RSC Advances

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. This *Accepted Manuscript* will be replaced by the edited, formatted and paginated article as soon as this is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

www.rsc.org/advances

Journal Name

ARTICLE

Discrimination between Lectins with Similar Specificities by Ratiometric Profiling of Binding to Glycosylated Surfaces; A Chemical 'Tongue' Approach

L. Otten and M. I. Gibson [a]

Carbohydrate-lectin interactions dictate a range of signalling and recognition processes in biological systems. The exploitation of these, particularly for diagnostic applications, is complicated by the inherent promiscuity of lectins along with their low affinity for individual glycans which themselves are challenging to access (bio)synthetically. Inspired by how a 'tongue' can discriminate between hundreds of flavours using a minimal set of multiplexed sensors and a training algorithm, here individual lectins are 'profiled' based on their unique binding profile (barcode) to a range of monosaccharides. By comparing the relative binding of a panel of 5 lectins to 3 monosaccharide-coated surfaces, it was possible to generate a training algorithm that enables correct identification of lectins, even those with similar glycan preferences. This is demonstrated to be useful for discrimination between the Cholera and Ricin Toxin lectins showing the potential of this minimalist approach to exploiting glycan complexity.

Introduction

Protein carbohydrate interactions are essential for many biological processes including cell-cell communication, fertilisation and innate immunity.[1] They are also readily exploited by pathogens during adhesion steps. These adhesion steps are mediated by carbohydrate binding proteins known as lectins. In Nature, multivalent presentation of glycans at cell surfaces increases the affinity towards its binding partner which has been widely exploited to create synthetic glycomimetics, such as glyco-polymers[2] and particles.[3-4] [5] This strategy however does not necessarily maintain or improve the selectivity complicating the design and application of multivalent glycoconjugates.[2, 5-7] [8]

Lectin interactions are mediated by the carbohydrate itself but also the linker between the carbohydrate, the cell surface and precise 3D presentation of carbohydrates on the cell surface.[6, 9] Many lectins show highly specific binding to oligosaccharides but show much more promiscuous binding characteristics on a mono- and di-saccharide level. For example, peanut agglutinin (PNA) is generally described as being β-galactose specific but microarray analysis shows that it will readily bind all monosaccharides with very little difference between them.[10] The same is also true for cholera toxin, this toxin is highly specific to the GM-1 ganglioside in the body and thus is described as being galactose specific but this lectin will indiscriminately bind all monosaccharides to one degree or another.[10]

This wide variety of roles played by glycans in the body's innate processes and their prevalence in nature means the interference or detection of these interactions could have an impact in combatting infectious diseases.[11] For example, FimH is a lectin involved in the binding of uropathogenic *Escherichia coli* to mannose rich residues and is a crucial virulence factor. Cholera is caused by cell internalisation of an AB_5 toxin, mediated by the 5 lectin subunits of the toxin initiating binding to GM-1 on epithelial cells in the small intestine. Ricin is a toxic protein extracted from *Ricinus communis* seeds, it consists of one subunit responsible for cleaving an adenine residue from the 28S ribosomal RNA (thus rendering the cell incapable of protein synthesis) and one subunit responsible for binding to galactose rich residues.[12] Differences in glycosylation of cells have also been implicated in tumour cells and determining metatastic potential of cancers[13-14] and the ABO blood system is also determined by different antigenic oligosaccharides.[11, 15] Serological blood groups have been implicated in individual susceptibility to many diseases and the severity of others including small pox, cholera and malaria.[15-17] As such rapid detection of lectins can aid in the early identification and prevention of diseases and also in the design of therapeutics. This broad window of binding partners means that the design of a sensor for a lectin based on glycans alone is immensely challenging.

Whilst proteomic and antibody based techniques can be used for identification of lectins these are not always suitable for robust, point of care applications, and require infrastructure for preparation, storage, distribution and deployment of the sensor. Such a challenge is indeed not unique to glycobiology, and the detection of cell phenotypes, which often have dynamic surface ligand displays which change with their environment. To address this nanoparticles multiplexed biosensing has attracted much interest especially for diagnostics. [18] Rotello et al. have developed the use of differentially functionalised gold nanoparticles for multiplexed diagnostics. For

example, 52 different mixtures of seven different proteins could be identified using just six distinct nanoparticles.[19] Gold particles coated with 3 different thiols enabled cancerous and healthy cells to be discriminated without the requirement for any specific binding epitopes.[20] Detection of pathogenic bacteria using a related system in under 5 minutes has also been demonstrated[21] as have MRI based detection of cancerous cells with differential lectin expression levels.[22] Jayawardena et al. have described the use of glycosylated gold nanoparticles and their characteristic shift in SPR frequencies upon protein binding to characterise lectins based on their response to a panel of sugars.[23] In this case, lectins with very different glycan specificities were used (e.g. Concanavalin A/Soybean Agglutinin) and discrimination was also possible without the need for multiplexing and just using individual glycans making it a less challenging analysis.

The goal of the present research was to evaluate the use of simple and synthetically accessible mono-saccharides as multiplexed sensors to enable discrimination between different lectins which have similar binding specificities. Such a system would have widespread application especially for low-cost selective detection/monitoring of toxins.

Results and Discussion

The key aim of this work was to probe the differential response of lectins to simple carbohydrates (monosaccharides), so it was essential to employ accessible (/facile) coupling chemistry. 96-multiwell plates with hydrazide functionality were used to couple a range of monosaccharides, and mixtures of different monosaccharides using and aniline catalyst at 50 °C to give glycosylated surfaces (Figure 1A). This coupling mechanism is known to result in attachment of the monosaccharides predominantly in their ring closed (pyranose) β-anomeric form.[24] It should be noted that the presence of some acyclic species does not affect our later analysis using a training algorithm (*vide infra*). It was not possible to interrogate the functionalised polypropylene surface of the microwell plate using traditional surface analysis techniques (such as elipsometry). As an alternative to contact angle, droplet spread was measured. Hydrophobic surfaces, when viewed from above should give reduced surface coverage at equal volume, compared to a hydrophilic surface with a low contact angle. The native, and glycosylated surfaces were therefore interrogated by addition of a drop of ultra-pure water with resorufin (a dye) added to enable visualisation of the droplet spread, and subsequent image analysis. The native surfaces resulted in only 30 % of the surface being covered by the drop, but the galactose functional surface resulted in droplet spreading over 50 % of the surface (Figure 1B). As a positive control, glyceraldehyde was added to generate very hydrophilic surface coating, and this resulted in spreading over ~ 90 % of the surface, confirming this (unconventional) analytical approach. Glycan-modified surfaces should also present an uncharged, non-fouling surface, compared to the native hydrazide/polypropylene surface. Therefore, non-specific fouling (adsorption) was tested using FITC-labelled bovine serum albumin. Compared to the native surface, the glycosylated surfaces showed significantly less binding than the native surface, with no significant

absorption observed at concentrations below 0.2 mg.mL^{-1} again confirming the surface modification (Figure 1C).
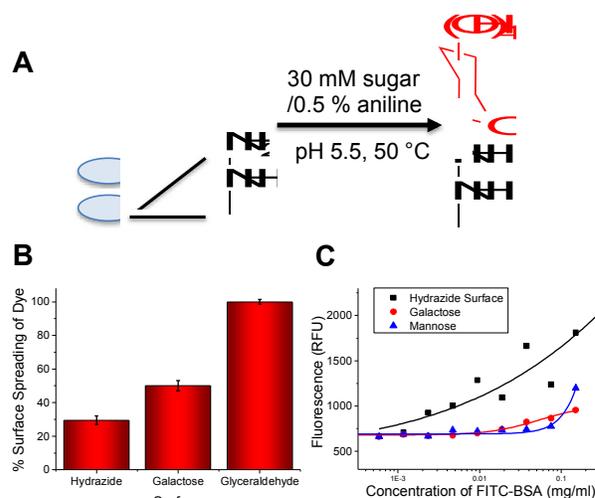

Figure 1. Fabrication of glycosylated 96-well plates. A) Hydrazide-carbohydrate coupling; B) Dye spreading assay showing relative hydrophilicity of surfaces; C) Non-specific binding of fluorescently-labelled bovine serum albumin onto different surfaces following 30 minutes incubation and washing.

To highlight the challenges faced in identification and profiling of lectins with similar binding specificities, a panel of 5, fluorescently labelled, galactose (or GalNAc) binding lectins were selected, exposed to a galactose microwell plate, washed and total fluorescence measured. Figure 2 shows the results of this, indicating that at any given concentration the total response recorded is not unique to any given lectin. Cholera toxin B subunit (CTx) gives higher binding than the others, but the absolute fluorescence intensity is obviously dependent on the concentration applied, which is not ideal for any realistic biosensory format as it requires significant prior knowledge of the solution being probed.
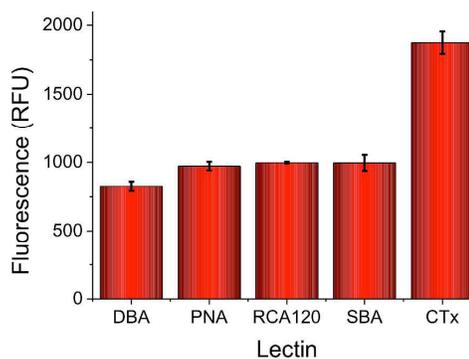

Figure 2. Relative binding of a panel of 5 lectins to a Galactose-functional surface as judged by fluorescence intensity. All lectins applied at 0.01 mg.mL^{-1}, with FITC labels.

Considering the low information content of these single sugar assays, we proceeded to extract information for a series of Gal-binding lectins from the CFG database (consortium for functional

glycomics) to range a small mono/disaccharides (Figure of this analysis included in ESI). The CFG data revealed that any single glycan cannot predict the identity of the lectins (i.e. a single peak is not present) due to their inherent promiscuity. However, if many different glycans are included, there is a unique pattern of binding of each lectin to the carbohydrates (a 'barcode'). Guided by this data, we rationalised that if we could identify the 'minimum basis set' of glycans that can provide a unique barcode for each lectin, it would be possible to distinguish between these, enabling protein identification without proteomics or associated methods. Using the hydrazide coupling chemistry described above, we generated 4 differently glycosylated surfaces; Gal, Man, Glc and a 1:1 mixture of Gal:Man (the latter was added as in our hands this improves the resolution of our subsequent analysis. Variable density glycan mixtures are known to give non-linear responses[7]). Pleasingly, these relatively low-affinity monosaccharides produced very unique binding profiles for each lectin, as shown in Figure 3. For example, *Ricinus communis* Agglutinin (RCA_{120}) had significantly higher binding to galactose, and the Gal/Man mixtures, than compared to Glc binding. Conversely, Soybean Agglutinin (SBA) had significantly depressed binding to the mixed surface. A summary of the relative binding of the lectins can be shown in a heat map to give a 'bar-code' which is unique to each protein.
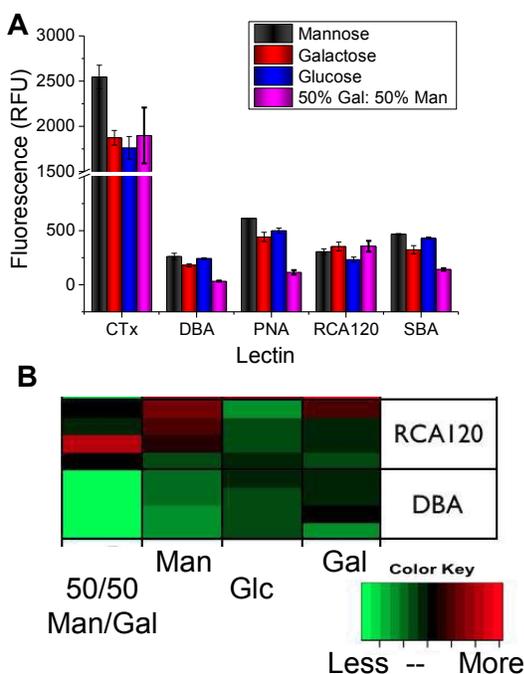

Figure 3. A) Relative lectin binding to glycosylated surfaces determined by fluorescence; B) Heatmap, demonstrating that each protein has a 'barcode' of responses to each glycan, each lectin displayed contains at least 4 independent replicates.

Analysis of the individual binding of one lectin to a sugar does not give much information, but when combined together, this differential response provides sufficient information to enable a linear discriminant analysis. Linear discriminant analysis is a training algorithm that inputs a matrix of data and produces a model in which all of the categories in the initial training matrix are grouped into distinct categories based on their linear discriminant factors (which are a linear combination of the initial inputs- in this case the surfaces used). Due to the high degree of separation between categories within the model produced it allows for greater confidence in the identification of lectins responsible for binding in unknown samples when compared to the raw data alone.
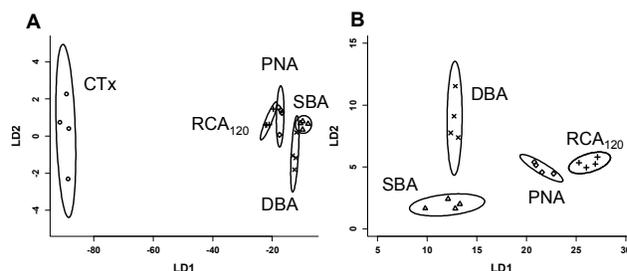

Figure 4. Linear discriminant analysis of lectin binding to the 4 different glycosyatled surfaces. A) lectins with CTx, and B) lectins without CTx.

Figure 4A shows the results of a linear discriminate analysis of these lectins to the four glycosylated surfaces, revealing highly resolved groupings for each lectin. The circles around each are indicative of a 95 % confidence boundary. Figure 4B shows the LD analysis for the lectins without CTx, as when this is included the other four lectins appear more tightly bunched (but are still perfectly resolved) due to the generally increased binding of CTx to all surfaces employed here. This simple, but powerful, multiplexed method enables separation and identification of lectins with similar binding profiles, but without the need for complex carbohydrates, in much the same way as a tongue has evolved to identify complex tastes based on only 5 different inputs. To test the predictive power of this, blind analysis of unknown lectin samples was also conducted, revealing 100 % predictive accuracy from this training matrix.

As a final test of this sensing approach, the differentiation between two different gal-binding, pathogenic, lectins was investigated. CTx is the toxin secreted by the bacteria *Vibrio cholera,* which causes cholera and is a huge problem in developing countries and disaster zones. RCA_{120} is a surrogate for ricin, which can be weaponised as a biological warfare agent. A training algorithm was again employed, but this time the RCA_{120}/CTx solutions were applied as mixtures of the two lectins, rather than as pure protein solutions. This provides a far more challenge test, which is closer to a real world sensing application. When CTx was present at > 50 % (by mass) the LD model correctly indicated its presence, and when the RCA_{120} concentration was above 50 %, this was correctly scored (see ESI for full details and LDA graphs).

Conclusions

Here we have reported the new concept of a 'chemical tongue' for multiplexed biosensing, and discrimination between carbohydrate-binding proteins (lectins). We show that using only

simple monosaccharides, which have very low intrinsic affinity and specificity, it is possible to discriminate between a panel of lectins with extremely similar binding preferences. The power of this method lies in the scalability, enabling many more (oligo)saccharides to be employed, and the use of the large glycan databases (which are freely available) to guide the design of each system. Using this approach we demonstrated that the chemical tongue can even distinguish the presence of cholera or ricin, in complex mixtures of the two lectins. Current and future work is focused on establishing the limits and scope of this method, and translating it into realistic sensory surfaces/components which would remove the need for labelled proteins.

Experimental

Full experimental detail including microplate functionalization LDA analysis is provided in the electronic supporting information
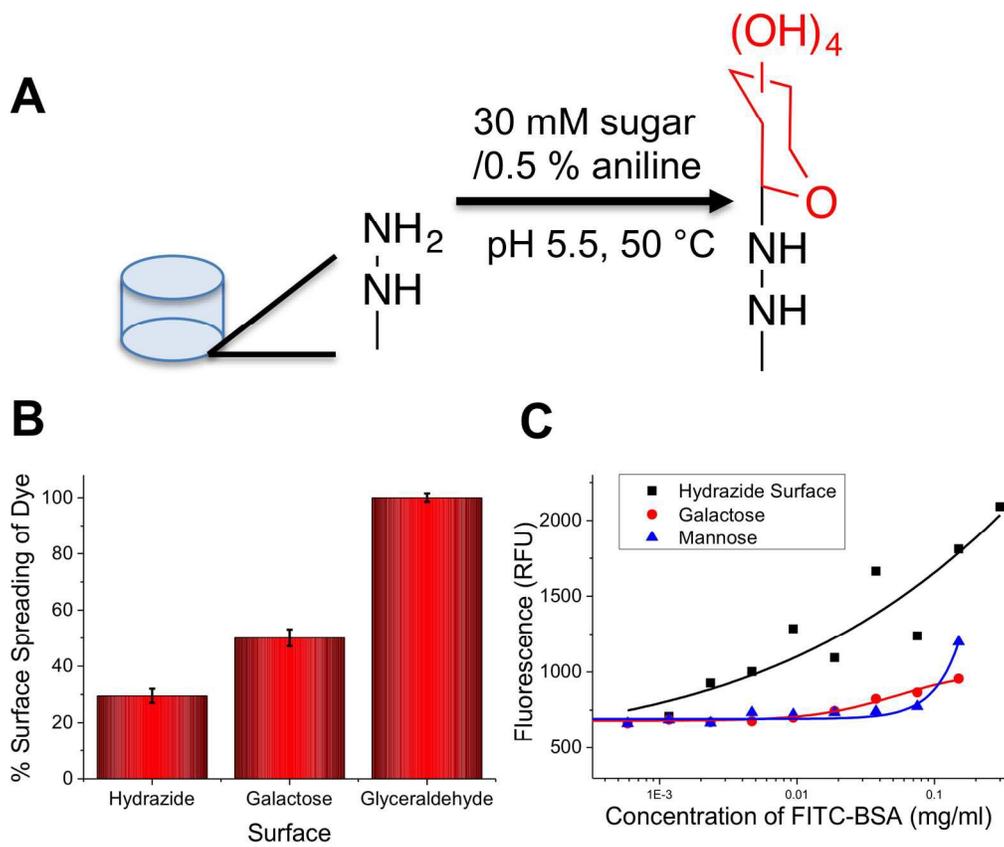
Acknowledgements

Notes and references

a Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL UK.
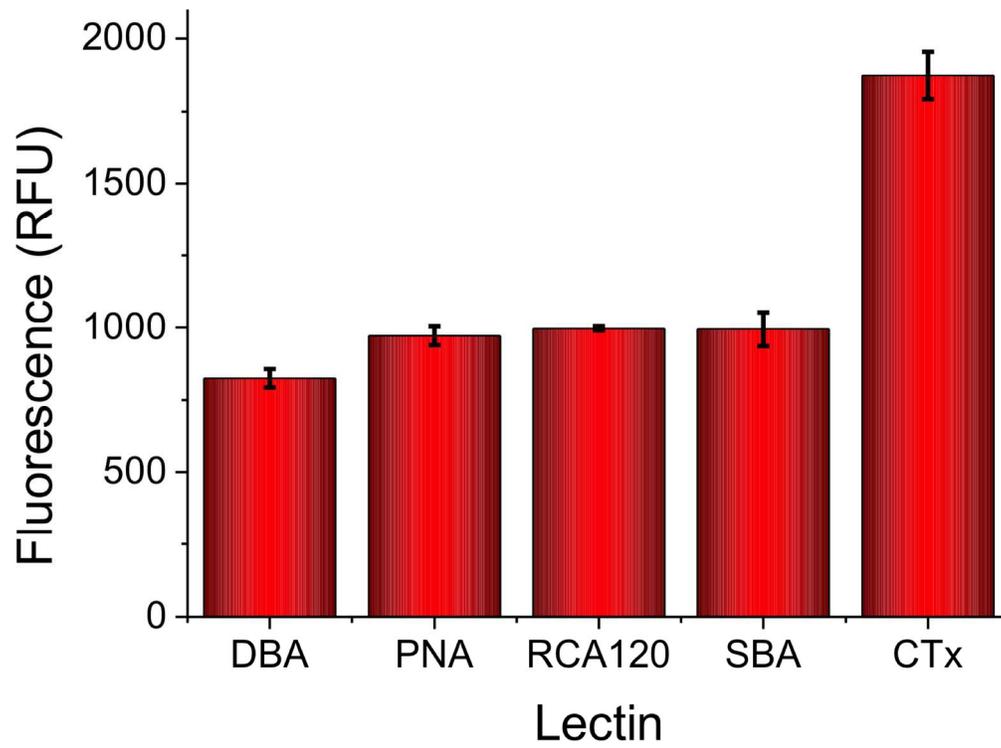*Corresponding author email; m.i.gibson@warwick.ac.uk
Electronic Supplementary Information (ESI) is available: this includes protein preparation, surface functionaliszation and LDA anlaysis. See DOI: 10.1039/b000000x/

[1] C. R. Bertozzi and L. L. Kiessling, *Science* **2001**, *291*, 2357-2364.

[2] M. W. Jones, L. Otten, S. J. Richards, R. Lowery, D. J. Phillips, D. M. Haddleton and M. I. Gibson, *Chemical Science* **2014**, *5*, 1611-1616.

[3] S.-J. Richards and M. I. Gibson, *ACS Macro Letters* **2014**, *3*, 1004-1008.

[4] M. J. Marin, A. Rashid, M. Rejzek, S. A. Fairhurst, S. A. Wharton, S. R. Martin, J. W. McCauley, T. Wileman, R. A. Field and D. A. Russell, *Org. Biomol. Chem* **2013**, *11*, 7101-7107.

[5] L. L. Kiessling, J. E. Gestwicki and L. E. Strong, *Angew. Chem. Int. Ed.* **2006**, *45*, 2348-2368.

[6] A. Bernardi, J. Jimenez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K.-E. Jaeger, M. Lahmann, T. K. Lindhorst, M. Marradi, P. Messner, A. Molinaro, P. V. Murphy, C. Nativi, S. Oscarson, S. Penades, F. Peri, R. J. Pieters, O. Renaudet, J.-L. Reymond, B. Richichi, J. Rojo, F. Sansone, C. Schaffer, W. B. Turnbull, T. Velasco-Torrijos, S. Vidal, S. Vincent, T. Wennekes, H. Zuilhof and A. Imberty, *Chemical Society Reviews* **2013**, *42*, 4709-4727.

[7] S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angewandte Chemie International Edition* **2012**, *51*, 7812-7816.

[8] J. L. Jimenez Blanco, C. Ortiz Mellet and J. M. Garcia Fernandez, *Chemical Society Reviews* **2013**, *42*, 4518-4531.
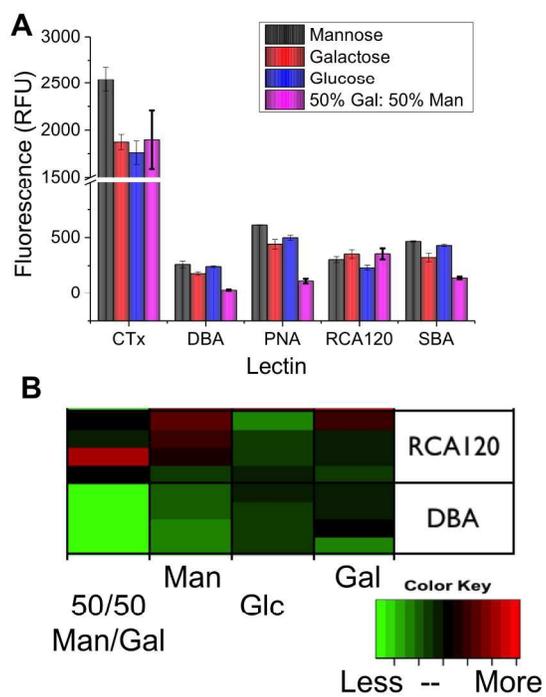
[9] H. Feinberg, D. A. Mitchell, K. Drickamer and W. I. Weis, *Science* **2001**, *294*, 2163-2166.

[10] http://www.functionalglycomics.org.

[11] A. Audfray, A. Varrot and A. Imberty, *Comptes Rendus Chimie* **2013**, *16*, 482-490.

[12] J. M. Lord, L. M. Roberts and J. D. Robertus, *The FASEB Journal* **1994**, *8*, 201-208.

[13] X. Zeng, C. S. Andrade, M. L. Oliveira and X.-L. Sun, *Anal Bioanal Chem* **2012**, *402*, 3161-3176.

[14] Y. Kim and A. Varki, *Glycoconj J* **1997**, *14*, 569-576.

[15] E. Hosoi, *The Journal of Medical Investigation* **2008**, *55*, 174-182.

[16] D. J. Anstee, *The relationship between blood groups and disease*, **2010**, p. 4635-4643.

[17] J. A. Rowe, D. H. Opi and T. N. Williams, *Current opinion in hematology* **2009**, *16*, 480-487.

[18] O. R. Miranda, B. Creran and V. M. Rotello, *Current Opinion in Chemical Biology* **2010**, *14*, 728-736.

[19] C.-C. You, O. R. Miranda, B. Gider, P. S. Ghosh, I.-B. Kim, B. Erdogan, S. A. Krovi, U. H. F. Bunz and V. M. Rotello, *Nat Nano* **2007**, *2*, 318-323.

[20] A. Bajaj, O. R. Miranda, I.-B. Kim, R. L. Phillips, D. J. Jerry, U. H. F. Bunz and V. M. Rotello, *Proceedings of the National Academy of Sciences* **2009**, *106*, 10912-10916.

[21] X. Li, H. Kong, R. Mout, K. Saha, D. F. Moyano, S. M. Robinson, S. Rana, X. Zhang, M. A. Riley and V. M. Rotello, *ACS Nano* **2014**.

[22] K. El-Boubbou, D. C. Zhu, C. Vasileiou, B. Borhan, D. Prosperi, W. Li and X. Huang, *Journal of the American Chemical Society* **2010**, *132*, 4490-4499.

[23] H. S. N. Jayawardena, X. Wang and M. Yan, *Analytical Chemistry* **2013**, *85*, 10277-10281.

[24] K. Godula and C. R. Bertozzi, *J. Am. Chem. Soc.* **2010**, *132*, 9963-9965.

A

30 mM sugar
/0.5 % aniline

pH 5.5, 50 °C

NH$_2$
NH

(OH)$_4$

O
NH
NH

B

% Surface Spreading of Dye

100
80
60
40
20
0

Hydrazide   Galactose   Glyceraldehyde

Surface

C

Fluorescence (RFU)

2000
1500
1000
500

■ Hydrazide Surface
● Galactose
▲ Mannose

1E-3   0.01   0.1
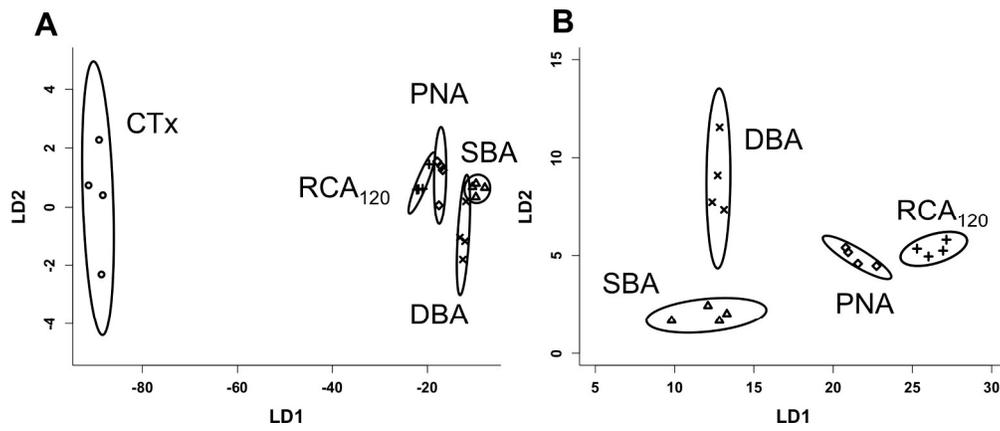
Concentration of FITC-BSA (mg/ml)

678x564mm (72 x 72 DPI)

750x551mm (72 x 72 DPI)

881x661mm (72 x 72 DPI)

847x352mm (72 x 72 DPI)