OVA over BSA suggests high non-specific, background binding for this SL, thereby decreasing its potential utility in a diagnostic array.

The newly reported SL3 was selected from screening the library against OVA, and showed only 2-fold selectivity towards OVA over BSM and PSM, while exhibiting relatively low background binding, as indicated by the 5-fold selectivity over BSA. SL4 and SL5 were identified from screening the library for PSM binders. Although SL4 displays an impressive 25-fold selectivity for PSM over BSM, it exhibits only ~6-fold selectivity for PSM over BSA. Thus, while exhibiting some degree of selectivity and showing a particular preference for binding certain analytes (*i.e.*, PSM *vs.* BSM), this SL can also be considered cross-reactive with respect to PSM *vs.* BSA. As such, this SL is an ideal candidate for inclusion in a sensor array because it possesses differential analyte binding. Note that SL4 shows virtually no affinity for OVA and as such the percent change in luminosity relative to the library control is very small (0.15%). Thus, for the discussion of selectivity, presented in Table 1, BSM was used as the weakest binder because it was not reasonable to use OVA and divide by such a small number (*e.g.* PSM selectivity *vs.* OVA is 250).

Similar to SL4, SL5 displayed exquisite selectivity, exhibiting 50-fold selectivity for PSM over OVA and ~15-fold selectivity over BSM. The excellent selectivity of SL4 and SL5 for PSM over BSM (~25- and ~15-fold selectivity, respectively) is particularly impressive because these two glycoproteins possess identical types of glycans, though to a different extent and differentially displayed.[41–43] These results suggest that these SLs not only bind to the saccharide, but also the protein. Nevertheless, it is important to recognize that we have previously shown that glycans are significant for the SL–glycoprotein interaction.[36]

Table 1 Selectivity factors for each SL screened against four different glycoproteins[a]

	OVA	BSM	PSM	BSA
SL1	1.0	1.3	1.9	1.3
SL2	4.7	1.4	1.0	2.3
SL3	5.1	2.9	2.5	1.0
SL4	0.1[b]	1.0	24.8	3.9
SL5	1.0	3.4	49.9	4.8

[a] The fold selectivity of an SL for one glycoprotein over another can be obtained by dividing their respective selectivity factors. SL1 and SL2 data from Bicker *et al.*[35] [b] The fold selectivity for SL4 was determined using BSM as the reference. OVA was not used as the reference because the %Δ luminosity was practically zero and dividing by such a small number resulted in fold selectivities that were quite meaningless.

The robustness of the SL-glycoprotein interaction was assessed using SL2 and SL5 with differing percentages of human serum (0, 25, 50 and 95%) in screening buffer. Both SLs retained excellent selectivity for the respective glycoproteins in all concentrations of serum (Fig. S2, ESI†). Control experiments confirmed that no serum components caused any changes in the assay response (Fig. S3, ESI†). To examine the contribution of valency, dissociation constants (K_d) were determined for both the bead-based polyvalent SL5 and a monovalent SL5. The dynamic nature of the beads[44] (*i.e.*, being a gel resin) allows for multiple interactions between bead-based SLs and the many glycans expressed on PSM. Therefore, incubating polyvalent, bead-based SL5 with varying concentrations of fluorescently labeled PSM (having a polyvalent display of glycans) yielded a K_d of 2.5 ± 0.29 μM (Fig. S4, ESI†).[45] A fluorescence polarization (FP) assay was used to measure the affinity of the fluorescently-labeled, monovalent SL5 (FITC-SL5) for PSM.[46] However, saturation of the FP signal was not observed because of limited glycoprotein solubility (Fig. S5, ESI†), thus K_d values could not be determined. Nevertheless, the observed response validated the assay and suggested that the K_d for the monovalent SL5-PSM interaction is significantly higher than 10 μM, the highest concentration tested. These results indicate that the polyvalent nature of the beads is critical for high affinity binding and suggest that multiple SLs on a single bead interact with each glycoprotein.

Glycan competition studies

Glycan competition assays were used to identify the glycan structure(s) that were responsible for SL2-OVA and SL5-PSM binding. For these studies, SL2 was selected over SL3 because of the higher selectivity shown for OVA over BSM and PSM, while SL5 was chosen over SL4 because of the larger signal intensity response. In this study, varying concentrations of different monosaccharides were independently incubated with equal portions of resin-bound SLs and a constant concentration of the FITC-glycoprotein (0.1 mg mL^{-1}) that the SL preferentially binds. The glycans used in the study of SL2 were those found on OVA, namely galactose, mannose and *N*-acetylglucosamine (GlcNAc).[36,47] For SL5, galactose, GlcNac, sialic acid, fucose and *N*-acetylgalactosamine (GalNAc),[36,48] which are all found on PSM, were used. Fructose was used to probe non-specific saccharide binding between the SLs and glycoproteins because it is one of the strongest known 1 : 1 boronic acid binders.[49,50] It was expected that effective competition between a monosaccharide and a FITC-glycoprotein, for binding to the resin-bound SL, would result in a decrease in luminosity. Such a decrease in the binding signal would suggest that a particular monosaccharide was important for glycoprotein binding to the SL. Note that the response values in Fig. 3 have been mathematically defined such that increasing bar height corresponds

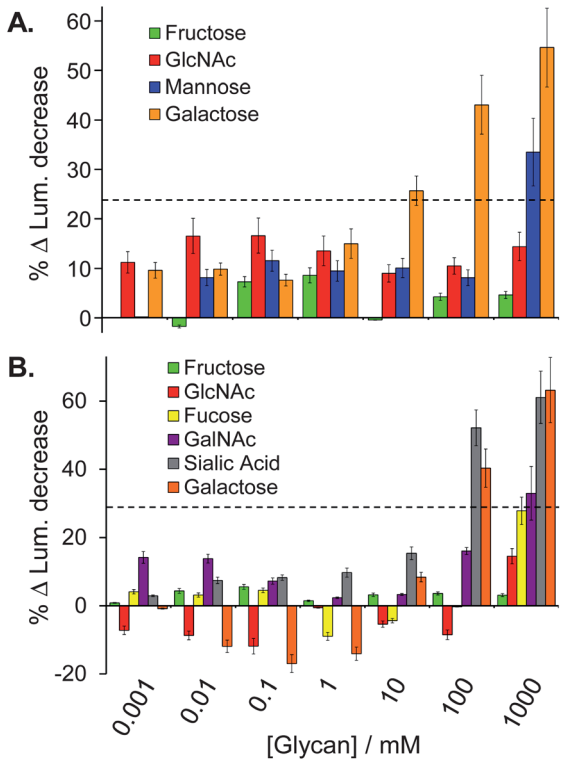


Fig. 3 Percent change in luminosity for the glycan competition studies used to explore the SL2-OVA (A) and SL5-PSM (B) binding interactions (analyte identification indicated in the legends above). Error bars represent the standard error of the percent change relative to the control as this propagated uncertainty is based on the variance between replicate measurements for the sample and control reference. The dashed lines in each panel indicate a competition threshold, based on three standard deviations above the noise. Signal response above this threshold indicates significant competition.

with more effective competition (intensity = (initial − final)/initial) to more clearly show the competition trends.

It is noteworthy that effective competition was only observed at high concentrations of the monovalent saccharides being studied. This result is likely due to the fact that these monosaccharide guests poorly compete with the multivalent display of saccharides found on the glycoproteins for binding to the multivalent display of SLs on the bead, as multivalent interactions are nearly always stronger than the sum of the monovalent interactions.[51] Also note that reducing glycosides and non-reducing monosaccharides (as found on the glycoproteins) were both used for these competition experiments, and that both classes of compounds showed similar trends in the data. The results from the competition studies with the reducing sugars are shown in Fig. 3 and the non-reducing sugar competition study data are summarized in the supporting information (Fig. S6, ESI†). Given that reducing monosaccharides can isomerize to the furanose form to provide a diol that more effectively binds to boronic acids in a 1 : 1 manner,[52–56] these monosaccharides provide a more stringent test of ligand binding than the non-reducing saccharides because they provide a "dual-competition" pathway. Namely *via* 1 : 1 furanose–boronic acid binding as well as the proposed pyranose–SL binding predicted for the saccharides found on the glycoproteins. Note that significant competition was defined as being three standard deviations above the noise (indicated by the dashed lines in Fig. 3). For this analysis the standard error for 1000 mM galactose was used because it displays the largest variance, thus for SL2 and SL5, the 'cut-off' percent change in luminosity was 23% and 29%, respectively.

For SL2, no appreciable decrease in luminosity was observed with *N*-acetylglucosamine (GlcNAc) even at concentrations as high as 1 M (Fig. 3A, red bars) indicating that *N*-acetylglucosamine does not interact with SL2, and thereby suggesting that this glycan is not critical for binding SL2 to OVA. In contrast, a significant decrease in luminosity was observed with both 1 M mannose and with as little as 10 mM galactose (Fig. 3A, blue and orange bars, respectively). These data indicate that SL2 is likely binding primarily with galactose, and to a lesser extent with mannose, both found on OVA. Competitive binding with non-reducing saccharides also showed significant competition with mannose (see ESI†). These results are particularly impressive because they suggest that SL2 interacts with both terminal (galactose) and core (mannose) glycan structures.[47] Given that galactose is typically considered to be a weak boronic acid binder for simple 1 : 1 binding, the observed competition suggests that the binding site in this system is organized in a manner suitable for binding this sugar.[31]

Particularly small changes in luminosity corresponding to the addition of GlcNAc or fucose to SL5 (Fig. 3B, red and yellow bars, respectively) suggest that these glycans were not crucial for SL5 binding to PSM. Conversely, GalNAc competed for binding at high concentrations (Fig. 3B, purple bars), while both sialic acid and galactose displayed significant competition with PSM for binding to SL5 at concentrations above 100 mM (Fig. 3B, gray and orange bars, respectively), suggesting that SL5 is likely interacting with these terminal glycans. The data for the non-reducing sugars also demonstrates that sialic acid and GalNAc compete for binding to SL5.

The fructose competition studies are particularly impressive because neither SL2 nor SL5 showed any significant competition with up to 1 M saccharide, *i.e.*, less than 10% observed decrease in the glycoprotein binding signal (Fig. 3, green bars). Since fructose is one of the strongest known 1 : 1 binders for boronic acids, the lack of competition with fructose provides further evidence that the SL–glycoprotein interactions are likely multivalent.

Discrimination of glycans

As an initial test of our approach towards binding biologically relevant targets, we used an array of SL1, SL3, SL4 and SL5 to distinguish between five structurally similar cancer associated glycans (TF antigen, Le^a, Le^x, sLe^a and sLe^x; Fig. 1A). These glycans were chosen because they represent some of the more common saccharide motifs overexpressed by cancerous cells as well as being composed of many of the same monosaccharides that were used in the above competitive binding assay with our SLs. SL2 was not included in the array to eliminate redundancy based on response similarities with SL3 and because of the high background binding to BSA as compared with SL3. It is worth noting that while SL1 has higher background binding to BSA

than SL3; it was still included in the array due to its broad yet differential, cross-reactive response to all glycoproteins assessed.

After screening each SL against a solution containing biotinylated glycan and fluorescently labeled streptavidin, luminosity values, from fluorescence microscope images, were analyzed (4 SLs by 5 glycans by 15 replicates). To account for differences in bead size and loading levels, luminosities were normalized against the highest luminosity within a given SL type (in this study the greatest degree of variability stems from bead-to-bead variations). The unique pattern generated for each different glycan based on the response of the four different SLs is shown in Fig. 4A. Note that the response for each glycan produces patterns that do not differ greatly between analytes, nevertheless the response is reproducible and the resulting patterns are unique and distinguishable within the limits of the associated error.

Though these patterns are similar they are nonetheless unique, and therefore statistical analyses were used to identify the most significant features necessary for classification of the analytes. Specifically, linear discriminant analysis (LDA) was used.[57] This analysis minimized variation within each glycan type while maximizing the differences between different glycans by creating linear combinations of each response pattern and transforming them into canonical discriminants. For this analysis, Discriminant 1 and Discriminant 2 contain 83.3% and 14.8% of the between group variation, respectively (Fig. 4B).[58] Therefore, each point in the plot contains information for an explicit measurement from the four different SLs responding to a specific glycan. Note that the different glycans are clustered into five groups with an average standard deviation of ~6%. Furthermore, the Wilks' lambda value for this analysis is 0.009 with a p-tail value of <0.000001, indicating that there is a statistically significant difference in the population means from this analysis at the 95% level of confidence.

While there is some overlap of the ellipses drawn in Fig. 4B, it is important to recognize that this plot only shows two dimensions out of the four dimensional data used for this analysis (displaying the data in three dimensions (four is not possible) does not visually enhance the ability of the plot to show discrimination).

Leave-one-out cross-validation was next used to assess the ability of the SL array to classify unknowns as the appropriate glycan.[58] This procedure sequentially removes one sample point at a time and uses the remaining points as a new training set to create a model analogous to that shown in Fig. 4B. The classification accuracy was determined by whether or not the "left-out" data point was assigned to the correct glycan grouping. Using this method each analyte response can be used as an unknown and the classification accuracy determined for the entire data set. Based on this analysis, the SL array correctly classified 71 of the 75 measured samples (94.7% classification accuracy, with a chance accuracy of only 20%). Significantly, the Lewis antigens and their sialylated forms (Le^a/Le^x and sLe^a/sLe^x) were efficiently discriminated while only differing by the addition of a terminal sialic acid moiety. Additionally, this SL-array impressively distinguished between Le^a and Le^x, as well as between sLe^a and sLe^x, glycans where the only structural difference is the regiochemistry of the linkage to the core GlcNAc moiety (Fig. 1A). Of the four misclassified glycans (Table 2), Le^a was twice identified as sLe^a, sLe^a was once classified as Le^a, and Le^x was once recognized as sLe^a.

To further evaluate the validity of our SL array for discriminating between these five structurally similar glycans, and to circumnavigate the disadvantages associated with leave-one-out cross-validation (also referred to as delete-one jackknife) the more statistically robust "boot-strapping" approach was used.[59]

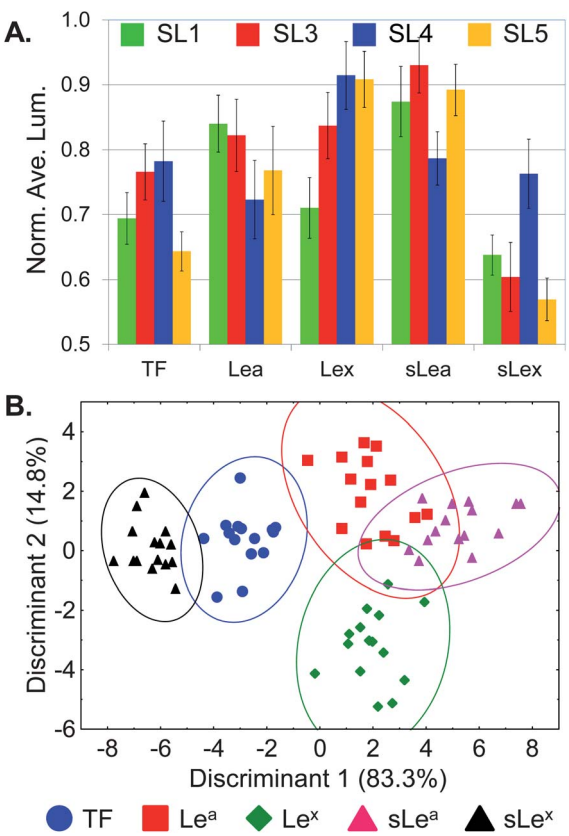


Fig. 4 Differentiation of five glycans using a SL array. (A) Fingerprint pattern of the average normalized luminosity intensities from SL1, SL3, SL4 and SL5 responding to five different glycans (TF, Le^a, Le^x, sLe^a and sLe^x). (B) The two-dimensional LDA score plot derived from the patterns shown in (A) for 15 replicates. Ellipses indicate 95% confidence level, analyte identification indicated in the legends above.

Table 2 Percent classification accuracies of glycans using an SL array as determined by different cross-validation techniques

	Jackknife	Boot-strap[a]	Training/test set[b]
Le^a	86.6	85.8	88.2
Le^x	93.3	95.3	96.0
TF	100	96.2	93.8
sLe^a	93.3	93.6	94.4
sLe^x	100	99.0	99.0
Total	94.6	94.2	93.9

[a] Average values were calculated from 50 replicate analyses of independently randomized samples with $N = 75$. [b] Average values were calculated from 25 replicate analyses of independently randomized samples at approximately 50% exclusion (randomized test samples accounted for, on average, 37 samples (49.5%), ranging from 26–43 samples).

In the approach, multiple data sets (typically 20–10 000) are generated by randomly selecting points from the original data set. During this sampling, the probability that a data point will appear '*n*' times is close to a Poisson distribution with mean unity.

The Mersenne–Twister random number generator[60] was used for random selection of data points in Systat and data sets were created with 75 elements, the same number as the original data set. Fifty (50) separate and unique data sets were generated using this approach and were then evaluated for classification accuracy. Overall, this analysis yielded a 94.2 ± 2.0% classification accuracy for the array identifying these five glycans. This is consistent with the leave-one-out accuracy of 94.6%. Significantly, individual glycans were accurately classified from 86–99% (Table 2). As with the leave-one-out analysis, the three greatest misclassifications were due to Le^a being misclassified as sLe^a (9.3%), sLe^a being misclassified as Le^a (6.7%), and Le^x being misclassified as sLe^a (4.7%).

Still further stressing the limits of this array for differentiating glycans, we chose to randomly split our data in half. Using one half as a training set, to create a statistical model, and the other half as a test set to assess the ability of this model to accurately identify these "unknowns." Training and test sets were chosen at random from the Normal distribution.[61] To minimize systematic error, random set generation and subsequent analyses were carried out 25 times to create replicates. The data in Table 2 represents the averages obtained for these replicate runs. Consistent with the previously described analyses, the overall classification accuracy of this approach was 93.9% ± 2.8%. This is by far the most stringent method used to assay the validity of the models generated from our SL array and still exhibits exceptional classification accuracy. The consistency displayed across the three methods further testifies to the strength of the outlined SL array design for discriminating structurally similar cancer associated glycans.

As indicated above, it is possible that the SLs interact, not only with the glycan, but also with the protein portion of glycoproteins. In this analysis the protein component, FITC-streptavidin, is the same for each glycan being analyzed. As such, any observed difference in the response from the array must be attributed to the glycan constituent. Given the structural similarities between these glycans, it is remarkable that there were not more misclassifications. In total, these results validate our ability to differentiate structurally similar cancer associated glycans with high accuracy using a small, cross-reactive SL array.

Discrimination of cancer cell lines

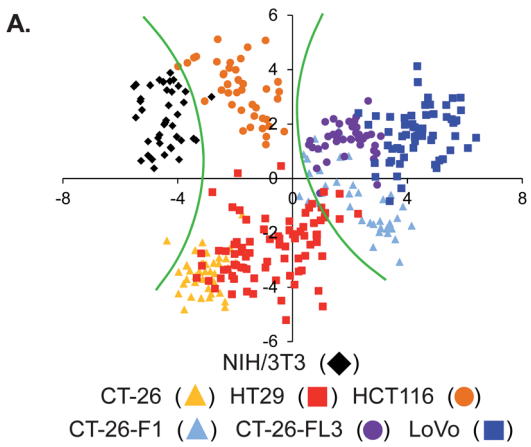
To further probe the utility of this four-component SL-array, we targeted an important goal in cancer diagnostics: to distinguish between healthy, cancerous/non-metastatic and cancerous/metastatic cell types. Specifically, we used our SL-array to discriminate between seven different cell types including: three colorectal carcinoma non-metastatic cell lines (HCT116, CT-26, HT-29), three colorectal carcinoma metastatic cell lines (CT-26-F1, CT-26-FL3, LoVo), and one murine fibroblast cell type (NIH/3T3) to serve as a "healthy" control cell line. Note that CT-26-F1 and FL3 cell lines were derived from the parental CT-26 cell line by *in vivo* education selection through serial passage in Balb/c mice and represent a series of highly similar isogenic cell lines that only differ in their metastatic potential (CT-26 <10% metastatic, CT-26-F1 ~50% metastatic and CT-26-FL3 ~95% metastatic).

Unlike the identification of discrete, structurally similar glycans, we predicted that cell type discrimination would result from a general response to the distinctive membrane protein composition of each cell type, thus affording a unique cellular signature, as previously demonstrated by Bunz and Rotello.[38] For this study, cell membrane proteins and glycoproteins were isolated[62] and fluorescently labeled to detect binding to the SL-array. While we note that this labeling approach is less than ideal for the development of a diagnostic, it does suffice to demonstrate the utility of using an SL array towards discriminating between cell lines. To account for differences in the extent of fluorescent labeling and protein concentration between each cell extract, luminosities were normalized against the highest luminosity within a given cell type (in this study the greatest degree of variability stems from cell line-to-cell line variations).[63] Note that replicates obtained for the LoVo, HCT116, NIH/3T3 and HT-29 cells were derived from multiple sample preparations of cell cultures grown by different researchers over the course of several months.

Fig. 5A shows the two-dimensional projection of the LDA results (4 SLs by 7 cell lines by 40 replicates each for NIH/3T3, CT-26, HT-29, CT-26-F1 and CT-26-FL3; 60 replicates for LoVo; and 80 replicates for HCT116). It is important to note that if all of the variance is captured in one discriminant then the statistical analysis is not really necessary; however successive discriminants containing large portions of the variance supports the validity of and the need for the statistical analysis. In this analysis Discriminant 1 contains 54% and Discriminant 2 contains 31% of the total variance, while the remaining 15% is partitioned between Discriminants 3 (11%) and 4 (4%) (*i.e.*, this is four-dimensional data). This distribution of variance suggests that each of the SLs in the array is important for discriminating between cell lines.

Note that each of the same colored points cluster together indicating the ability of the statistical model to define similarity between replicates of a specific analyte. However, some of these different clusters are closely packed and some groups overlap suggesting that there are strong similarities between some of the analytes, as would be expected. Nevertheless, it important to recognize that the data is in fact four dimensional; therefore the overlap between groups shown in this two dimensional figure (Fig. 5A) is not necessarily indicative of poor classification.

To quantitatively evaluate the accuracy of this approach, leave-one-out cross-validation was used and demonstrated that this statistical model exhibited 92.1% accuracy, correctly identifying 313 out of 340 measured samples. Fig. 5B presents the LDA classification results matrix for the assay. The cross-diagonal of the matrix corresponds to the number of accurately identified samples (set in bold). Any numbers that fall off this diagonal represent the number of misclassifications for that cell type and correspond to the misclassified cell type identity. The column on the right of the matrix provides the classification accuracy for each cell type. While the overall classification accuracy for the array is 92.1%, the accuracy for each individual cell type varies between 81–100%.



B. Leave-One-Out Cross Validation

Classification Results

Group	◆	▲	●	■	▲	●	■	% Correct
◆	**40**	0	0	0	0	0	0	**100**
▲	0	**39**	1	0	0	0	0	**98**
●	0	9	**65**	1	5	0	0	**83**
■	0	0	0	**39**	0	1	0	**93**
▲	0	0	0	0	**33**	7	0	**81**
●	0	0	0	0	2	**37**	1	**98**
■	0	0	0	0	0	0	**60**	**100**
Total	**40**	**48**	**40**	**65**	**40**	**45**	**62**	**340**

Classification accuracies = **92%**

Classification Error = 8%

Expected Chance Accuracy = 14.3%

Fig. 5 (A.) The two-dimensional LDA score plot of the response of the SL array for discriminating seven cell types. Green curves indicate boundaries between healthy, cancerous/non-metastatic and cancerous/metastatic cell types. For clarity, the Discriminant 1 *vs.* Discriminant 2 data was rotated 20° about the *z*-axis (analyte identification indicated in the legends above). (B.) Leave-one-out cross validation classification matrix for the SL-array based assay.

Given the diversity of protein and glycan structures present on the cell membrane for each of these different cell types, it is difficult to speculate on the specific glycans that are recognized by the SLs and that contribute to the discrimination of these different cell lines. Still, there are clear trends in the statistical output that support the validity of this analysis. As one moves from left to right along the *x*-axis in Fig. 5A the metastatic potential of the cell lines increases. Specifically, the green curves in Fig. 5A provide boundaries between the "healthy" 3T3 cells (black) at the far left of this plot; the cancerous/non-metastatic cell lines (HCT116, CT-26 and HT-29 – orange, yellow, red, respectively) in the middle and the cancerous/metastatic cell lines (CT-26-F1, CT-26-FL3 and LoVo – light blue, purple, blue, respectively) to the right. This clustering of cell types with similar metastatic potential suggests that the basis upon which the first two discriminants are derived correlate highly with this attribute. Additionally, the Wilks' lambda value for this analysis is 0.003 with a p-tail value of <0.000001, thus indicating that there is a statistical difference in the population means from this analysis at the 95% level of confidence. Further MANOVA treatment of the data provided a Wilks' lambda value of 0.004 with a p-value of <0.000001 and sequential univariate F-Tests for each variable provided p-values of <0.000001 for each.

To further validate this approach, boot-strapping and training/test set analyses (at a 50% exclusion split) were carried out. The results provided in Table 3 indicate that these more rigorous validation methods provide classification accuracies consistent with those obtained for the leave-one-out cross-validation, 92.1 ± 1.1% and 92.7 ± 1.8%, respectively. As seen for the glycan analysis above, cell-line misclassifications were consistent across all three validation methods. Furthermore, the misclassified cell-lines were not random but often had a structural basis behind the result. For example, in the boot-strap analysis, CT-26-F1 displayed the lowest classification accuracy at 80.5%; and all of the misclassifications were as CT-26-FL3, an isogenic, highly metastatic cell line. Similarly, from the training/test set analysis, CT-26-FL3 has one of the lower classification accuracies (87.4%); here all of the misclassifications in this analysis were attributed to CT-26-F1 (85%) and LoVo (15%). Recall that both CT-26-FL3 and LoVo are highly metastatic and that CT-26-FL3 is isogenic with CT-26-F1. Finally, while the classification error for HCT116 is relatively large across the validation methods (classification accuracies from 81.3–89.2%), the majority of misclassifications are CT-26 and HT-29 cells. Since all three cell lines are cancerous non-metastatic, these misclassification are not unexpected because classification accuracy, in this model, correlates with metastatic potential.

Directed partitioning for enhanced cancer cell discrimination

With the advancement of cross-reactive sensor arrays, numerous statistical and non-statistical approaches have become available to evaluate the array responses; however, many do not scale well with increasing numbers of analyte classes. For analysis of these multi-class systems, the most common statistical approaches rely on multivariant analysis, such as feature selection algorithms. Alternatively, the analysis can be reduced to a series of multiple

Table 3 Percent classification accuracies of cell lines using an SL array as determined by different cross-validation techniques

	Jackknife	Boot-strap[a]	Training/test set[b]
3T3/NIH	100	100	100
CT-26	97.5	97.0	96.7
CT-26-F1	82.5	80.5	82.3
CT-26-FL3	92.5	92.7	87.4
HCT116	81.3	83.6	89.2
HT-29	97.5	96.8	96.7
LoVo	100	99.9	100
Total	92.1	92.1	92.7

[a] Average values were calculated from 100 replicate analyses of independently randomized samples with $N = 340$. [b] Average values were calculated from 25 replicate analyses of independently randomized samples at approximately 50% exclusion (randomized test samples accounted for, on average, 173 samples (50.9%), ranging from 153–191 samples).

binary classification problems run in parallel, such as one-from-*n* (one-against-rest), pairwise (one-against-one) or hierarchical (decision trees) processes.

We have previously presented a hybrid approach, for the identification and discrimination of biogenic amines,[64] where the multi-class system is simplified in a manner analogous to the binary classification routines. However, this class reduction did not rely on statistical methods; instead, we used insight into the chemical nature of the analytes to group these compounds into structurally related categories.

In training the array using this directed partitioning technique, previous knowledge about the nature of the samples is required, for example whether the cell lines are cancerous or not. However, as described above, no specific information about the exact identity of the analytes is necessary, for example the glycan being bound. This method is in direct contradiction with traditional routines that rely solely on statistical models. The quality of the results from this approach is often enhanced because logical reasoning, based on the inherent nature of the samples, is involved as part of the partitioning. Once classified into groups, these subsets could be further categorized as the individual components using a hierarchical, group–ungroup, multi-layered analysis approach to achieve enhanced classification. Therefore, directed partitioning was used to reduce classification error and the data were grouped according to their metastatic potential, *i.e.* healthy, cancerous/non-metastatic and cancerous/metastatic.

When the analysis was performed using these new groups, classification accuracies, based on leave-one-out cross-validation, improved to 97.1%, correctly identifying 330 out of 340 samples (Fig. 6A). The classification accuracy is unchanged using the training/test set analysis at 50% exclusion (97.3 ± 1.5%). From a diagnostic perspective, this is perhaps the most important classification; to determine whether the cancer is present or not. Of the 10 misclassified samples, 8 were cancerous/non-metastatic that were identified as cancerous/metastatic and the remaining 2 were cancerous/non-metastatic that were considered healthy, thus producing a 0.6% "false negative" rate. Additionally, note that the data for the 3T3 cells seems "bimodal," showing two distinct clusters within the category. This separation results from combining data acquired by different experimentalists from different culture broths. Most significantly, while this separation is noticeable, the overall clustering is still quite tight and the 3T3 classification is 100% in the leave-one-out analysis. Based on the training/test set analysis, the within group misclassification is 6.7%, resulting in an overall 0.8% "false-positive" rate. These results clearly support the validity of this approach to identify cancerous from noncancerous cell lines. Furthermore, the low false negative rate compares quite favorably with current diagnostic tests such as the CEA test, where the false negative rate is 16%.[22]

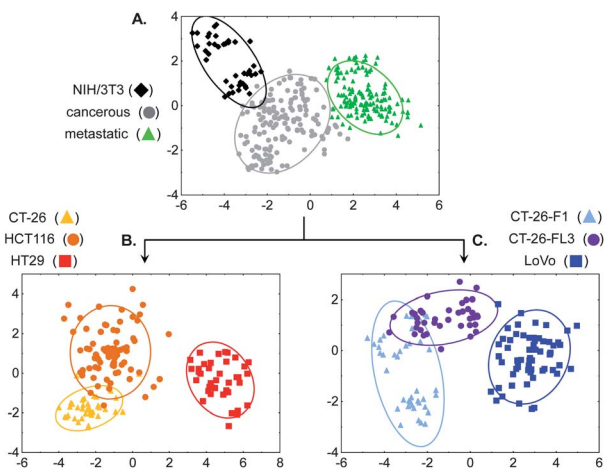


Fig. 6 (A) The 2-D LDA score plot of the response of the SL array for discriminating grouped healthy, grouped cancerous/non-metastatic and grouped cancerous/metastatic cell types. (B) 2-D LDA score plot of the array response to ungrouping the cancerous/non-metastatic cells: HCT116, CT-26 and HT-29. (C) 2D LDA score plot of the array response to ungrouping the cancerous/metastatic cells: CT-26-F1, CT-26-FL3 and LoVo. Ellipses indicate 95% confidence level, analyte identification indicated in the legends above.

By successively ungrouping each subset, a multi-layered analysis could be carried out to identify the individual cell type. The two-dimensional projections of the four-dimensional LDA results for these subset categorizations are shown in Fig. 6B–C. In Fig. 6B cancerous/non-metastatic cell lines were accurately discriminated in 150 out of 160 samples or 94%; an improvement from 89% in the single-layer analysis. Specifically, HT-29 cells were classified with 100% accuracy; CT-26 cells achieved 98% classification accuracy and HCT116 were classified with 89% accuracy. For the 10 misclassified analytes, 9 of the HCT116 samples were identified as CT-26 while one CT-26 was classified as HCT116. Given that all three of these cell lines are cancerous non-metastatic, these misclassification are not extraordinary because classification accuracy, in this model, correlates with metastatic potential.

Similarly, the cancerous/metastatic cell lines were separated into the individual components with 92% classification accuracy (129 out of 140 samples, Fig. 6C). In this analysis, 91% of the misclassifications resulted from mis-assignments between CT-26-F1 and CT-26-FL3. It is important to recall that these are highly similar isogenic cell lines, derived from the parental CT-26 cell line, and differ only in their metastatic potential. The impressive 88% classification accuracy, between the highly metastatic cell lines CT-26-F1 and CT-26-FL3, as well as 92% classification accuracy between the parent CT-26, and metastatic CT-26-F1 and CT-26-FL3 cell lines further validates our approach while indicating that there are distinct glycosylation patterns associated with metastatic potential. These results highlight the adaptability of this array-based approach for classifying cell types based on complex mixtures rather than a specific analyte, thereby mimicking the mammalian senses of taste and smell.[39,40]

Conclusions

In summary, selective and cross-reactive SLs have been identified by screening a resin-based SL library binding to glycoproteins. Selectivities as high as ~50-fold, for one glycoprotein over another, have been observed. The selectivity of the SL-glycoprotein interactions are maintained in 95% human serum, demonstrating their robustness. Significantly, SLs were assembled into an array format to distinguish between five structurally similar cancer associated glycans with 94% accuracy. Additionally, the same array was used to discriminate seven cell types,

including three colorectal carcinoma non-metastatic cell lines, three colorectal carcinoma metastatic cell lines, and one healthy control cell line with high accuracy. Two statistical methods were employed for this analysis. In a single layered approach, analysis of all seven analytes at once provided overall classification accuracy above 92%. Using directed partitioning afforded 97% accuracy for distinguishing between cancerous non-metastatic, cancerous metastatic and healthy cells. By sequentially ungrouping these subsets the overall accuracy of the analysis was improved compared with the single-layer analysis. Current work is focused on identifying SLs for specific cancer associated targets to enhance detection sensitivity and discrimination ability, as well as expanding the array to discriminate between other glycans and cell types. Finally, we note that SLs themselves may possess therapeutic utility as targeting agents and metastatic inhibitors, as has been shown with natural lectins.[65,66]

Acknowledgements

We thank Dr J. E. Jones and Dr O. Obianyo for their help with fluorescence polarization. This work was supported by funds provided from NIH COBRE grant P20RR17698.

Notes and references

1 D. H. Dube and C. R. Bertozzi, *Nat. Rev. Drug Discovery*, 2005, **4**, 477–488.
2 V. Harmat and G. Naray-Szabo, *Croat. Chim. Acta*, 2009, **82**, 277–282.
3 J. J. Lavigne and E. V. Anslyn, *Angew. Chem., Int. Ed.*, 2001, **40**, 3118–3130.
4 F. P. Schmidtchen, *Chem. Soc. Rev.*, 2010, **39**, 3916–3935.
5 A. P. Umali and E. V. Anslyn, *Curr. Opin. Chem. Biol.*, 2010, **14**, 685–692.
6 J.-P. Behr, The Lock and Key Principle: The State of the Art 100 Years on, in *Perspect. Supramol. Chem.*, 1994; 1.
7 M. Brattoli, G. de Gennaro, V. de Pinto, A. D. Loiotile, S. Lovascio and M. Penza, *Sensors*, 2011, **11**, 5290–5322.
8 M. Cole, J. A. Covington and J. W. Gardner, *Sens. Actuators, B*, 2011, **156**, 832–839.
9 R. Paolesse, D. Monti, F. Dini and C. Di Natale, *Top. Curr. Chem.*, **300**, 139–174.
10 R. K. Ranjan and K. Prasad, *Anal. Chem–Indian J.*, 2008, **7**, 739–742.
11 F. Roeck, N. Barsan and U. Weimar, *Chem. Rev.*, 2008, **108**, 705–725.
12 A. D. Wilson and M. Baietto, *Sensors*, 2009, **9**, 5099–5148.
13 J. Yinon, *Anal. Chem.*, 2003, **75**, 98A–105A.
14 M. A. Hollingsworth and B. J. Swanson, *Nat. Rev. Cancer*, 2004, **4**, 45–60.
15 T. Nakagoe, T. Sawai, T. Tsuji, M. A. Jibiki, A. Nanashima, H. Yamaguchi, T. Yasutake, H. Ayabe and K. Arisawa, *Hepatogastroenterology*, 2003, **50**, 696–699.
16 W. S. Wang, J. K. Lin, T. C. Lin, T. J. Chiou, J. H. Liu, C. C. Yen, W. S. Chen, J. K. Jiang, S. H. Yang, H. S. Wang and P. M. Chen, *Hepatogastroenterology*, 2002, **49**, 388–392.
17 J. L. Magnani, *Arch. Biochem. Biophys.*, 2004, **426**, 122–131.
18 S. E. Baldus, T. K. Zirbes, S. P. Monig, S. Engel, E. Monaca, K. Rafiqpoor, F. G. Hanisch, C. Hanski, J. Thiele, H. Pichlmaier and H. P. Dienes, *Tumor Biol.*, 1998, **19**, 445–453.
19 M. M. Fuster, J. R. Brown, L. Wang and J. D. Esko, *Cancer Res.*, 2003, **63**, 2775–2781.
20 J.-i. Ogawa, A. Sano, S. Koide and A. Shohtsu, *J. Thorac. Cardiovasc. Surg.*, 1994, **108**, 329–336.
21 S. Nakamori, M. Kameyama, S. Imaoka, H. Furukawa, O. Ishikawa, Y. Sasaki, Y. Izumi and T. Irimura, *Dis. Colon Rectum*, 1997, **40**, 420–431.
22 M. G. Fakih and P. Aruna, *Oncology*, 2006, **20**, 579–587.
23 D. Walker, G. Joshi and A. Davis, *Cell. Mol. Life Sci.*, 2009, **66**, 3177–3191.
24 S. Jin, Y. Cheng, S. Reid, M. Li and B. Wang, *Med. Res. Rev.*, 2010, **30**, 171–257.
25 T. D. James, K. R. A. S. Samankumara and S. Shinkai, *Angew. Chem., Int. Ed.*, 1997, **35**, 1911–1922.
26 T. D. James and S. Shinkai, *Top. Curr. Chem.*, 2002, **218**, 159–200.
27 J. J. Lavigne and E. V. Anslyn, *Angew. Chem., Int. Ed.*, 1999, **38**, 3666–3669.
28 M. Li, N. Lin, Z. Huang, L. Du, C. Altier, H. Fang and B. Wang, *J. Am. Chem. Soc.*, 2008, **130**, 12636–12638.
29 Yamamoto, M. M. Takeuchi and S. Shinkai, *Tetrahedron*, 1998, **54**, 3125–3140.
30 W. Yang, H. Fan, X. Gao, S. Gao, V. V. R. Karnati, W. Ni, W. B. Hooks, J. Carson, B. Weston and B. Wang, *Chem. Biol.*, 2004, **11**, 439–448.
31 T. D. James, M. D. Phillips and S. Shinkai, *Boronic acids in saccharide recognition*, Royal Society of Chemistry, Cambridge, UK, 2006.
32 P. J. Duggan and D. A. Offermann, *Tetrahedron*, 2009, **65**, 109–114.
33 A. Pal, M. Bérubé and D. G. Hall, *Angew. Chem., Int. Ed.*, 2010, **49**, 1492–1495.
34 T. D. James, H. Shinmori and S. Shinkai, *Chem. Commun.*, 1997, 71–72.
35 K. L. Bicker, J. Sun, J. J. Lavigne and P. R. Thompson, *ACS Comb. Sci.*, 2011, **13**, 232–243.
36 Y. Zou, D. L. Broughton, K. L. Bicker, P. R. Thompson and J. J. Lavigne, *ChemBioChem*, 2007, **8**, 2048–2051.
37 N. Y. Edwards, T. W. Sager, J. T. McDevitt and E. Anslyn, *J. Am. Chem. Soc.*, 2007, **129**, 13575–13583.
38 A. Bajaj, O. R. Miranda, I.-B. Kim, R. L. Phillips, D. J. Jerry, U. H. F. Bunz and V. M. Rotello, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10912–10916, S10912/10911–S10912/10910.
39 J. J. Lavigne and E. V. Anslyn, *Angew. Chem., Int. Ed.*, 2001, **40**, 3118–3130.
40 A. T. Wright and E. V. Anslyn, *Chem. Soc. Rev.*, 2006, **35**, 14–28.
41 N. G. Karlsson, H. Nordman, H. Karlsson, I. Calstedt and G. C. Hansson, *Biochem. J.*, 1997, **326**, 911–917.
42 S. M. D'Arcy, C. M. Donoghue, C. A. Koeleman, D. H. V. d. Eijnden and A. V. Savage, *Biochem. J.*, 1989, **260**, 389–393.
43 S. Martensson, S. B. Levery, T. T. Fang and B. Bendiak, *Eur. J. Biochem.*, 1998, **258**, 603–622.
44 *Combinatorial Chemistry Catalog and Solid Phase Organic Chemistry (SPOC) Handbook*, Novabiochem, Laufelfingen, 1996.
45 *GraFit*, Erithacus Software Limited, Version 5.0.11 edn, 2004. Note that when the data were fit to a two-site binding model no significant differences in the calculated K_d values were apparent, $K_d1 = 0.47 \pm 40.51$ μM and $K_d2 = 43.47 \pm 41.40$ μM. However, the errors are quite large for this later analysis while the single site model afforded a significantly better fit to the data.
46 N. J. Moerke, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 1–15.
47 D. J. Harvey, D. R. Wing, B. Kuster and I. B. Wilson, *J. Am. Soc. Mass Spectrom.*, 2000, **11**, 564–571.
48 K. T. Pilobello, L. Krishnamoorthy, D. Slawek and L. K. Mahal, *ChemBioChem*, 2005, **6**, 985–989.
49 G. Springsteen and B. Wang, *Tetrahedron*, 2002, **58**, 5291–5300.
50 S. Jin, C. Zhu, Y. Cheng, M. Li and B. Wang, *Bioorg. Med. Chem.*, 2010, **18**, 1449–1455.
51 M. Mammen, S.-K. Choi and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 1998, **37**, 2754–2794.
52 M. Bielecki, H. Eggert and J. C. Norrild, *J. Chem. Soc., Perkin Trans. 2*, 1999, 449–456.
53 S. P. Draffin, P. J. Duggan, S. A. M. Duggan and J. C. Norrild, *Tetrahedron*, 2003, **59**, 9075–9082.
54 H. Eggert, J. Frederiksen, C. Morin and J. C. Norrild, *J. Org. Chem.*, 1999, **64**, 3846–3852.
55 J. C. Norrild and H. Eggert, *J. Am. Chem. Soc.*, 1995, **117**, 1479–1484.
56 J. C. Norrild and H. Eggert, *J. Chem. Soc., Perkin Trans. 2*, 1996, 2583–2588.
57 *Systat*, Version 11.00.01, Systat Software, Inc., 2004.
58 K. R. Beebe, R. J. Pell and M. B. Seasholtz, *Chemometrics. A Practical Guide.*, John Wiley & Sons, Inc., New York, 1998.

59 C. F. J. Wu, *Ann. Stat.*, 1986, **14**, 1261–1295.
60 M. Matsumoto and T. Nishimura, *ACM Transactions on Modeling and Computer Simulation*, 1998, **8**, 3–30.
61 G. Casella and R. L. Berger, *Statistical Inference*, Thomas Learning, Pacific Grove, CA, 2002.
62 T. Nakamura, T. Hayashi, Y. Nishimura-Nasu, F. Sakaue, Y. Morishita, T. Okabe, S. Ohwada, K. Matsuura and T. Akiyama, *Genes Dev.*, 2008, **22**, 1244–1256.
63 Control studies were carried out to assure that the array response was independent of concentration or extent of glycoprotein labeling with the fluorophore.
64 T. L. Nelson, I. Tran, T. G. Ingaliinera, M. S. Maynor and J. J. Lavigne, *Analyst*, 2007, **132**, 1024–1030.
65 G. Mannori, D. Santoro, L. Carter, C. Corless, R. M. Nelson and M. P. Bevilacqua, *Am. J. Pathol.*, 1997, **151**, 233–242.
66 T. Minko, *Adv. Drug Delivery Rev.*, 2004, **56**, 491–509.

