Introduction

Glycans play important roles inside eukaryotic cells by binding to proteins and lipids, and they are also found in the extracellular space between cells [1]. Glycans can be grouped into two classes; linear sugars and polysaccharides. The polysaccharides consist of repeating pyranose monosaccharide rings and branched sugars, which are formed by linking various monosaccharide units [2]. Through non-covalent interactions with lectins, glycans control biochemical reactions by engaging in various biological processes such as development [3,4], coagulation [5] and response to infection by bacterial and viral agents [6]. The size of the cellular glycome is believed to be in range of 100000–500000 glycans [7]. This large size of glycomic contents could be attributed to the combinatorial aspect that oligosaccharide chains come in either linear or branched form, monosaccharide building blocks are either in α or in β anomeric configurations and monosaccharides can be linked via various carbon atoms in their sugar rings [8]. Using the complexity of the glycome, cells adopt to encode a massive amount of biological information, and it is a great challenge to decode this hidden information to understand the biology of lectins and their interactions with carbohydrates.

Protein-carbohydrate interactions are involved in a variety of biological and biochemical processes, and, recently, attempts to understand the molecular basis of such interactions have appeared [9]. Traditional methods to probe glycan–protein recognition events include X-ray crystallography, NMR spectroscopy, the hemagglutination inhibition assay [10], enzyme-linked lectin assay [11], surface plasmon resonance [12] and isothermal titration calorimetry [13]. Although these methods have been successfully applied to elucidate the details of carbohydrate–protein interactions, they are rather labor intensive and require large amounts of carbohydrate samples. These shortcomings make the aforementioned traditional approaches unsuitable as high-throughput analytic methods [14]. On the other hand, recently, many computational methods have been suggested to study protein carbohydrate interactions [15–21].

Conventional methods for carbohydrate ligand detection are often cumbersome and we need sensitive and high-throughput technologies that can analyze carbohydrate–protein interactions in order to discover and differentiate oligosaccharide sequences interacting with carbohydrate binding proteins [8]. Carbohydrate micro-array based technology can serve as an appropriate method [22–25]. However, at present, one of the biggest limiting factors in utilizing the complete potential of the glycan microarray data is the lack of efficient analysis tools to extract relevant information.

For complete utilization of a glycan microarray data, we need a systematic computational method [26]. Large quantities of data are generated from the analysis of the Consortium for Functional Glycomics (CFG) glycan microarray [27]. Also, predicting the glycan-binding specificity or binding motif can be a time consuming step of scrutinizing and evaluating the linear sequences of monosaccharides in glycans [27]. The CFG offers glycan microarray data for various lectins (both plant and animal origin) and glycan binding antibodies. Recently computational methods have been developed for analyzing the glycan-binding specificity...
from glycan array data such as the motif-segregation method [26] and the outlier motif analysis (OMA) method [28].

In this work, we have developed a method to group various plant lectins and their interacting carbohydrates by the community detection analysis of a lectin-glycan network generated by the glycan microarray data from CFG. The lectin-glycan network consists of 1119 nodes (lectins and glycans) and 16769 edges (interactions). From this network, we have identified 3 lectins having large degrees of connectivity playing the roles of hubs. Additionally, we compared the results of our community detection method with other well known clustering algorithms. We show that our method outperforms existing clustering methods in terms of both modularity score as well as the number of statistically significant (p-value ≤ 0.05) glycan specific lectin groups. We propose that this study can reveal a global organization of lectin-glycan interactions, and help to identify strongly correlated lectin and glycan clusters.

Methodology

Data Generation

A total of 786 glycan array files for plant lectins were downloaded using a custom made script from Consortium for Functional Glycomics (CFG) as of Dec 2013. CFG provides extensive glycomics resources so that one can explore functions of glycans and glycan-binding proteins that play important roles in human health and disease [http://www.functionalglycomics.org/static/consortium/consortium.shtml]. All of these 786 files were further processed into a single input file, which consists of rows of protein-carbohydrate pairs. Three datasets were generated by filtering the protein-carbohydrate pairs using the cutoff values of relative fluorescence units (RFU) 5000, 10000 and 20000. These three datasets were used for network construction and their community detection. Figure 1 shows the histogram of the RFU values collected from 786 glycan array files. The data corresponding to RFU larger than 5000 constitutes only about 3.5% of the whole data. All the data is available to researchers upon request.

Network Construction

To perform a systematic analysis of protein-carbohydrate interaction, we have constructed a bipartite network, where unweighted edges are assigned between proteins and carbohydrates. Each node represents a lectin or a glycan and its identity is indicated by its array ID or glycan ID at a given condition. A glycan array ID represents a specific protein under a specific condition. Therefore, two different nodes in the network may represent two different concentrations of a protein in the glycan array experiment. The strength of a lectin-glycan interaction is represented by its RFU value and three networks are generated using three cutoff values of RFU of 5000, 10000 and 20000.

Community Detection of a Network

We have identified the community structure of the lectin-glycan network by using the Mod-CSA method, which is a highly effective modularity optimization method [29,30,31]. The modularity is a widely used measure to determine the community structures of various networks. From a given community structure it measures the difference between the number of inter-community edges and its expected value from a randomly re-wired counterpart preserving the degrees of nodes. Modularity (Q) is defined as:

\[
Q = \sum_{i=1}^{N_c} \left( \frac{l_i}{M} - \frac{D_i}{2M} \right)^2,
\]

where \(M\) is the total number of edges in the network, \(N_c\) is the number of communities, \(l_i\) is the number of edges within community \(i\) and \(D_i\) is the sum of degrees of nodes in community \(i\). The value of \(Q\) ranges between −1 and 1 and it becomes close to 1 for a highly modular community structure and 0 for a random community structure [32].

Network Visualization and Comparison with other Clustering Methods

Three lectin glycan array networks constructed in this study were exported to the Cytoscape 2.8.2, a bioinformatics package for biological network visualization and data integration [33]. To compare our clustering method with other widely used network clustering algorithms such as MCL [34,35], MCODE [36] and greedy algorithm [32], we have used clusterMaker [37] and GLay plugins [38], a multi-algorithm clustering plugins for Cytoscape.

Enrichment of Glycan-Specific Proteins

Enriched glycan-specific lectins within each cluster were investigated by annotating each lectin with a predetermined glycan binding specificity. Reported specificities of various lectins were extracted from literature [39,40] and Uniprot database [41] as summarized in Table 1. The full list of all 513 protein nodes used in this study with annotations (wherever possible) are listed in Table S1.

The enrichment of glycan-specificities of lectins in each cluster was assessed by calculating the hypergeometric p-value. The p-value corresponds to the probability that a given lectin cluster sharing the same glycan-specificity can be obtained by chance. The p-value was calculated as follows:

\[
P = 1 - \sum_{k=0}^{K-1} \frac{K}{k} \frac{N-K}{n-k} \frac{N}{n},
\]

where \(N\) is the total number of lectins in the network, \(K\) is the number of all lectins having a particular glycan-specificity, and \(k\) is the number of lectins having the particular glycan-specificity in a cluster with the size of \(n\).
Table 1. List of glycan binding specificities of lectins investigated in this study is shown. Specificities are collected from literature and uniprot database.

| S. No. | Protein Name                      | Reported Specificity                                      |
|--------|-----------------------------------|----------------------------------------------------------|
| 1.     | Pokeweed Agglutinin               | (GlcNAcb1-4)n                                            |
| 2.     | Datura Stramonium Lectin          | (GlcNAcb1-4n, Galb1-4GlcNAc)                             |
| 3.     | Soybean Agglutinin                | a- or b-linked terminal GalNAc, GalNAca1-3Gal            |
| 4.     | LBA Lima Bean Agglutinin/LBL      | a-D-GalNAc                                               |
| 5.     | Griffonia Simplicifolia Lectin I  | a-Linked Gal                                              |
| 6.     | Agglutinin                        | a-Linked terminal GalNAc                                  |
| 7.     | Psophocarpus tetragonolobus Agglutinin/Basic agglutinin | a-Linked terminal GalNAc                                  |
| 8.     | Psophocarpus Tetragonolobus Lectin I | a-Linked terminal GalNAc                                  |
| 9.     | Vicia Villosa Lectin (VVL)        | a-Linked terminal GalNAc, GalNAca1-3Gal                  |
| 10.    | Griffonia simplicifolia II/Insecticidal | Agalactosylated tri/tetra antennary glycans, GlcNAc          |
| 11.    | Phaseolus vulgaris Erythrogglutinin/ | Bi-antennary complex-type N-glycan with outer Gal and bisecting GlcNAc |
| 12.    | Wheat Germ Agglutinin (WGA)      | Chitin oligomers, Sia                                     |
| 13.    | Laburnum alpinum Agglutinin/     | Di-N-acetylcibitobiose specific lectin.                  |
| 14.    | Ulex europaueus Agglutinin OR    | Di-N-acetylcibitobiose specific lectin.                  |
| 15.    | Trichosanthes japonica Agglutinin II | Fuca1-2Galb1- -> or GalNAcb1- -> groups at their nonreducing terminals |
| 16.    | Cholera Toxin B                   | Fuca1-2Galb1-3GalNAc1-4(Neu5Aca2-3)Galb1-4Glcb OR Galb1-3GalNAc1-4(Neu5Aca2-3)Galb1-4Glcb |
| 17.    | Ulex Europaueus Agglutinin OR    | Fuca1-2Galb1-4GlcNAc                                       |
| 18.    | Lotus Tetragonolobus Lectin/     | Fuca1-3(Galb1-4)GlcNAc, Fuca1-2Galb1-4GlcNAc              |
| 19.    | Aspergillus oryzae Lectin        | Fuca1-6GlcNAc (core fucose)                               |
| 20.    | Lens Culinaris Agglutinin        | Fuca1-6GlcNAc, a-D-Glc, a-D-Man                           |
| 21.    | Pisum Sativum Agglutinin         | Fuca1-6GlcNAc, a-D-Glc, a-D-Man                           |
| 22.    | Aleuria Aurantia Lectin AAL      | Fuca1-6GlcNAc, Fuca1-3(Galb1-4)GlcNAc                     |
| 23.    | Pseudomonas aeruginosa lectin/   | Fucose Anywhere                                           |
| 24.    | Psofoporus Tetragonolobus Lectins | Fucose binding lectin                                     |
| 25.    | Fucose-binding lectin protein     | Fucose binding lectin                                     |
| 26.    | Euonymus europaeus Agglutinin    | Galb1-3Gal, blood group B antigen                         |
| 27.    | Cytisus sscoparius Agglutinin    | Galactose binding lectin                                  |
| 28.    | Discoidin-2                      | Galactose binding lectin                                  |
| 29.    | Polyporus Squamosus Lectin       | Galactose binding lectin                                  |
| 30.    | Discoidin-1 subunit B/C          | Galactose- and N-acetylgalactosamine-binding             |
| 31.    | SRL- strong binding to di-saccharide | Galb1-3GalNAca-a- similar to Agaricus bisporus lectin     |
| 32.    | Agaricus bisporus Agglutinin     | Galb1-3GalNAc                                             |
| 33.    | Amaranthus Caudatus Lectin       | Galb1-3GalNAc                                             |
| 34.    | Galactose-binding lectin (Agglutinin PNA) | Galb1-3GalNAc                                             |
| 35.    | Jacalin/Agglutinin alpha chain   | Galb1-3GalNAc, GalNAc                                     |
| 36.    | Bauhinia Purpurea Lectin         | Galb1-3GalNAc, GalNAc                                     |
| 37.    | Maclura Pomifera Lectin/Agglutinin alpha chain/MPA | Galb1-3GalNAc, GalNAc                                     |
| 38.    | Erythrina crista-galli Lectin    | Galb1-4GlcNAc                                             |
| 39.    | Ricinus Communis Agglutinin I    | Galb1-4GlcNAc                                             |
| 40.    | Dolichos biflorus Agglutinin/Seed lectin subunit I | GalNAc1-3GalNAc, blood group A antigen                     |
| 41.    | Wisteria floribunda Agglutinin   | GalNAc1-4GlcNAc, Galb1-3(6)GalNAc                          |
| 42.    | Marasmium oraeas agglutinin      | Galb1(3,1,3)Gal                                            |
Enrichment analysis was also attempted by using DAVID functional annotation cluster tool [http://david.abcc.ncifcrf.gov/home.jsp], which did not yield any statistical significant clusters. We then manually searched each lectin in InterPro database [42] but only 8 unique GO terms such as chitin-binding, carbohydrate-binding, protein binding, endopeptidase inhibitor activity, etc, were retrieved. However, these GO terms are too general to signify any detailed glycan binding specificities of corresponding proteins.
lectins. Therefore, in this study, the enrichment analysis for each cluster was performed based on the annotations listed in Table 1. Only those clusters with at least 10 protein nodes were analyzed for statistical significance.

Identification of Hub Proteins

In general, biological networks possess the scale-free property [43] in which only a few nodes in the network have many connections serving as hubs in the network. Hub proteins were identified by calculating the node degree distribution [44] by using the NetworkAnalyzer plugin of Cytoscape. Top three highest degree protein nodes were assigned as hubs (see Figure 2).

Results and Discussion

We constructed three lectin-glycan interaction networks by using the plant lectin-glycan micro array data filtered by three RFU cut-offs. The network where the interactions were filtered by RFU <5000 consists of 1119 nodes (513 proteins and 606 carbohydrates) and 16769 edges. Similarly, the second network filtered by RFU <10000 has 1035 nodes and 12169 edges, and the third one (filtered by RFU <20000) consists of 901 nodes and 8042 edges. Since the first network has the maximum number of nodes and edges, and shows more statistically significant glycan specific groups (discussed later) than the other two networks, the results specified henceforth represent the first network if not specifically indicated. The first network is shown in Figure 2, where proteins are represented as diamonds and glycans as circles and the interactions between them are represented as edges.

The network representation enables a quick visual inspection of the glycans bound to a lectin of interest. Additionally, in order to identify hub lectins from the lectin-glycan array, the node degree distribution of the network was calculated and is shown in Figure 2. In an interaction network, proteins that interact with a large number of partners are considered as hubs [45], and are essential components of biological networks [46]. The definition of the hub node is rather subjective, but based on the observation of the biggest gap between the 3rd and 4th largest degree nodes in Figure 2, we assigned hub proteins as those three with degree larger than 220. The 3 hubs are Phloem Proteinc2 (PP2A1) from Arabidopsis thaliana, wheat germ agglutinin (WGA) from Triticum vulgaris (wheat), and Ricinus communis agglutinin (RCA) from Ricinus communis (castor bean).

By using the Mod-CSA method, the lectin-glycan network is clustered into 4 modules (communities), which are represented by separate colors in Figure 3. The largest module consists of 168 protein nodes and 215 glycan nodes, and the smallest community contains 90 protein nodes and 133 glycan nodes.

To validate the lectin-glycan interaction network and its detected community-structure, we investigated the binding specificities of the first neighbors of two plant lectins, Sambucus nigra agglutinin (SNA) and concanavalin A (ConA) whose glycan binding specificities are well known. The first lectin is a well-characterized plant lectin, elderberry bark agglutinin from Sambucus nigra, which is known to recognize the Neu5Acα2-6Gal linkage [47]. The second one is concanavalin A (ConA), which is known to have specificity for mannose sugars [48,49,50]. Proper categorization of the specificities of glycan-binding proteins plays a significant role in understanding protein-glycan interactions and utilizing glycan-binding proteins as analytical reagents.

Binding Specificities of SNA

It is well known that some plants contain more than one lectin with different sugar binding specificities [51]. The bark of the elderberry (Sambucus nigra) has two lectins SNA-I and SNA-II with different glycan binding specificities. Sambucus nigra agglutinin I (SNA-I), is the first lectin identified from the elderberry bark which has been conventionally employed to recognize Neu5Acα2-6Gal [47] or Neu5Acα2-6Galβ1-4GlcNAc sequence [27]. SNA-I is composed of two polypeptides, namely chain A of 33 kDa with enzymatic activity, and chain B of 35 kDa with carbohydrate-binding activity [52]. Molecular modeling studies have indicated that the overall structure of SNA-I is quite similar to that of Ricin [53] and SNA-I belongs to the group of type 2 ribosome-inactivating proteins [52]. SNA-II is the second lectin isolated from the elderberry bark tissue, and it exhibits high affinity for glycoconjugates and Type 14 pneumococcal polysaccharides having multiple terminal D-Gal groups [51]. SNA-II consists of two identical carbohydrate-binding B-chains [51,52].

In the current lectin glycan array network, nineteen nodes represent both SNA-I and SNA-II lectins. Out of these nineteen SNA nodes, fifteen SNA-I nodes are from community 1 (1000180, 1000181, 1000183, 1000184 and 1000725), and community 3 (1002793, 1004421, 1004422, 1004701, 1004702, 1004703, 1004704, 1004705, 1004706 and 1004780). Similarly, SNA-II is represented by four nodes (1004707, 1004708, 1004709 and 1004710) enriched in community 3.

Figure 2. The node degree distribution of the lectin-glycan network is shown. We observe a large gap between 3 hub nodes and the other nodes. The degree distribution was plotted using plotly [https://plot.ly/plot].

doi:10.1371/journal.pone.0095480.g002
The 10 SNA-I nodes in community 3 show specificity for complex-type biantennary N-glycans (Table 2A). From this table we observe that almost all of the interacting glycans possess the determinant Neu5Acα2-6Gal or Neu5Acα2-6Galβ1-4GlcNAc (shown by bold text in the table). Another interesting point to notice is that the glycans 527 and 479 exhibit low RFU values in Table 2. This could be due to the fact that these glycans contain Neu5Acα2-3 sequence, which is known to decrease the binding of SNA [27]. On the other hand, 316 (Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manβ1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manβ1-4GlcNAcβ1-4GlcNAcβ1-6Manβ1-4GlcNAcβ1-4GlcNAcβ1-6) contains two sequences, one (Neu5Acα2-6Galβ1-4GlcNAc) increasing the binding and the other (Neu5Acα2-3) decreasing the binding.

Compared to SNA-I nodes in community 3, five SNA-I nodes in community 1 (1000180, 1000181, 1000183, 1000184 and 1000725) interact with a smaller number of complex glycans (see Table 2B). Top 3 glycans possess either Neu5Acα2-6Gal or Neu5Acα2-6Galβ1-4GlcNAc and show RFU values greater than 40000. Two glycans from the second half of the table (glycans 60 and 59) show lower values of RFU because of the presence of the Neu5Acα2-3Gal sequence, which is known to decrease glycan binding. All these results are consistent with existing studies on the SNA specificity [27].

The 4 SNA-II nodes (1004707, 1004708, 1004709 and 1004710) in community 3 show preference for mainly mannose glycans or terminal GlcNAcb1-4GlcNAcb. Only two glycans (347 and 349) possess the determinant of Neu5Acα2-6Galβ1-4GlcNAc (Table 2C). In general, SNA-II is known to be Gal/ GalNAc specific and is precipitated by glycoproteins, which consist of terminal GalNAc oligosaccharide chains [51]. Specifically, it shows higher affinity for D-GalNAc- and terminal N-acetyl-D-galactosaminyI disaccharides as compared to D-Gal. Conversely, the affinity exhibited by SNA-I for D-Gal and D-GalNAc- is identical [51]. However, SNA-I recognizes Neu5Acα2-6Gal [47] or Neu5Acα2-6Galβ1-4GlcNAc glycan sequence [27] with high specificity. Despite the differences in their glycan binding specificities, SNA-I and SNA-II share some similarities. For example, both lectins contain similar amino acid composition, while SNA-II contains more asparagine/aspartic acid, glycine and methionine residues [51]. Additionally, the carbohydrate-binding B-chains of both lectins show caspase-dependent apoptosis in different insect cell lines [52]. Considering their characteristic glycan binding specificities, SNA-I and SNA-II may play different functional roles in plants.

### Binding Specificities of ConA

Concanavalin A (ConA) binds to a variety of eukaryotic cells through specific interactions with saccharide-containing cellular receptors, and has been widely used as a molecular probe in studies of cell membrane dynamics and cell division [54]. ConA typically binds to glucosyl and mannosyl residues at the non-reducing termini of oligo- or polysaccharides [48,49] and it can also bind to non-terminal mannosyl residues [50]. The current network contains sixteen nodes of ConA (1000158 and 1000165 in community 1; 1000356 and 1000699 in community 2; and 1004459, 1004460, 1004461, 1004462, 1004464, 1004465, 1004466, 1004467, 1004468, 1002791, 1004412 and 1004413 in community 3) which mainly interacts with mannose containing glycans.
Table 2. Three types of complex glycans for SNA proteins are listed.

| Glycan No. | Glycan Name | Avg. RFU |
|------------|-------------|----------|
| 268        | Neu5Ac2-6Gal| 51134    |
| 467        | Neu5Ac2-6Gal| 48246    |
| 465        | Neu5Ac2-6Gal| 43978    |
| 346        | Neu5Ac2-6Gal| 43812    |
| 327        | Neu5Ac2-6Gal| 41668    |
| 320        | Neu5Ac2-6Gal| 41588    |
| 302        | Neu5Ac2-6Gal| 41500    |
| 483        | Neu5Ac2-6Gal| 41106    |
| 55         | Neu5Ac2-6Gal| 40488    |
| 348        | Neu5Ac2-6Gal| 39574    |
| 606        | Neu5Ac2-6Gal| 39290    |
| 482        | Neu5Ac2-6Gal| 39202    |
| 57         | Neu5Ac2-6Gal| 38592    |
| 56         | Neu5Ac2-6Gal| 37417    |
| 609        | Neu5Ac2-6Gal| 36652    |
| 457        | Neu5Ac2-6Gal| 36616    |
| 325        | Neu5Ac2-6Gal| 36221    |
| 314        | Neu5Ac2-6Gal| 35848    |
| 503        | Neu5Ac2-6Gal| 33405    |
| 298        | Neu5Ac2-6Gal| 32632    |
| 287        | Neu5Ac2-6Gal| 31718    |
| 354        | Neu5Ac2-6Gal| 30544    |
| 557        | Neu5Ac2-6Gal| 28273    |
| A) Glycan No. | Glycan Name                                                                 | Avg. RFU |
|-------------|------------------------------------------------------------------------------|----------|
| 366         | Fuc(1→4)(Gal(1→3)GlcNAc(1→2Man)1-6 (Fuc(1→4)(Gal(1→3)GlcNAc(1→2Man)1-3) | 27993    |
|             | Man(1→4)GlcNAc(1→4(Fuc(1→6)GlcNAc)1-SP22)                                    |          |
| 319         | Neu5Ac(2→6)Gal(1→4)GlcNAc(1→2Man)1-6 (Neu5Ac(2→3)Gal(1→4)GlcNAc(1→2Man)1-3) | 27611    |
|             | Man(1→4)GlcNAc(1→4GlcNAc)1-SP12                                              |          |
| 54          | Neu5Ac(2→6)Gal(1→4)GlcNAc(1→2Man)1-3 (Neu5Ac(2→6)Gal(1→4)GlcNAc(1→2Man)1-6) | 27447    |
|             | Man(1→4)GlcNAc(1→4GlcNAc)1-SP12                                              |          |
| 321         | Neu5Ac(2→6)Gal(1→4)GlcNAc(1→2Man)1-3 (Neu5Ac(2→3)Gal(1→4)GlcNAc(1→2Man)1-3) | 26481    |
|             | Man(1→4)GlcNAc(1→4GlcNAc)1-SP12                                              |          |
| 274         | Neu5Ac(2→6)Gal(1→4Glc)1-SP8                                                 | 25380    |
| 53          | Neu5Ac(2→6)Gal(1→4)GlcNAc(1→2Man)1-3 (Neu5Ac(2→6)Gal(1→4)GlcNAc(1→2Man)1-6) | 25345    |
|             | Man(1→4GlcNAc(1→4GlcNAc)1-SP12                                              |          |
| 48          | [9NAc]Neu5Ac(2→6)Gal(1→4GlcNAc)1-SP8                                       | 21935    |
| 488         | Neu5Ac(2→6)Gal(1→4GlcNAc)1-6 (Fuc(1→2)Gal(1→4(Fuc(1→3)GlcNAc(1→3) | 21783    |
|             | Gal(1→4Glc)1-SP21                                                           |          |
| 328         | Neu5Ac(2→6)Gal(1→4GlcNAc(1→3Gal)1-3GalNAc(1→3) | 21014    |
|             | Man(1→4)GlcNAc(1→4GlcNAc(1→3)Man1-1)                                        |          |
| 324         | Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-3 (Neu5Ac(2→3)Gal(1→4GlcNAc(1→2Man)1-6) | 19830    |
|             | Man(1→4GlcNAc(1→4GlcNAc)1-SP12                                              |          |
| 58          | Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-6 (Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-3) | 18639    |
|             | Man(1→4GlcNAc(1→4GlcNAc)1-SP24                                              |          |
| 347         | Man1-6 Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-3) MAN1-4GlcNAc(1→4GlcNAc)1-SP12 | 16329    |
| 464         | Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-6 (GlcNAc(1→4)Neu5Ac(2→6)Gal(1→4 | 16237    |
|             | GlcNAc(1→2Man)1-3)Man1-4GlcNAc(1→4GlcNAc)1-4GlcNAc)1-SP21                   |          |
| 466         | Neu5Ac(2→6)Gal(1→4GlcNAc(1→1-6 Neu5Ac(2→6)Gal(1→4GlcNAc(1→1-2Man)1-6) | 13858    |
|             | Man1-6(GlcNAc(1→4)Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-3)Man1-4GlcNAc(1→4GlcNAc)1-4GlcNAc)1-SP21 |          |
| 409         | Neu5Ac2(1→4)(Gal(1→3)GlcNAc(1→2Man)1-3)Man1-4GlcNAc(1→4GlcNAc)1-SP21 | 12386    |
| 270         | Neu5Ac(2→6)Gal(1→4GlcNAc)1-3GalNAc(1→3)Man1-4GlcNAc(1→4GlcNAc)1-SP8         | 11415    |
| 317         | Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-3 (Gal(1→4GlcNAc(1→2Man)1-6)Man1-4GlcNAc(1→4GlcNAc)1-SP12 | 11124    |
| 360         | KDN(1→3)Gal(1→3)GalNAc(1→3)Man1-4GlcNAc(1→4GlcNAc)1-SP12            | 11019    |
| 485         | Man1-6 Man1-3 (Man1-4GlcNAc(1→4(Fuc(1→6)GlcNAc)1-SP19) | 10968    |
| 427         | Fuc(1→2)Gal(1→3)GlcNAc(1→2Man)1-6(Fuc(1→6)GlcNAc(1→2Man)1-3)Man1-4GlcNAc(1→4(Fuc(1→6)GlcNAc)1-SP22 | 10833    |
| 458         | Neu5Ac(2→6)Gal(1→4GlcNAc)1-6 (Fuc(1→2)Gal(1→3)GlcNAc(1→3GalNAc(1→3)GlcNAc)1-SP21 | 10202    |
| 52          | Neu5Ac(2→6)Gal(1→4GlcNAc)1-2Man(1→3) (Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-6)Man1-4GlcNAc(1→4GlcNAc)1-SP8 | 9467    |
| 309         | Neu5Ac(2→6)Gal(1→4GlcNAc(1→2Man)1-6 (GlcNAc(1→2Man)1-3)Man1-4GlcNAc(1→4GlcNAc)1-SP12 | 9381    |
| 376         | Neu5Ac(2→6)Gal(1→4GlcNAc)1-3GalNAc(1→3)Man1-4GlcNAc(1→4GlcNAc)1-SP14 | 8974    |
### Table 2. Cont.

| Glycan No. | Glycan Name                                                                 | Avg. RFU |
|------------|-----------------------------------------------------------------------------|----------|
| 521        | NeuSAc2-6Gal|j1-4GlcNAcj1-2Man-Sp0                                                        | 8470     |
| 313        | Neu5Acz2-3Gal|j1-4GlcNAcj1-2Manz1-3                                                         | 8322     |
|            | /NeuSAc2-6Gal|j1-4GlcNAcj1-2Manz1-6                                                          |          |
|            | Manz1-4GlcNAcj1-4GlcNAcj-Sp12                                                  |          |
| 316        | NeuSAc2-3Gal|j1-4GlcNACj1-2Manz1-3                                                          | 8189     |
|            | /NeuSAc2-6Gal|j1-4GlcNAcj1-2Manz1-6                                                          |          |
|            | Manz1-4GlcNAcj1-4GlcNAcj-Sp12                                                  |          |
| 353        | GlnNACj1-2Mans1-6(GlcNACj1-2Manz1-3)                                          | 7768     |
|            | Manz1-4GlcNAcj1-4(Fucz1-6)GlcNACj-Sp22                                         |          |
| 527        | NeuSAc2-3Gal|j1-3GlcNAcj1-2Manz-Sp0                                                         | 6941     |
| 478        | NeuSAc2-6Gal|j1-4GlcNAcj1-6                                                                  | 6901     |
|            | (Galz1-3GlcNAcj1-3)Galz1-4GlcSp21                                              |          |
| 315        | NeuSAc2-6Gal|j1-4GlcNAcj1-2Manz1-3                                                          | 6606     |
|            | (GlcNACj1-2Manz1-3)Manz1-4GlcNAcj1-4GlcNAcj-Sp12                              |          |
| 358        | KDNaz2-6Gal|j1-4GlcNAc-Sp0                                                                  | 6532     |
| 333        | NeuSAc2-6Gal|j1-4GlcNAcj1-3Galz1-1-4GlcNAcj1-4GlcNAcj-Sp0                                  | 6339     |
| 349        | NeuSAc2-6Gal|j1-4GlcNAcj1-2Manz1-3                                                          | 6178     |
|            | 3Manz1-4GlcNAcj1-4GlcNAcj-Sp12                                                  |          |
| 607        | NeuSAc2-6Gal|j1-4GlcNAcj1-3Galz1-1-4GlcNAcj1-4GlcNAcj1-2Manz1-6                            | 6161     |
|            | (NeuSAc2-6Gal|j1-4GlcNAcj1-3Galz1-1-4GlcNAcj1-2Manz1-3)                                      |          |
|            | Manz1-4GlcNAcj1-4GlcNAcj-Sp12                                                  |          |
| 479        | Neu5Ac2-3Gal|j1-4GlcNAcj1-2Manz-Sp0                                                         | 6154     |
| 51         | NeuSAc2-6Gal|j1-4GlcNAcj1-2Manz1-3                                                          | 5582     |
|            | /NeuSAc2-6Gal|j1-4GlcNAcj1-2Manz1-6                                                          |          |
|            | Manz1-4GlcNAcj1-4GlcNAcj1-6(NLT)AVL                                            |          |
| 49         | Neu5,9Ac2a-6Gal|j1-4GlcNAcj-Sp8                                                       | 5207     |

### B)

| Glycan No. | Glycan Name                                                                                     | Avg. RFU |
|------------|--------------------------------------------------------------------------------------------------|----------|
| 2          | AGP-A (AGP ConA flowthrough)                                                                    | 52286.06 |
| 246        | NeuSAc2-6Gal|j1-4GlcNAcj-Sp8                                                                                   | 49625.09 |
| 263        | NeuSGc2-6Gal|j1-4GlcNAcj-Sp0                                                                                   | 48932.24 |
| 250        | NeuSAc2-6Gal|j1-4Glcj-Sp8                                                                                      | 48814.18 |
| 6          | Transferrin                                                                                       | 47533.26 |
| 248        | NeuSAc2-6Gal|j1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0                                                                 | 47165.41 |
| 42         | [6OSO3]Galz1-4Glcj-Sp0                                                                            | 34505.6  |
| 44         | [6OSO3]Galz1-4GlcNAcj-Sp8                                                                         | 32444.86 |
| 247        | NeuSAc2-6Gal|j1-4GlcNAcb1-3Galb1-4(Fucz1-3)GlcNAcb1-3Galb1-4(Fucz1-3)GlcNAcb-Sp8                             | 30612.85 |
| 45         | [6OSO3]Gald1-4(6OSO3)Glcj-Sp8                                                                      | 27055.41 |
| 245        | NeuSAc2-6Gal|j1-4GlcNAcj-Sp0                                                                                   | 26857.74 |
| 1          | Alpha1-acid glycoprotein (AGP)                                                                    | 25869.33 |
| 43         | [6OSO3]Galz1-4Glcj-Sp8                                                                            | 22740.63 |
| 86         | GalNAc1-3Galb-Sp8                                                                                 | 21300.14 |
| 20         | [3]GalNAc-Sp8                                                                                     | 20559.23 |
| 3          | AGP-B (AGP ConA bound)                                                                            | 17780.79 |
| 72         | Fucz1-2Gal|j1-4GlcNAcj-Sp8                                                                                   | 15937.51 |
### Table 2. Cont.

| Glycan No. | Glycan Name                                                                 | Avg. RFU |
|------------|------------------------------------------------------------------------------|----------|
| 70         | Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0                 | 14916.26 |
| 69         | Fuc1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0                                   | 13165.86 |
| 87         | GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp0                                       | 12866.52 |
| 242        | Neu5Ac-O6GalNAcb-Sp8                                                        | 12071.82 |
| 90         | GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0                                         | 11384.24 |
| 60         | Fuc1-2Galb1-3GalNAcb1-4(Neu5AcO2a-3)Galb1-4Glc-Sp9                          | 10906.21 |
| 120        | Galb1-3(Galb1-4GlcNAcb1-6)GalNAcb-Sp8                                       | 10546.55 |
| 150        | Galb1-4GlcNAcb1-6(Galb1-3)GalNAcb-Sp8                                        | 9937.06  |
| 251        | Neu5Ac-O6Galb1-3GalNAcb-Sp8                                                 | 9853.36  |
| 12         | Galb1-3(Neu5AcO2a-6)GalNAcb-Sp8                                             | 7693.01  |
| 10          | α-GalNAcb-Sp8                                                                | 6840.72  |
| 40         | [4OSO3]Galb1-3GalNAcb-Sp8                                                    | 6574.62  |
| 39         | [4OSO3]Galb1-4GlcNAcb-Sp0                                                    | 6514.83  |
| 241        | Neu5Ac-O6Galb1-3GalNAcb-Sp8                                                 | 6184.61  |
| 87         | GalNAcb1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8                                       | 5469.91  |
| 41          | 6+H2PO3Man−-Sp8                                                             | 5447.01  |
| 249        | Neu5Ac-O6Galb1-4Glc-SP0                                                      | 5434.93  |

| Glycan No. | Glycan Name                                                                 | Avg. RFU |
|------------|------------------------------------------------------------------------------|----------|
| 51         | Manb1-6(Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb1-3Manb1-4GlcNAcb1-4GlcNAcb-Sp13   | 38866    |
| 352        | Manb1-6(Manb1-4GlcNAcb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp12  | 37659    |
| 216        | Manb1-6(Manb1-3)Manb1-6(Manb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12       | 36933    |
| 347        | Manb1-6(Neu5AcO2a-6Galb1-4GlcNAcb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12  | 35539    |
| 212        | Manb1-2Manb1-6(Manb1-3)Manb1-6(Manb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12 | 35267    |
| 213        | Manb1-2Manb1-6(Manb1-2Manb1-3)Manb1-6(Manb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp12 | 18208   |
| 217        | Manb1-6(Manb1-3)Manb1-6(Manb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12       | 15856    |
| 485        | Manb1-6(Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb1-4(Fuca1-4GlcNAcb1-4GlcNAcb-Sp19) | 12002    |
| 211        | Manb1-6(Manb1-2Manb1-3)Manb1-6(Manb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12 | 10800   |
| 417        | Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Manb1-4GlcNAcb-Sp14                        | 10154    |
| 477        | Galb1-3GlcNAcb1-2Manb1-6(GlcNAcb1-4(Galb1-3GlcNAcb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21 | 7265    |
Community-Based Lectin-Glycan Interaction Network

All ConA interacting glycan nodes from community 1, 2 and 3 are shown in Table 3A, 3B and 3C, respectively. ConA interacting glycan nodes in community 1 are either mannose sugars or biantennary complex glycans such as transferrin and AGP-B. On the other hand, the ConA nodes in community 2 show preference for terminal glucose glycans.

In comparison to communities 1 and 2, the ConA nodes in community 3 show high preference for mannose containing sugars especially “N-glycan, high mannose” (Table 3C). These results agree with existing reports on ConA’s binding structure and specificity for mannose containing structures [55-57], in addition to the recognition of biantennary glycans, complex N-glycans [58] and terminal glucose [57].

Existing studies on SNA-I [47] and ConA [55-57] demonstrate the validity of the lectin-glycan interaction network and its detected community structure. Once a network is constructed, it is fairly easy to identify a lectin that explicitly binds to a certain glycan sequence by just selecting the lectin node of interest and its first neighbors in the network. The lectins in different communities show a dramatic difference in their glycan binding specificities. The current network-based approach should provide quick overall analysis and the use of glycan microarray data on the lectin-glycan interaction without time-consuming calculations.

Community Detection of the Lectin-glycan Interaction

We performed community detection of the lectin-glycan interaction network by using Mod-CSA [28], and compared the results with existing methods such as MCL [34,35], MCODE [36] and greedy algorithm [32,38]. The number of identified communities and the modularity values obtained by various community detection algorithms are shown in Table 4, Figure 4 and Figure 5.

From Table 4, Figure 4 & Figure 5a–d, it is clear that Mod-CSA [29] outperforms the other clustering methods in terms of the modularity score as well as the number of nodes left unclassified. The only method comparable to our modularity score of 0.37 obtained by Mod-CSA was the fast greedy algorithm [32,38] with a modularity score of 0.30. The algorithm recognizes clusters by repetitively eliminating edges from the network and then checks again which nodes are still connected [59]. The method detected 6 communities with the largest community containing 223 protein nodes and 298 glycan nodes (community 1) whereas the three smallest communities consist of either 4 nodes (community 4) or 3 nodes (community 5 & 6) only (see Figure 5b).

To compare the biological significance of modules (communities) obtained by Mod-CSA and by the greedy algorithm, we calculated the numbers of statistically meaningful enriched clusters of lectins that bind to the same specific glycan. The glycan binding specificity of each protein node was identified either from the literature or from Uniprot database as described in the methods section, and the significance of each glycan specific clusters was assessed by calculating its p-value (p ≤ 0.05). From Table 5, we observe that 44 statistically meaningful enriched clusters of lectins are identified with p-values ≤ 0.05. Whereas only 33 enriched clusters are identified by the greedy algorithm. This result suggests that many additional functionally related lectin clusters are identified by Mod-CSA, than detected by greedy algorithm.

For example, the greedy algorithm failed to identify 13 glycan specific lectin clusters (shown in bold in Table 5) that were identified by Mod-CSA. On the contrary, 3 glycan specific clusters (shown in italic bold in Table 5) were not detected by Mod-CSA, which are found by the greedy algorithm result. Specifically, the greedy algorithm failed to identify all fucose specific lectins, while Mod-CSA [29] successfully detected almost all fucose specific lectins and grouped them in community 1. Similarly, the greedy algorithm identified only five mannose related specificities in community 3, which is the major mannose binding community detected by greedy algorithm. However, Mod-CSA recognized eight mannose related specificities in community 1.

We compared our method with other popular clustering algorithms such as MCODE [36] and MCL [34,35]. MCODE method divided the network into a total of 23 clusters with the modularity score of −0.036. The largest cluster consists of 36 nodes whereas the smallest cluster contains only 4 nodes. However, only 3 clusters contain more than 10 protein nodes and they were further analyzed for enrichment of glycan specific lectin groups. The statistical analysis of these 3 clusters resulted in only 4 statistically meaningful lectin groups. From Figure 5c, we observe that a large number of single nodes (791) are not clustered into any groups. This is because MCODE identifies clusters of tightly connected nodes and does not intend to assign every node in the network to a cluster [59]. The main reason for this could be the fact that the MCODE algorithm is sensitive to noise in the network, particularly to false positive interactions [60]. Consequently, only a small number of strongly connected clusters are identified by MCODE and the rest of the nodes remain unclustered, which makes it hard to extract information from the network.

Table 2. Cont.

| Glycan No. | Glycan Name | Avg. RFU |
|-----------|-------------|----------|
| 50        | Man1-6(Man1-3Manj1-4GlcNAc)j1-4GlcNAcj1-4GlcNAcSp24 | 6298     |
| 349       | Neu5Ac2-6Galj1-4GlcNAcj1-2Man1-3Manj1-4GlcNAcj1-4GlcNAcSp24 | 6173     |
| 561       | Gala1-3Galj1-4GlcNAcj1-2Man1-6(Man1-4GlcNAc)Sp24 | 5565     |

A) Complex N-glycans for 10 SNA nodes (1002793, 1004421, 1004422, 1004780, 1004701, 1004702, 1004703, 1004704, 1004705 and 1004706) belonging to community 3, which is the major mannose binding community. B) Four SNA (SNA-II) nodes (1004707, 1004708, 1004709 and 1004710) in the community 3 show preference for mainly mannose glycans. Only two glycans (glycan 347 and 349) possess the determinant Neu5Ac2-6Galj1-4GlcNAc. C) less complex glycans for protein nodes 1000180, 1000181, 1000183, 1000184 and 1000725 in community 1. Few glycan show the determinant Neu5Ac2-6Galj1-4GlcNAc.

doi:10.1371/journal.pone.0095480.t002

All ConA interacting glycan nodes from community 1, 2 and 3 are shown in Table 3A, 3B and 3C, respectively. ConA interacting glycan nodes in community 1 are either mannose sugars or biantennary complex glycans such as transferrin and AGP-B. On the other hand, the ConA nodes in community 2 show preference for terminal glucose glycans.

In comparison to communities 1 and 2, the ConA nodes in community 3 show high preference for mannose containing sugars especially “N-glycan, high mannose” (Table 3C). These results agree with existing reports on ConA’s binding structure and specificity for mannose containing structures [55-57], in addition to the recognition of biantennary glycans, complex N-glycans [58] and terminal glucose [57].

Existing studies on SNA-I [47] and ConA [55-57] demonstrate the validity of the lectin-glycan interaction network and its detected community structure. Once a network is constructed, it is fairly easy to identify a lectin that explicitly binds to a certain glycan sequence by just selecting the lectin node of interest and its first neighbors in the network. The lectins in different communities show a dramatic difference in their glycan binding specificities. The current network-based approach should provide quick overall analysis and the use of glycan microarray data on the lectin-glycan interaction without time-consuming calculations.

Community Detection of the Lectin-glycan Interaction

We performed community detection of the lectin-glycan interaction network by using Mod-CSA [28], and compared the results with existing methods such as MCL [34,35], MCODE [36] and greedy algorithm [32,38]. The number of identified communities and the modularity values obtained by various community detection algorithms are shown in Table 4, Figure 4 and Figure 5.

From Table 4, Figure 4 & Figure 5a–d, it is clear that Mod-CSA [29] outperforms the other clustering methods in terms of the modularity score as well as the number of nodes left unclassified. The only method comparable to our modularity score of 0.37 obtained by Mod-CSA was the fast greedy algorithm [32,38] with a modularity score of 0.30. The algorithm recognizes clusters by repetitively eliminating edges from the network and then checks again which nodes are still connected [59]. The method detected 6 communities with the largest community containing 223 protein nodes and 298 glycan nodes (community 1) whereas the three smallest communities consist of either 4 nodes (community 4) or 3 nodes (community 5 & 6) only (see Figure 5b).

To compare the biological significance of modules (communities) obtained by Mod-CSA and by the greedy algorithm, we calculated the numbers of statistically meaningful enriched clusters of lectins that bind to the same specific glycan. The glycan binding specificity of each protein node was identified either from the literature or from Uniprot database as described in the methods section, and the significance of each glycan specific clusters was assessed by calculating its p-value (p ≤ 0.05). From Table 5, we observe that 44 statistically meaningful enriched clusters of lectins are identified with p-values ≤ 0.05. Whereas only 33 enriched clusters are identified by the greedy algorithm. This result suggests that many additional functionally related lectin clusters are identified by Mod-CSA, than detected by greedy algorithm.

For example, the greedy algorithm failed to identify 13 glycan specific lectin clusters (shown in bold in Table 5) that were identified by Mod-CSA. On the contrary, 3 glycan specific clusters (shown in italic bold in Table 5) were not detected by Mod-CSA, which are found by the greedy algorithm result. Specifically, the greedy algorithm failed to identify all fucose specific lectins, while Mod-CSA [29] successfully detected almost all fucose specific lectins and grouped them in community 1. Similarly, the greedy algorithm identified only five mannose related specificities in community 3, which is the major mannose binding community detected by greedy algorithm. However, Mod-CSA recognized eight mannose related specificities in community 1.

We compared our method with other popular clustering algorithms such as MCODE [36] and MCL [34,35]. MCODE method divided the network into a total of 23 clusters with the modularity score of −0.036. The largest cluster consists of 36 nodes whereas the smallest cluster contains only 4 nodes. However, only 3 clusters contain more than 10 protein nodes and they were further analyzed for enrichment of glycan specific lectin groups. The statistical analysis of these 3 clusters resulted in only 4 statistically meaningful lectin groups. From Figure 5c, we observe that a large number of single nodes (791) are not clustered into any groups. This is because MCODE identifies clusters of tightly connected nodes and does not intend to assign every node in the network to a cluster [59]. The main reason for this could be the fact that the MCODE algorithm is sensitive to noise in the network, particularly to false positive interactions [60]. Consequently, only a small number of strongly connected clusters are identified by MCODE and the rest of the nodes remain unclustered, which makes it hard to extract information from the network.
Table 3. The table shows all types of glycans interacting with ConA protein nodes.

| Glycan No. | Glycan Name                                                                 | Avg. RFU |
|------------|-----------------------------------------------------------------------------|----------|
| A)         |                                                                             |          |
| 144        | Man1-2Man1-6(Mans1-3)Man1-6(Mans2Man2Man1-3)Man1-6(4GlcNAc)1-4GlcNAc| 31059    |
| 139        | Man1-3(Mans1-6)Manx-3Sp3                                                    | 23784    |
| 136        | Man1-2Man1-3(Mans1-6)Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 23161    |
| 138        | Man1-3(Mans1-2Man1-2Man1-3Man1-6Man1-6)Man1-6Man1-6(4GlcNAc)1-4GlcNAc| 17347    |
| 137        | Man1-2Man1-3Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 14700    |
| 135        | Man1-2Man1-3Man1-6Man1-3Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 14334    |
| 143        | Man1-6(Mans1-2Man1-3Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 12786    |
| 145        | Man1-2Man1-2Man1-3Man1-2Man1-3Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 12581    |
| 112        | ω-D-Glc–Sp8                                                                 | 10407    |
| 75         | Gal1-4GlcNAc1-3Gal1-4Glc1-3Sp8                                              | 8329     |
| 151        | Neu5Gca2-3Gal1-6(Fuc1-3)GlcNAc–Sp0                                          | 8141     |
| 59         | Gal1-3GalNAc1-4Gal1-4Glc1-3Sp8                                              | 7380     |
| 113        | mixed glycans: Man5-9-N–Sp1                                                  | 6646     |
| 114        | Man1-6Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 6600     |
| 146        | Man1-3(Man1-6Man1-6Man1-6(4GlcNAc)1-4GlcNAc1-4GlcNAc| 6209     |
| 6          | Transferrin                                                                 | 5981     |
| 130        | Man1-2Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 5406     |
| 129        | Man1-6Man1-3(Man1-6Man1-2Man1-2Man1-3)Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 5259     |
| 199        | Man5-9mix-Asn                                                                | 5076     |
| 192        | Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 51705    |
| 198        | Man1-6Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 52238    |
| 196        | Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 49576    |
| 191        | Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 48888    |
| 190        | Man1-2Man1-3Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 44830    |
| 189        | Man1-2Man1-3Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 43717    |
| 197        | Man1-6Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 40190    |
| 195        | Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 38636    |
| 191        | Man1-2Man1-3Man1-6Man1-3Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 35442    |
| 177        | Glc1-4Glc1-3Sp8                                                             | 18139    |
| 179        | Glc1-6Glc1-6Glc1-3Sp8                                                        | 13465    |
| 178        | Glc1-4Glc1-3Sp8                                                              | 12700    |
| 180        | Glc1-4Glc1-3Sp8                                                              | 6825     |
| 186        | Glc1-6Gal1-6Sp8                                                              | 6057     |

B) Glycan Name

| Glycan No. | Glycan Name                                                                 | Avg. RFU |
|------------|-----------------------------------------------------------------------------|----------|
| 193        | Man1-2Man1-6(Man1-3)Man1-6(Man2Man2Man1-3)Man1-6(4GlcNAc)1-4GlcNAc| 52832    |
| 194        | Man1-2Man1-2Man1-3Man1-6Man1-3Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 52705    |
| 199        | Man5-9mix-Asn                                                                | 52238    |
| 198        | Man1-6Man1-3(Man1-6Man1-3)Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 51705    |
| 196        | Man1-3Man1-6Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 49576    |
| 192        | Man1-3Man1-6Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 48888    |
| 190        | Man1-2Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 44830    |
| 189        | Man1-2Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 43717    |
| 197        | Man1-6Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 40190    |
| 195        | Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 38636    |
| 191        | Man1-2Man1-3Man1-6Man1-6Man1-6Man1-6Man1-6Man1-3Man1-6(4GlcNAc)1-4GlcNAc| 35442    |
| 177        | Glc1-4Glc1-3Sp8                                                             | 18139    |
| 179        | Glc1-6Glc1-6Glc1-3Sp8                                                        | 13465    |
| 178        | Glc1-4Glc1-3Sp8                                                              | 12700    |
| 180        | Glc1-4Glc1-3Sp8                                                              | 6825     |
| 186        | Glc1-6Gal1-6Sp8                                                              | 6057     |
| Glycan No. | Glycan Name | Avg. RFU |
|-----------|-------------|---------|
| 609       | Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6 | 11536 |
| 607       | Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3-3Manb1-4GlcNAcb1-4GlcNAcb-Sp12 | 26362 |
| 577       | GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6 | 15181 |
| 576       | Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6 | 37709 |
| 575       | GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6 | 5650 |
| 561       | Gala1-3Galb1-4GlcNAcb1-2Mana1-6 | 7563 |
| 541       | GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6 | 40427 |
| 528       | Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0 | 5564 |
| 527       | Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0 | 19479 |
| 485       | Manb1-6Mana1-3Manb1-4GlcNAcb1-4Fucal1-6GlcNAcb-Sp19 | 17039 |
| 484       | Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6 | 34863 |
| 483       | Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6 | 6282 |
| 477       | Gala1-3GlcNAcb1-2Mana1-6 | 35401 |
| 476       | GlcNAcb1-6GlcNAcb1-2Mana1-6 | 10251 |
| 474       | Fucal1-2Galb1-3Fucal1-4GlcNAcb1-2Mana1-6 | 8363 |
| 458       | Gala1-3Fucal1-2Galb1-3GlcNAcb1-2Mana1-6 | 7281 |
| 456       | Gala1-3Fucal1-2Galb1-3GlcNAcb1-2Mana1-6 | 31036 |
| 455       | Gala1-3Fucal1-2Galb1-4GlcNAcb1-2Mana1-6 | 9449 |
| 422       | GlcNAcb1-2Mana1-6 | 19106 |
| 421       | Fucal1-2Galb1-3GlcNAcb1-2Mana1-6 | 5130 |
| 420       | Gala1-3Fucal1-2Galb1-4GlcNAcb1-2Mana1-6 | 21266 |
| 422       | GlcNAcb1-2GlcNAcb1-3GlcNAcb1-2Mana1-6 | 32937 |
| 421       | Fucal1-2Galb1-3GlcNAcb1-2Mana1-6 | 10863 |
| 418       | Gala1-4GlcNAcb1-2Mana1-6 | 15686 |
| 417       | Gala1-4GlcNAcb1-2Mana1-6 | 36858 |
Table 3. Cont.

| Glycan No. | Glycan Name | Avg. RFU |
|------------|-------------|----------|
| 404        | Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19 | 8992 |
| 399        | Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 37441 |
| 398        | GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 5577 |
| 397        | Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19 | 36771 |
| 396        | Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19 | 6293 |
| 395        | GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21 | 28946 |
| 394        | Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 6161 |
| 393        | Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 15799 |
| 392        | Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 8469 |
| 391        | Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 21683 |
| 390        | Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp22 | 41588 |
| 389        | Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21 | 12174 |
| 388        | Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 42514 |
| 387        | Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 5414 |
| 386        | Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 21541 |
| 385        | Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 33065 |
| 384        | Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 27555 |
| 383        | Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 (Fuca1-6)GlcNAcb-Sp22 | 40353 |
| 382        | Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 (Fuca1-6)GlcNAcb-Sp22 | 10006 |
| 381        | GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp22 | 42532 |
| 380        | Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 14973 |
| 379        | Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 (Fuca1-6)GlcNAcb-Sp22 | 38876 |
| 378        | Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12 (Fuca1-6)GlcNAcb-Sp22 | 48582 |
| 377        | GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp22 | 41416 |
| 376        | Neu5Ac2-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-4)GlcNAcb1-4GlcNAc-Sp12 | 58408 |
| 375        | Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 7604 |
| 374        | Neu5Ac2-2Galb1-4GlcNAcb1-2Mana1-6(Neu5Ac2-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 48590 |
| 373        | Fuca1-3Galb1-4GlcNAcb1-2Manb1-3(Fuca1-3(Galb1-4GlcNAcb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp20 | 19979 |
| 372        | Neu5Ac2-2Galb1-4GlcNAcb1-2Mana1-6(Neu5Ac2-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 42652 |
| 371        | Neu5Ac2-2Galb1-4GlcNAcb1-2Manb1-6(Neu5Ac2-2Galb1-4GlcNAcb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 29652 |
| 370        | Neu5Ac2-2Galb1-4GlcNAcb1-2Manb1-6(Neu5Ac2-2Galb1-4GlcNAcb1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12 | 45597 |
| 369        | Mana1-2Manb1-6(Mana1-2Manb1-3)Manb1-6(Mana1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp9 | 25444 |
| 368        | Mana1-2Manb1-6(Mana1-2Manb1-3)Manb1-6(Mana1-2Manb1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp9 | 35802 |
Among all four methods tested, the MCL algorithm performed worst in terms of its modularity value of \(-0.815\). MCL detected 33 clusters with the largest cluster consisting of 340 nodes while the smallest cluster has 2 nodes (Figure 5d). Similar to MCODE, the MCL method detected only 3 clusters containing more than 10 protein nodes and many nodes (689) in the network were not assigned to any group, again making it difficult to interpret these unassigned nodes. Therefore, these unassigned nodes were left out for further analysis. The MCL method resulted in only 12 statistically significant glycan specific groups.

If the performances of MCL and MCODE are hindered by false positive interactions, MCL and MCODE may perform better with networks generated using only reliable data. To find out if the Mod-CSA method outperforms the other methods regardless of

Table 3. Cont.

| Glycan No. | Glycan Name                                                                 | Avg. RFU |
|------------|------------------------------------------------------------------------------|----------|
| 315        | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-3(GlcNAcb1-2Man a1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12 | 37868    |
| 314        | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-3(Manb1-4GlcNAcb1-2GlcNAcb1-4GlcNAcb-Sp12) | 37828    |
| 313        | Man a1-2Man a1-3(Man a1-2Man a1-6Man a1-3Man a1-6Man a2-Sp9)                 | 16574    |
| 312        | Man a1-6(Mana1-3Man a1-6Man a1-3)Man a1-6Man a2-Sp10                        | 6909     |
| 309        | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(GlcNAcb1-2Man a1-3Manb1-4GlcNAcb1-4GlcNAcb-Sp12) | 41526    |
| 302        | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Galb1-4GlcNAcb1-2Man a1-3Manb1-4GlcNAcb1-4GlcNAcb-Sp12) | 6413     |
| 217        | Man a1-6(Man a1-3Man a1-6Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12)             | 38826    |
| 216        | Man a1-6(Man a1-3Man a1-6Man a1-2Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12)     | 28738    |
| 215        | Man a1-2Man a1-2Man a1-6Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12               | 34121    |
| 214        | Man a1-6(Man a1-3Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12)                    | 39736    |
| 213        | Man a1-6(Man a1-3Man a1-6Man a1-2Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12)     | 8657     |
| 212        | Man a1-2Man a1-6(Man a1-3Man a1-6Man a1-2Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12) | 53582    |
| 211        | Man a1-3Man a1-6Man a2-Sp9                                                  | 18045    |
| 210        | Man a1-2Man a1-3Man a2-Sp9                                                  | 35979    |
| 209        | Man a1-2Man a1-6(Man a1-3Man a1-2Man a1-6Man a1-3Man b1-4GlcNAcb1-4GlcNAcb-Sp12) | 7614     |
| 208        | Man a1-2Man a1-2Man a1-3Man a2-Sp9                                           | 37383    |
| 207        | Man a1-2Man a1-3Man a2-Sp9                                                  | 6620     |
| 205        | Man a1-3Man a1-2Man a1-6Man a1-6Man a2-Sp9                                   | 24990    |
| 58         | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-3)     | 8188     |
| 57         | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-3)     | 38425    |
| 56         | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-3)     | 5338     |
| 55         | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-3)     | 61880    |
| 54         | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6)     | 40848    |
| 53         | GlcNAcb1-2Man a1-6GlcNAcb1-2Man a1-6Man b1-4GlcNAcb1-4GlcNAcb-Sp13             | 42923    |
| 52         | Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man a1-6)     | 54802    |
| 51         | Man a1-6(Man a1-3Man a1-6Man a2-Sp13)                                        | 42505    |
| 50         | Man a1-3(Man a1-6Man a2-Sp13)                                                | 8976     |
| 49         | Glna1-2Man a1-3Man a1-6Glna1-2Man a1-6Man b1-4GlcNAcb1-4GlcNAcb-Sp13         | 5961     |
| 48         | Man a1-3(Man a1-6Man b1-4GlcNAcb1-4GlcNAcb-Sp13)                              | 37798    |

A) ConA interacting glycan nodes from community 1 are shown. These ConAs interact either with mannose nodes or biantennary complex glycans such as Transferrin and AGP-B. B) ConA interacting glycan nodes from community 2 are shown. They show preference for terminal glucose glycans. C) ConA nodes in community 3 show high preference for “N-glycan, high mannose” sugars.
doi:10.1371/journal.pone.0095480.t003
the amount of potentially false information, we performed the enriched cluster analysis on two additional networks generated using more stringent RFU criteria, RFU ≥10000 and RFU ≥ 20000 (see Table S2). The results remain same regardless of the RFU cutoff values used to generate the network. For example, the numbers of statistically significant glycan specific groups identified by Mod-CSA are 41 and 35 using RFU cutoff values of 10000 and 20000, respectively. However, the greedy algorithm provides 23 and 20 statistically significant glycan specific groups. Similarly, with the MCL method, 20 and 14 statistically significant glycan specific groups were identified (see Table S3). Surprisingly, MCODE detected no statistically significant glycan specific lectin groups from more stringent networks.

Finally, we compared the clusters obtained by Mod-CSA with random clusters. We divided the nodes into four random clusters, which have the same number of nodes with those detected by Mod-CSA. This process was iterated 20 times and the average number of statistically enriched glycan-specific groups detected by random clustering was compared with that by Mod-CSA. The maximum and minimum number of significantly enriched lectin groups was 11 and 1, respectively. On average, these 20 random permutations of clusters resulted in about 7 glycan-specific lectin groups having p-value $<0.05$ (see Table S4). A comparison of the number of significantly enriched lectin groups detected by the different clustering methods is shown in Figure 6. All these results demonstrate that Mod-CSA extracts more information than the other widely used clustering methods, and it can serve as a powerful tool for investigating the lectin-glycan interaction.

The Optimal Community Structure of the Lectin-glycan Interaction Network

It has been shown that Mod-CSA can provide globally optimal modularity partitioning of a network containing up to 2000 nodes [31]. Since our lectin-glycan network has 1119 nodes, we believe that the Mod-CSA result corresponds to the optimal grouping of the network in terms of its modularity. The optimal modularity grouping of lectins and glycans results in 4 communities with the modularity score of 0.37. We attempted to explore the relationship between all nodes within the same community on the basis of structure and function of each lectin and the type of glycan binding specificity. Each lectin node was assigned with its known glycan binding specificity, and the statistical significance of their grouping was assessed by calculating its p-value ($p<0.05$) (see Table 5 and Figure 4). A brief description of each community is given below:

**Community 1 (Fucose specific).** This is the largest community of the lectin-glycan network detected by Mod-CSA analysis and contains 168 protein nodes and 215 glycan nodes, respectively. This community is dominated by protein nodes with fucose specific lectins, such as ulex europaeus agglutinin I (UEA-I), alcuria aurantia lectin (AAL), ralstonia solanacearum lectin (RSL), etc. The fucose binding sites of RSL are very similar to those of previously reported five fucose-binding sites of AAL [61]. Fucose-containing xyloglucans are known to promote signaling consequences on plant tissues [62]. The other types of overrepresented lectins in this community have specificity for Galactose- and N-acetylgalactosamine binding with cell adhesion as their main function. The most common protein domains correspond to these galactose specific lectins are H_lectin (PFAM ID: PF09458) domain, which is involved in self/non-self recognition of cells through binding with carbohydrates [63], and Galactose-binding domain-like domain known as Discoidin domain (PFAM ID: PF00754), which is found in many blood coagulation factors. The galactose specific lectins in this community include agglutinin from Helix pomatia, Dictoind I and Dictoind II from Dictostelium discoideum (Slime mold). Additionally, the unannotated lectins in this cluster such as 6RG, Tap1, Mubin1 show specificity for galactose or fucose sugars (see Table S3), which strongly indicates that these proteins are related to cell adhesion.

This community contains the top hub PP2A1 (1001943) with the largest node degree of 257. The other three PP2A1 nodes (1002090, 1002091 and 1002092) belong to community 2. The list of unique glycans that interact with these PP2A1 nodes are given below:

| Table 4. A summary of various clustering methods tested in this work. |
|-----------------|-----------------|-----------------|-----------------|
| **Method** | **No. of Clusters** | **Modularity** | **Description** |
|---------------|-----------------|-----------------|-----------------|
| Mod-CSA | 4 | 0.37 | The conformational space annealing based modularity optimization method. |
| Greedy | 6 | 0.30 | Fast greedy community detection algorithm. |
| MCODE | 23 | -0.04 | Bader and Hogue algorithm for finding modules in networks. |
| MCL | 33 | -0.81 | Markov clustering algorithm from van Dongen that uses random walks to simulate flow. |

Mod-CSA outperforms the other popular clustering methods in terms of the modularity score. doi:10.1371/journal.pone.0095480.t004
summarized in Table S6. From this table it can been seen that PP2A1 nodes show specificity for a diverse range of glycans such as GlcNAc, high-mannose N-glycans and sialic acid containing glycans. Recently, Beneteau et al., (2010) [64] in their glycan array experiments have shown that PP2A1 binds to different types of carbohydrates. This indicates the possibility that the phloem PP2 lectin plays roles in numerous functions, recognizing either endogenous glycoproteins or glycosylated receptors of pathogens. This diversity in glycan binding by PP2A1 could be attributed to the presence of several carbohydrate-binding sites in PP2A1 [64].

Community 2 (Galb1-3GalNAc specific). This is the smallest community with 98 protein nodes and 133 glycan nodes. Community 2 is rich in N-acetylgalactosamine and N-acetylgalactosamine binding lectins such as Wheat Germ Agglutinin (WGA), Griffonia simplicifolia II (GS-II), and Sclerotium rolfsii lectin (SRL). WGA belongs to a highly conserved family of chitin-binding lectins from cereals (Gramineae), such as rye, barley, rice and wheat [65]. Chitin, a polymer of β-1,4-N-acetylglucosamine is present in the cell wall of many fungi, in the exoskeleton and digestive tract of some insects, and in some nematodes [66]. Similarly, GS-II, also an N-acetylglucosamine-specific legume lectin, has insecticidal activity against cowpea weevil [67]. In contrast to WGA and GS-II, SRL displays strong binding to O-linked galactose-beta-1,3-N-acetylgalactosamine, disaccharide (Thomsen Friedenreich antigen) similar to Agaricus bisporus lectin [68]. Similarly, the other N-acetylgalactosamine specific lectins in this group are involved in the binding of T-antigen structure Gal-beta1,3-GalNAc e.g. Agglutinin alpha chain (Jacalin alpha chain) from Artocarpus integer (Jack fruit) and Agglutinin alpha chain (MPA) from Maclura pomifera (Osage orange).

Figure 5. Communities generated by four methods are shown. (a) Mod-CSA generated communities are shown. In each community, glycans nodes are represented by circles whereas the protein nodes are shown as diamonds. From the figure it can be seen that all the nodes in a network have been assigned to a community. Community 1 has PP2A1 as hub node where as Community 4 has two hub nodes, WGA1 and RCA. (b) Greedy algorithm generated communities are shown. The nodes are color coded as per the Mod-CSA result. Each of the first three communities (community 1 to 3) contain a hub node where as communities 4-6 have only a few nodes. (c) MCODE generated communities are shown. Many nodes are not clustered, and the three hubs are grouped into one community. (d) MCL generated communities are shown many nodes are not clustered at all. Hub nodes are not clustered with any other nodes.

doi:10.1371/journal.pone.0095480.g005
Unannotated protein nodes are represented by lectins such as Protein PHLOEM PROTEIN 2-LIKE A1 (PP2A1) from Arabidopsis thaliana and Codium fragile lectin (CFT) from Codium fragile [Dead man’s fingers] (Green alga). PP2A1 is known to interact with diverse types of carbohydrates and may be involved in numerous recognition functions [64]. On the other hand, CFT shows preference for the α-anomer of GalNAc and recognizes GalNAca1 sequences as well as high affinity for the Forssman pentasaccharide and for Galb1-3GalNAc-a- [69], which is one of the overrepresented (p-value, 0.05) glycan specific group in this community. Lists of unique glycans for PP2A1 and CFT nodes are summarized in Table S7.

Community 3 (Mannose specific). Protein nodes in this group are dominantly mannose binding lectins and nine out of twelve statistically significant glycan groups are mannose specific. Many members of these mannose specific lectins have B_lectin (PFAM ID: PF01453) structural domain. The members of this family are mannose specific and belong to Bulb lectin super-family (Amaryllidaceae, Orchidaceae and Aliaceae). For example, Galanthus nivalis agglutinin (GNA), a mannose-specific lectin from snowdrop bulbs, is a tetrameric member of the family of Amaryllidaceae lectins that exhibit antiviral activity towards HIV [70]. Other mannose binding lectins in this group have Lecitn_legB (PFAM ID: PF00139) structural domain and require metal ions like Ca and Mn ions for carbohydrate binding and cell-agglutinating activities. Examples include ConA and Garden pea lectin. The group also includes various high mannose binding lectins such as Hippeastrum hybrid lectin (HHL), Narcissus pseudo-narcissus agglutinin (NPA), Salt stress-induced protein, Allium sativum agglutinin (ASA), etc. Another mannose binding lectin in this group which has an antiviral activity is Cyanovirin-N (CV-N). The antiviral activity of CV-N is mediated through specific interactions with the viral surface envelope glycoproteins gp120 and gp41, as well as to high-mannose oligosaccharides found on the HIV envelope [71]. Other lectins that were grouped in this community for which we could not find the reported glycan specificity include Arum maculatum agglutinin (AMA), Caragana arborescens agglutinin (CAA), Colchicum autumnale lectin (CA), and Arisaema helloborifilium schott lectin (AHL). All these lectins also show high specificity for mannose sugars (Table S8). Overall the community consists of 147 protein nodes and 124 glycan nodes.

Community 4 (GalNAc specific). From Table 5 it can be observed that this community is enriched in GalNAc specific lectins such as Datura stramonium agglutinin (DSA), Soybean agglutinin (SBA), Vicia villosa agglutinin (VVA), Bauhinia purpurea lectin (BPL), etc. These galactose specific lectins may play a significant role in cell-agglutinating activities e.g. VVA (Lectin B4) from Vicia villosa (Hairy vetch). Another galactose-specific lectin in this group is a legume lectin known as Erythrina cristagalli lectin (ECL) [72]. Although its function in the legume is unknown, it has been shown that ECL possesses hemagglutinating activity and it is believed to be mitogenic for human T lymphocytes [73]. A large number of plant and fungal proteins (e.g. solanaceous lectins of tomato and potato, plant endochitinases, the wound-induced proteins: hevein, win1 and win2, and the Kluyveromyces lactis killer toxin alpha subunit) that bind N-acetylglucosamine contain chitin-binding domain (PFAM ID: PF00187). These proteins might function as a defence against chitin containing pathogens, e.g. Chitin-binding lectin 1 of Solanum tuberosum (Potato). This community also includes lectins such as Macrolepiota procera agglutinin (MPA) and Laccaria bicolor lectin both of which show high specificity for complex GalNAc glycans (Table S9). This community consists of 100 protein and 134 glycan nodes.

Additionally, this community includes 2 out of three hub nodes identified in the lectin-glycan array network. One of the hubs...
Table 5. Lists of statistically meaningful enriched clusters (p≤0.05) of lectins binding to the identical glycan are shown.

| Cluster No. | No. of members | Reported Specificity | P-value | Cluster No. | No. of members | Reported Specificity | P-value |
|-------------|----------------|----------------------|---------|-------------|------------------|----------------------|---------|
| 1           | 168            | a-Linked terminal GalNAc | 0.0055  | 2           | 223             | a-Linked terminal GalNAc | 0.0028  |
|             |                | Chitin oligomers, Sia  | 0.0109  |             |                  | Chitin oligomers, Sia  | 0.0006  |
|             |                | **Fuca1-2Galb1⁻→ or GalNAcb⁻→** | 0.0347  |             |                  | **Fuca1-2Galb1-3GalNAcb⁻→4(Neu5Ac⁻2-3)Galb1-4GlcNcb OR Galb1-3GalNAcb⁻→4(Neu5Ac⁻2-3)Galb1-4GlcNcb** | 0.0352  |
|             |                | Fuca1-2Galb1-3GalNAcb⁻→4(Neu5Ac⁻2-3)Galb1-4GlcNcb OR Galb1-3GalNAcb⁻→4(Neu5Ac⁻2-3)Galb1-4GlcNcb | 0.0112  |             |                  | Fuca1-2Galb1-4GlcNcb | 0.0001  |
|             |                | Fuca1-2Galb1-4GlcNcb | 3.49E-06 |             |                  | Fuca1-2Galb1-4GlcNcb | 0.0001  |
|             |                | **Fuca1-6GlcNcb (core fucose)** | 0.0347  |             |                  | Fucose binding lectin | 0.0028  |
|             |                | Fuca1-6GlcNAc, Fuca1-3(Galb1-4)GlcNAc | 0.0004  |             |                  | Galactose binding lectin | 0.0117  |
|             |                | Fucose binding lectin | 0.0004  |             |                  | Galactose- and N-acetylgalactosamine-binding | 0.0065  |
|             |                | Galactose- and N-acetylgalactosamine-binding | 0.0147  |             |                  | Galb1-3GalNAc | 0.0052  |
|             |                | Mannose binding lectin | 0.0001  |             |                  | Galb1-3GalNAc, GalNAc | 0.009   |
|             | 2              | **terminal N-acetylgalactosamine (GalNAc)** | 0.0347  |             |                  | Galactosylated tri/tetra antennary glycans, GlcNAc | 0.0004  |
|             | 98             | Chitin oligomers, Sia  | 5.27E-10 |             |                  | Mannose binding lectin | 3.47E-06 |
|             |                | Galb1⁻→3GalNAc-a- | 4.40E-08 |             |                  | N,N’-diacetyllactosidediamine (GalNAc)⁻1-4GlcNcb, LacdiNAc) | 0.0069  |
|             |                | Galb1-3GalNAc | 0.0007  |             |                  | Siaa2-3Galb1⁻→ | 0.0151  |
|             |                | Galb1-3GalNAc, GalNAc | 0.0021  | 2             | 190             | (GlcNAc)⁻1-4n, Galb1-4GlcNcb | 0.0257  |
|             |                | Mannose binding lectin | 0.0207  |             |                  | Agalactosylated tri/tetra antennary glycans, GlcNAc | 0.0257  |
|             |                | **N-acetylglucosamine and N-acetyleneuraminic acid** | 0.0068  |             |                  | Chitin oligomers, Sia | 3.16E-07 |
| 3           | 147            | Fuca1-6GlcNAc, a-D-Glc, a-D-Man | 3.36E-05 |             |                  | Fuca1-2Galb1-4GlcNcb | 0.0119  |
|             |                | Galb1-3GalNAc | 0.0114  |             |                  | Fuca1-6GlcNAc, a-D-Glc, a-D-Man | 0.0005  |
|             |                | High-mannose, Mana1-3Man | 0.0223  |             |                  | Galactose binding lectin | 0.0177  |
|             |                | High-mannose, Mana1-3Man, Mana1-6Man | 0.0162  |             |                  | Galb1⁻→3GalNAc-a- | 4.17E-05 |
|             |                | High-mannose, Mana1-6(Mana1-3)Man | 8.12E-07 |             |                  | High-mannose, Mana1-6(Mana1-3)Man | 0.0485  |
|             |                | High-mannose, Mana1-6Man | 0.0026  |             |                  | Mannose binding lectin | 0.0196  |
|             |                | **Mana1-3(Mana1-6)Man, bi- and tri-antennary complex-type N-glycan, GalNAc** | 0.0232  |             |                  | Siaa2-6Gal/GalNAc | 0.0469  |
|             |                | **Manb Anywhere** | 0.0232  |             |                  | Tri/tetra-antennary complex-type N-glycan | 0.0184  |
|             |                | Mannose binding lectin | 5.86E-06 | 3             | 93              | High-mannose, Mana1-3Man | 0.0092  |
|             |                | N-acetylglucosamine | 0.0162  |             |                  | High-mannose, Mana1-3Man, Mana1-6Man | 0.0105  |
|             |                | Siaa2-6Gal/GalNAc | 0.0044  |             |                  | High-mannose, Mana1-6Man | 5.32E-06 |
Table 5. Cont.

| Mod-CSA (Q = 0.366) | Community (Glay) (Q = 0.3) |
|---------------------|---------------------------|
| Cluster No. | No. of members | Reported Specificity | P-value | Cluster No. | No. of members | Reported Specificity | P-value |
| 4 100 | (GlcNAcb1-4)n, Galb1-4GlcNAc | 0.0013 | N-acetylglucosamine | 0.0105 |
| | a- or b-linked terminal GalNAc, GalNAca1-3Gal | 0.0003 | Siaa2-6Gal/GalNAc | 0.0049 |
| | Agalactosylated tri/tetra antennary glycans, GlcNAc | 0.0013 | Subterminal Mannose | 0.0058 |
| | Bi-antennary complex-type N-glycan with outer Gal and bisecting GlcNAc | 0.0056 | 4 4 NA | NA |
| | Galb1-4GlcNAc | 0.0004 | 5 3 NA | NA |
| | GalNAca1-3GalNAc, blood group A antigen | 0.0056 | 6 3 NA | NA |
| | GalNAcb1-4GlcNAc, Galb1-3(-6)GalNAc | 0.0056 |
| | GlcNAc oligomers, oligosaccharide containing GlcNAc and LacNAc | 0.0474 |
| | GlcNAc trimers/tetramers | 0.0056 |
| | Mannose binding lectin | 0.0003 |
| | N,N-diacetlylactosidemine (GalNAc) | 0 |
| | Siaa2-3Galb1-3[Siaa2-6]GalNAc | 0.0036 |
| | Siaa2-3Galb1-4GlcNAc | 0.0377 |
| | Tri/tetra-antennary complex-type N-glycan | 0.0234 |

Communities generated by Mod-CSA and greedy algorithm are used. The statistical significance of each reported glycan binding lectin was calculated by hypergeometric distribution using p < 0.05. For each glycan listed in Table S1, interacting lectin nodes were identified to calculate the significance of the community structure determined in this study. The number of statistically significant glycan-specific groups according to Mod-CSA partitioning is 44 (p-value < 0.05) while greedy algorithm provides only 33 groups. 15 glycan-specific groups generated by Mod-CSA but not by greedy algorithm are shown in bold, whereas 3 groups generated by greedy algorithm but not by Mod-CSA are shown in italic bold.

doi:10.1371/journal.pone.0095480.t005

represent protein node (1004763) for wheat germ agglutinin (WGA) from Triticum vulgaris (wheat), whereas the second node (1004668) represents Ricinus communis agglutinin (RCA) from Ricinus communis (castor bean). WGA is a stable homodimer protein and exhibits specificity for N-acetylgalactosaminic acid and N-acetylgalactosamine (GlcNAc) sugars. The glycans for WGA hub node are summarized in Table S10 and it can be observed that almost all these glycans have GlcNAc group, while few others contain N-acetylgalactosaminic acid. Each monomeric unit of WGA consists of four domains (A–D) which can be further classified into “primary” (B and C domains) and “secondary” (A and D domains) binding sites showing dissimilar affinities for GlcNAc containing moieties [74]. These structural characteristics and the closeness of binding sites make WGA a worthy candidate to explore multivalent protein-carbohydrate interactions and to assess the impact of structural modifications of glycoconjugates [75]. These multivalent interactions are favorable as compared to monomeric ones and are frequently employed by nature to control the closeness of binding sites making WGA a worthy candidate to explore multivalent protein-carbohydrate interactions and to assess the impact of structural modifications of glycoconjugates [75]. These multivalent interactions are favorable as compared to monomeric ones and are frequently employed by nature to control the number of statistically significant glycan-specific groups according to Mod-CSA partitioning is 44 (p-value < 0.05) while greedy algorithm provides only 33 groups. 15 glycan-specific groups generated by Mod-CSA but not by greedy algorithm are shown in bold, whereas 3 groups generated by greedy algorithm but not by Mod-CSA are shown in italic bold.

RCA as well as ECL recognize carbohydrate chains with non-reducing terminal β-d-galactose (Galβ) and show preference to Galβ1-4GlcNAc instead of Galβ1-3GlcNAc sequence [77,78]. The diverse types of glycans including Galβ1-4GlcNAc that interact with RCA hub node are listed in Table S11. The table also shows many Neu5Aca2-6Galβ1 sugars having large RFU values. RCA is a glycoprotein from seeds of castor plants and one of the most important applied lectins that have been widely used as a tool to study cell surfaces and to purify glycans [79]. RCA promotes binding and agglutination of polysaccharides and glycoproteins in addition to liposomes and micelles containing glycolipids with galactosyl residues [80,81]. Furthermore, the specificities of interactions of RCA with neutral and sialylated oligosaccharides have been well established and is consistent with our results as summarized in Table S11 [82].

The current community-based network study of the lectin-glycan microarray data provides not only a quick and systematic analysis of lectin specificities, but also global organization and grouping of biologically related lectins along with their binding partners (glycans). Such information will be vital to identify lectins that bind to particular glycan structures or to catalogue lectins according to the similarity in specificities. Another important significance of the community-based network analysis is the identification of a novel lectin and the initial guess about its specificity. For this, a sequence database should be constructed for each community identified and a target lectin under investigation should be fed into the databases to get an idea about the structural-functional role of the query lectin and the type of glycans it might bind to. This approach will be more practical when the communities have a large number of different lectins and might help in...
determining the glycan binding nature of a given lectin. There are many network-based protein function prediction methods along with approaches utilizing structural or sequence information of proteins. Recently, when dealing with a protein-protein-interaction network, it has been shown that more accurate protein function prediction results were obtained by modularity based community detection of the network. The current study provides the first attempt to study lectin-carbohydrate interactions via community detection of a network.

**Conclusion**

We have constructed a bipartite lectin-glycan interaction network from the collection of glycan microarray data. The network itself provides a quick and global view of the lectin-glycan interaction from which hub proteins are identified. We find that the hub proteins match well with the characteristics of known biological relevance. Using Mod-CSA, a recently developed efficient community detection method, 4 modules are identified. The clustering results are shown to be biologically more meaningful than those obtained by other widely used methods. Most significantly, 44 statistically significant glycan specific groups are identified including fucose and mannose binding ones, some of which could not be detected by alternative methods. Even with more strict RFU cut-offs, clusters generated by Mod-CSA provide consistently better results as compared to other methods. We provide overall analysis of 4 communities identified in the lectin-glycan microarray network. We also show how multiple lectins from the same plant, such as *Sambugus nigra* (SNA-I and SNA-II) are grouped into different communities based on their glycan binding specificities. The network study provides a framework to get a broad picture of data containing many interacting components. These capabilities of a community-based network analysis allow researchers to explore, analyze and compare a variety of proteins and glycans within the context of modules/communities identified in the network. We expect that this will trigger interest in the prediction of protein-carbohydrate interactions using biological networks and will have wider applications as additional glycan binding proteins are identified. The method can also be applied to study other types of lectins as well as other interaction networks.

**Supporting Information**

Table S1 List of all protein nodes, their clusters and reported specificity in the lectin-glycan network. (XLS)

Table S2 The list of meaningful glycan-specific groups and their P-values detected by Mod-CSA and greedy algorithm (GLAY) at RFU ≥10000 and RFU ≥20000. (XLS)

Table S3 The list of meaningful glycan-specific groups and their P-values detected by MCL and MCODE at RFU ≥5000, RFU ≥10000 and RFU ≥20000. (XLS)

Table S4 List of randomly identified statistically significant glycan-specific groups. (DOCX)

Table S5 List of unique galactose and fucose sugars that interact with unannotated 6RG, Tap1, and Mubin at RFU ≥5000 in the lectin-glycan network. (XLS)

Table S6 List of diverse glycans that interact with the hub PP2A1 at RFU ≥5000 in the lectin-glycan network. (XLS)

Table S7 Lists of unique glycans for PP2A1 and CFT at RFU ≥5000 in the lectin-glycan network. (XLS)

Table S8 List of unique glycans for unannotated lectins Arum maculatum agglutinin (AMA), Caragana arborescens agglutinin (CAA), Colchicum autumnale lectin (CA), and Arisaema helleborifolium schott lectin (AHL) that show high specificity for mannose sugars at RFU ≥5000 in the lectin-glycan network. (XLS)

Table S9 List of complex glycans that show high specificity for lectins such as Macrolepiota procera agglutinin (MPA) and Laccaria bicolor lectin. (XLS)

Table S10 List of diverse glycans that interact with the hub WGA at RFU ≥5000 in the lectin-glycan network. (XLS)

Table S11 List of diverse glycans that interact with the hub RCA at RFU ≥5000 in the lectin-glycan network. (XLS)

**Author Contributions**

Conceived and designed the experiments: AM Juyoung Lee Jooyoung Lee. Performed the experiments: AM Juyoung Lee. Analyzed the data: AM Juyoung Lee Jooyoung Lee. Contributed reagents/materials/analysis tools: AM Juyoung Lee Jooyoung Lee. Wrote the paper: AM Juyoung Lee Jooyoung Lee.

1. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, et al. (1999) Essentials of Glycobiology. Cold Spring Harbor Laboratory Press, New York.
2. Shirver Z, Ragham S, Susskehkan R (2006) Glycomics: a pathway to a class of new and improved therapeutics. Nat Rev Drug Discov 3: 863–873.
3. Perrimon N, Bernfield M (2000) Specificities of heparin sulphate proteoglycans in developmental processes. Nature 404: 725–728.
4. Ioffe E, Stanley P (1994) Mice lacking N-acetylglucosaminyltransferase activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. Proc Natl Acad Sci 91: 728–732.
5. Jin L, Abrahams JP, Skinner R, Piteau M, Pike RN, et al. (1997) The anticoagulant activation of antithrombin by heparin. Proc Natl Acad Sci 94: 14603–14608.
6. Fu X, Albermann C, Jiang J, Liao J, Zhang C, et al. (2003) Antibiotic optimization via in vitro glycoreandomization. Nat Biotechnol 21: 1467–1469.
7. Freeze HH (2006) Genetic defects in the human glycome. Nat Rev Genet 7: 537–551. Erratum in: Nat Rev Genet, 7: 660.
8. Feizi T, Fazio F, Chai W, Wong CH (2003) Carbohydrate microarrays - a new set of technologies at the frontiers of glycomics. Curr Opin Struct Biol 13: 637–645.
9. Imberty A, Lorret-Jacob H, Perez S (2007) Structural view of glycosaminoglycan-protein interactions. Carbohydr Res 342: 430–439.
10. Sharon N, Lis H (1972) Lectins: cell-agglutinating and sugar-specific proteins. Science 177: 949–959.
11. McCoy JP, Varani J, Goldstein IJ (1983) Enzyme-linked lectin assay (ELLA): use of alkaline phosphatase-conjugated Griffonia simplicifolia B4 isoelectric for the detection of alpha-D galactopyranosyl end groups. Anal Biochem 130: 437–444.
12. Duverger E, Frison N, Roche AC, Monnaguy M (2003) Carbohydrate-lectin interactions assessed by surface plasmon resonance. Biochimica 83: 167–179.
13. Dain TK, Brewer CF (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. Chem Rev 102: 387–429.
14. Park S, Lee MR, Shin I (2007) Fabrication of carbohydrate chips and their use to probe protein-carbohydrate interactions. Nat Protoc 2: 2747–2738.
15. Tarora C, Jones S, Thornton JM (2000) Analysis and prediction of carbohydrate binding sites. Protein Eng 13: 89–98.

16. Shionryu-Miyazawa C, Shirai T, Ishida H, Yamane T (2003) An empirical approach for structure-based prediction of carbohydrate binding sites on proteins. Protein Eng Des Sel 16: 467–478.

17. Malik A, Ahmad S (2007) Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a new methodology. BMC Struct Biol 7: 1.

18. Davis S, Alexander D, King A, Khuri S, Keirouz W (2009) Predicting the specificity of lectins from protein sequences. Proteins 70: 67–82.

19. Kulkarni M, Bridgett SJ, Goody RS, Jackson RM (2009) InCaSiteFinder: a method for structure-based prediction of inositol and carbohydrate binding sites in proteins. J Mol Graph Model 28: 297–303.

20. Malik A, Fries A, Jia V, Ahmad S (2010) PROCARG: A Database of Known and Modelled Carbohydrate-Binding Protein Structures with Sequence-Based Prediction Tools. Adv Bioinformatics 2010: 168603.

21. Agarwal S, Mishra NK, Singh H, Raguram GP (2011) Identification of mannos binding residues using local composition. PLoS One 6(6): e24095.

22. Park S, Shin I (2002) Fabrication of carbohydrate chips for studying protein-carbohydrate interactions. Angew Chin Int Ed Engl 41: 3180–3182.

23. Wang D, Liu S, Trumner JJ, Deng C, Wang A (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. Nat Biotechnol 20: 275–281.

24. Fukui S, Feizi T, Galustian C, Lawson AM, Chai W (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-receptor interactions. Nat Biotechnol 20: 1011–1017.

25. Houseman BT, Mckish M (2002) Carbohydrate Arrays for the Evaluation of Protein Binding and Enzyme Activity. Chem Biol 9: 443–454.

26. Porter A, Yue T, Herrling L, Day S, Suh E, et al. (2010) A motif-based analysis of glycan array data to determine the specificities of glycan-binding proteins. Glycobiology 20: 369–380.

27. Smith DF, Song X, Cummings RD (2010) Use of glycan microarrays to explore carbohydrate-protein binding. Glycobiology 20: 121–138.

28. Nakao K, Takeda M, Hashi BB (2012) The fine specificity of mannos-binding and galactose-binding lectins revealed using outlier motif analysis of glycan array data. Glycobiology 22: 160–169.

29. Lee J, Gross SP, Lee J (2012) Mod-CSA: Modularity optimization by consensus space annealing. Phys Rev E Stat Nonlin Soft Matter Phys 85: 056702.

30. Lee J, Lee J (2013) Hidden information revealed by original community structure from a protein-complex bipartite network improves protein function prediction. PLoS One 8(5): e60572.

31. Lee J, Gross SP, Lee J (2013) Improved network community structure improves function prediction. Sci Rep 3: 2197.

32. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69: 026113.

33. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, et al. (2006) Crystallographic and Modelled Carbohydrate-Binding Protein Structures with Sequence-Based Prediction Tools. Adv Bioinformatics 2010: 168603.

34. van Dongen S (2000a) Graph Clustering by Flow Simulation. Unpublished doctoral dissertation. Centre for Mathematics and Computer Science, University of Utrecht, The Netherlands.

35. Smith DF, Song X, Cummings RD (2010) Use of glycan microarrays to explore specificity of glycan-binding proteins. Methods Enzymol 488: 417–444.

36. Masson KA, Lennarz WJ, Hashi BB (2012) The fine specificity of mannos-binding and galactose-binding lectins (1987) The elderberry (Sambucus nigra L.) bark lectin recognizes the conformational space annealing. Phys Rev E Stat Nonlin Soft Matter Phys 85: 056702.

37. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, et al. (2005) The UniProt protein knowledgebase. Nucleic Acids Res 33: D194–D197.

38. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, et al. (2012) InterPro: A consortium database for protein families, domains and sites. Nucleic Acids Res 40: D221–D225.

39. Kaku H, Peumans WJ, Goldstein IJ (1990) Isolation and characterization of a lectin from the endosperm of rye grain. Biochim Biophys Acta 1015: 103–113.

40. Poretz RD, Goldstein IJ (1970) An examination of the topography of the saccharide-binding sites of concanavalin A and of the forces involved in complexation. Biochemistry 9: 2890–2896.

41. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411: 41–42.

42. Hardman KD (2002) Structure of concanavalin A at 2.4-A resolution. Biochemistry 11: 4910–4919.

43. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, et al. (2012) InterPro: A consortium database for protein families, domains and sites. Nucleic Acids Res 40: D221–D225.

44. Saiki R, Gotoya M, Misawa Y, Yamashita M, Hashimoto C, et al. (2010) Molecular design of N-linked tetravalent glycosides bearing N-acetylglucosamine, N,N'-diacetyltartaric acid and N-acetylated lactose: Analysis of cross-linking activities with WGA and ECA lectins. Bioorg Med Chem 18: 621–629.
77. Itakura Y, Nakamura-Tsuruta S, Kominami J, Sharon N, Kasai K, et al. (2007) Systematic comparison of oligosaccharide specificity of Ricinus communis agglutinin I and Erythrina lectins: a search by frontal affinity chromatography. J Biochem 142: 459–469.
78. Tateno H, Mori A, Uchiyama N, Yahe R, Isaki J, et al. (2008) Glycoconjugate microarray based on an evanescent-field fluorescence-assisted detection principle for investigation of glycan-binding proteins. Glycobiology 18: 789–798.
79. Wu AM, Wu JH, Singh T, Lai JJ, Yang Z, et al. (2006) Recognition factors of Ricinus communis agglutinin 1 (RCA1). Mol Immunol 43: 1700–1715.
80. Kawaguchi T, Tagawa K, Senda F, Matsunaga F, Kitan H (1999) Recognition of Amphiphiles with Many Pendent Galactose Residues by Ricinus communis Agglutinin. J. Colloid Interface Sci 210: 290–295.
81. Cartellieri S, Helmholz H, Niemeyer B (2001) Preparation and evaluation of Ricinus communis agglutinin affinity adsorbents using polymeric supports. Anal Biochem 295: 66–75.
82. Wang Y, Yu G, Han Z, Yang B, Hu Y, et al. (2011) Specificities of Ricinus communis agglutinin 120 interaction with sulfated galactose. FEBS Lett 585: 3927–3934.