Spring 5-8-2022

Lectin-Glycan Complexes: A Comprehensive Analysis of Docking Calculations

Ashton Custer

LECTIN-GLYCAN COMPLEXES: A COMPREHENSIVE ANALYSIS OF DOCKING CALCULATIONS

By

Ashton Taylor Custer

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College as well as for the B.S. Chemistry, Biochemistry Emphasis, degree track requirements.

Oxford, MS

May 2022

Approved By

_____

Advisor: Dr. Robert J. Doerksen

_____

Reader: Dr. Sushil Mishra

_____

Reader: Dr. Ryan Fortenberry

DEDICATION

This thesis is dedicated to everyone who guided and encouraged me throughout the year, including my professors, advisors, and family members. Thank you.

ACKNOWLEDGEMENTS

I would like to thank Dr. Robert Doerksen for taking me into his GlyCORE research group and for getting me excited about glycoscience. I would also like to thank Dr. Sushil Mishra for teaching me, guiding me, and overseeing all of my research in this thesis. I would also like to thank Ayooluwa Aderibigbe and Priyanka Samanta for supporting me throughout my research. I would also like to thank GlyCORE and NIH for funding and supporting my project.

ABSTRACT

ASHTON TAYLOR CUSTER: Lectin-Glycan Complexes: A Comprehensive Analysis of Docking Calculations

Lectins are a type of glycan-binding protein that noncovalently bind glycans. Carbohydrates are molecules consisting of sugar units joined together. Glycans are carbohydrates. Hence, glycans are also sugars. Lectins and lectin-glycan complexes have a range of biological roles and can be found in animals (including humans), plants, bacteria, viruses, and yeasts and fungi.[1] Many scientists focus on the computational study of these complexes due to their intricate roles in many living organisms. Computational study is important in furthering our knowledge of lectin-glycan complexes and other such protein complexes. However, computational study is not perfect. There are many challenges in computational study, especially in the docking of ligands to receptors. In glycans, there are usually a high number of hydroxyl (OH) groups that affect docking; there could be surrounding ions; the rings in glycans can have CH-π stacking interactions with aromatic residues and cause issues; and glycans larger than one subunit have bonds between subunits that allow them to twist. These represent just a few challenges in docking. It is difficult for the software to accurately dock glycans to corresponding receptors because of these challenges. So, the purpose of this study was to try to evaluate the performance of the docking program Autodock Vina (referred as Vina) and Vina-carb on a large dataset of docking problems and propose a workflow for effective docking of glycan ligands.[2] We have looked into the effect of glycan size, seed value (a random starting point for docking calculations), and Carbohydrate Intrinsic (CHI) energy functions in glycosidic linkages.[3] We tried using CHI values in Vina-Carb that mimicked Autodock Vina. We saw that in almost all cases Vina-Carb did better, even if it was a marginal difference. Then we tried optimizing the CHI values CHI coefficient and CHI cutoff.[3] We did see some patterns emerge for specific values. We also used a random seed for calculations but did not see much of a difference in using a random seed for calculations. There were some improvements and surprises. Overall, we know that optimizing docking software is a challenge, but doing so will improve research for many scientists. More calculations will be done in the future because they will be worthwhile. We originally sought to analyze Vina-Carb to make it better. Such research will help improve future computational study. We have already seen some parameters that have promise for further investigation.

PREFACE

I learned a lot during my time in the Doerksen lab. I never thought I would study computational chemistry, but here we are. I am very proud of the work I have accomplished and this thesis I have written.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Terms. Before delving into the study, some basic definitions are important. Carbohydrates are made up of sugar subunits connected together. Glycans are sugar molecules, and sugar molecules are carbohydrates. Therefore, glycans are also carbohydrates. The terms "glycan" and "carbohydrate" can be used interchangeably, but for the purpose of this study, we will refer to them as glycans. The term "ligand" is also synonymous for "glycan" in this paper, as glycans are the only ligands being studied in these experiments. Lectins are a specific type of glycan-binding protein that are very sensitive to binding to sugar molecules, such as fucose. In fact, they are often classified based on the carbohydrates to which they bind.[1] Lectins may have one binding site or multiple. Metal ions such as calcium and magnesium are often present as well and can affect the affinity and binding of these ligands to the lectins.[1] Glycoproteins covalently bind carbohydrates, such as glycans, forming chemical bonds to their amino acid side chains. This means there is a physical bond between the ligand and the receptor, and this process is known as glycosylation. However, glycan-binding proteins bind to carbohydrates non-covalently and reversibly.[1] Glycan-binding proteins, especially L-fucose-binding lectins, are the focus of this study. Figure 1 below shows an example of an L-fucose lectin-glycan complex. In green is the lectin PDB ID: 1FWU sourced from the RCSB Protein Data Bank (PDB).[4]

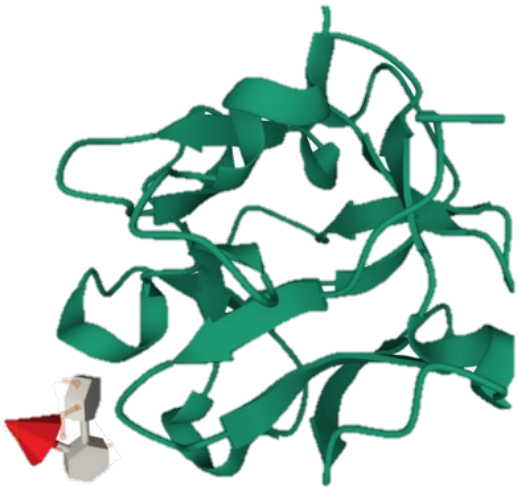

Figure 1. The lectin PDB ID: 1FWU sourced from the RCSB Protein Data Bank (PDB).[20] Also shown is the carbohydrate non-covalently bound to PDB ID: 1FWU. The lectin is shown in green and the carbohydrate is shown in gray and red in SNFG (NCBI Symbol Nomenclature for Glycans) form.

1.2 Biological roles. Lectins are present in a variety of organisms and microorganisms including animals, plants, bacteria, viruses, yeasts, and fungi.[1] They engage in both cell-molecule and cell-cell interactions.[5] Some lectins depend on the presence of metal ions such as calcium and magnesium in order to effectively bind carbohydrates.[1] Lectins form lectin-glycan complexes with sugars, but they often also aid in agglutination, or clumping, of cells by attaching to a carbohydrate on each cell.[1,5]

Cell signaling is an especially important and prominent mechanism, and lectins often do this in other organisms, such as animals. Lectins are present in vertebrates and invertebrates alike, including humans. Integral lectins are present in membranes as structural components while soluble lectins are present in intracellular and intercellular fluids.[1] This means lectins can be found practically anywhere! In animals, lectins are often used in endocytosis – the process of allowing materials into a cell – the transport of glycoproteins, apoptosis – one kind of cell death

– cell adhesion, migration of cells, and the immune response, including the capture of viruses for degradation.[1,6] Plants often use lectins to preserve nitrogen, to regulate cell signaling, and as defense mechanisms against various outsiders such as bacteria, viruses, and insects.[1,7] In fungi, yeasts, bacteria, and viruses, lectins are often used to interact with host cells, recognize immune processes, and for phagocytosis – essentially eating another cell – cell adhesion, and symbiotic processes.[1,6] Fungal lectins have broad carbohydrate specificity, ranging from simple sugars to glycoproteins to more unusual glycans that other lectins do not often bind to; this makes these lectins quite flexible with a range of functions including mycorrhization, which is the symbiosis between the fungus and the roots of a plant.[8] Lectins on the surfaces of bacteria, however, are often used to initiate infection of host cells.[9] Similarly, viruses often use glycan-lectin interactions in order to attack and enter host cells.[6]

1.3 Computational study. With so many functions and mechanisms across a broad range of organisms, docking is a great tool in computational and biochemical research. Computational research in and of itself is a great verifiable tool in research as it is often faster, uses less resources, and is great at predicting models such as glycan-lectin complexes.[10] Lectin research can also be applied to more areas of science including molecular biology, pharmacology, immunology, medicine, and clinical analysis, including drug design.[1,2] Lectins can be found in several databases including the RCSB PDB and UniLectin. There, you can find a lectin's structure, chemical components, and biological significance. This information is very useful when considering computational study to aid in the understanding of lectin-glycan interactions. Reported experimental or computed structures can be downloaded and loaded into software such as PyMOL for viewing and performing calculations.[2] Within this program, you can get a 360˚ view of the structure and see exactly where and how the ligand(s) bind to the lectin. You can also

see if there are any ions present around the binding site(s) and also see what type of solution the lectin was found in. You can even load more than one complex into PyMOL at one time. This allows the user to see how multiple complexes interact with one another. Other software such as AutoDock Vina and Vina-Carb can be added to PyMOL to run calculations on the complex, such as how well the software predicts docking of the ligand(s) to the lectin. Docking is a quick way to see the interactions between the glycan and the lectin and to determine the key factors present in the binding itself as well as in the binding site.[11] However, there are a number of unusual features in lectin-glycan complexes that make it difficult to model these interactions computationally: large ligand size (often greater than four subunits in size); high number of OH groups on the ligands; CH-π stacking interactions between glycans and aromatic amino acids in lectins (tryptophan, tyrosine, phenylalanine, and histidine); and torsional glycosidic linkages.[12] The softwares are competent at predicting the binding of the glycan(s) to the lectin, but they do not perform these calculations perfectly. It is important to note that docking software may work better for proteins and small molecules other than carboproteins and sugars because the software was developed and optimized for these particular kinds of molecules. In this study, we aim to see how well docking software Autodock Vina and Vina-Carb behave when docking glycans to lectins, and we seek to optimize them to get better docking, especially with ligands with the challenges discussed above.

2 Initial data collection and organization

To begin understanding how sugars bind to glycan-binding proteins, I downloaded lectin-glycan complexes from the database called UniLectin.[13] This database is much newer than the esteemed RSCB PDB, but it still contains viable molecules to study. Moreover, the molecules listed in UniLectin can also be found in the RCSB PDB. UniLectin was used instead of the

RCSB PDB because we could easily find and categorize the lectin-glycan complexes with UniLectin whereas the RCSB PDB would require a more extensive search simply due to its size. We particularly focused on fucose binding lectins due to the relatively large number of the known fucose-binding lectins and the difference in the nature of binding among them. Some L-fucose binding lectins are known to have highly electrostatic binding sites (for example PA-IIL, LecB), but some have CH-π stacking derived binding where very few polar interactions are essential for binding. Lectin RSL from the bacterium *Ralstonia solenerserum* is one such nonpolar example.[14]

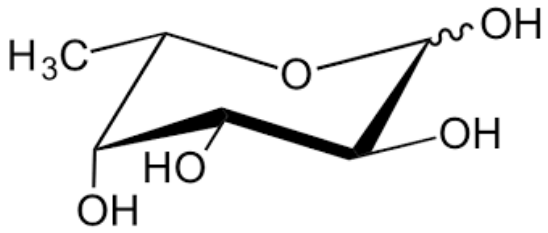

Figure 2. Monosaccharide L-fucose. The squiggled line shows the hydroxyl group that is attached to the anomeric carbon, denoting either alpha or beta orientation. This anomeric carbon denotes the L enantiomer. For the purposes of this study, L-fucose binding lectins were studied rather than D-fucose-binding lectins.

The L orientation means the molecule rotates plane polarized light counterclockwise, or left-handedly. This is in contrast with the D enantiomer which rotates plane polarized light clockwise, or right-handedly. "L-fucose binding" does not mean these lectins only bind to the monosaccharide L-fucose, but rather, they are very specific and highly selective towards an L-fucose subunit within a glycan molecule.[15] The can often bind to some other similar monosaccharides but with relatively weak affinity. There were 246 L-fucose binding lectin-glycan structures present in the UniLectin database, but some of the complexes were the same,

just listed at different X-ray crystallography resolutions. So, I removed the repeated lectins at the lower resolution, giving a total of 220 lectins for analysis. To analyze the proteins, I used a protein modeling software known as PyMOL, and I stored data in a Linux system. In PyMOL, I accessed the RCSB PDB to download the lectins.[16] I used an Excel spreadsheet to organize information on each lectin for easy reference throughout the study. The lectins were organized alphabetically based on their four-character PDB ID. With their ID, other information was noted including their origin, the lectin family, the resolution in Ångstroms (Å), the condensed International Union of Pure and Applied Chemistry (IUPAC) name of the glycan, the size of the ligand, whether or not any ions were present, whether or not any mutations were present, and any other distinctions to note. There were five origins of lectins studied: animal, bacterial, fungal and yeast, plant, and virus. Fungal and yeast were grouped together due to their biological similarity. There were several types of families within these origins. Table 1 is an example of a lectin and its information input in the spreadsheet (see Appendix A for more details):

Table 1. An example of the categorical information obtained for all 220 lectins. 2JDP is a bacterial lectin belonging to the family LecB/PA-IIL, RSIIL. It was obtained at a resolution of 1.3Å. It binds to the monosaccharide fucose, and it is surrounded by eight calcium ions. Lastly, it has an S23A mutation.

PBD ID	Origin	Family	Resolution (Å)	IUPAC Condensed	Ligand Size	Ion	Mutation	Other
2JDP	Bacterial	LecB / PA-IIL, RSIIL	1.3	Fuc	Monosaccharide	8Ca	Mutant S23A	None

The same information was input for all 220 complexes. The most useful tabulated information was origin, ligand name, ligand size, and ion presence. These formed the bases of

the organization and analyzation of our calculations. Table 2 represents the breakdown of all of the lectins in the different categories:

Table 2. An example of the categorical information obtained for all 220 complexes. The origins are in the left-most column with the number of lectins from each origin, and then the top row breaks down all 220 lectins into ion presence and ligand size – mono for monosaccharide, di for disaccharide, tri for trisaccharide, tetra for tetrasaccharide and oligo for oligosaccharide. For example, there are 22 animal lectins in total, 19 of which contain ions, 3 of which do not contain ions, etc.

Categories							
	Ion	*No Ion*	*Mono*	*Di*	*Tri*	*Tetra*	*Oligo*
Animal							
22	19	3	5	1	9	4	3
Bacterial							
89	59	30	47	1	9	14	18
Fungal and Yeast							
28	10	18	12	2	5	6	3
Plant							
10	10	0	2	0	4	2	2
Virus							
71	25	46	3	1	43	17	7
Total							
220	123	97	69	5	70	43	33

Figure 3 contains pie charts to aid in visualizing the breakdown of the lectins into categories.

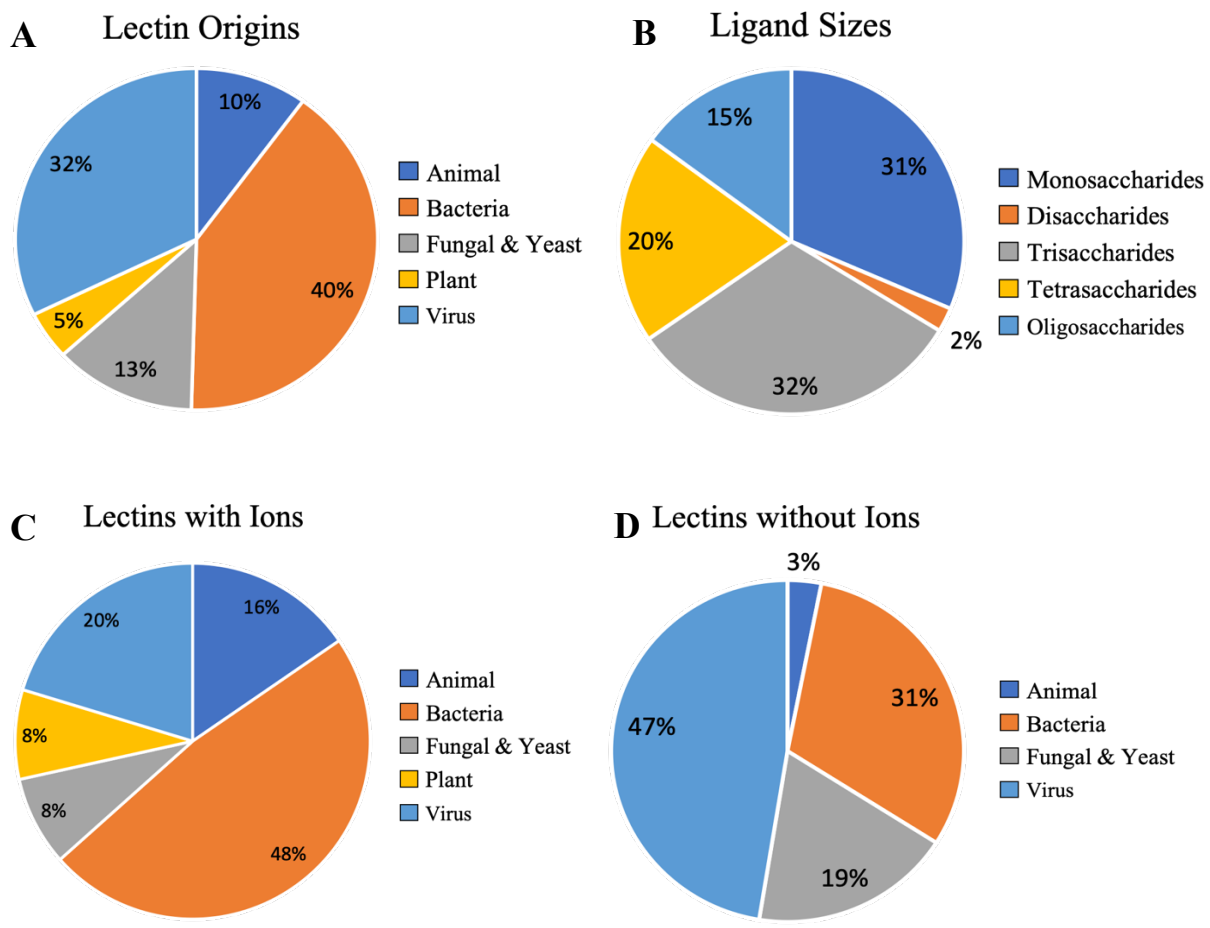

Figure 3. The pie charts show the breakdown of all the 220 complexes in different categories: origin, ligand size, and ion presence. Pie chart A shows the different origins of 220 complexes. Pie chart B shows the sizes of ligands bound to the lectins, from monosaccharides to oligosaccharides (greater than four subunits in this case). Pie chart C shows the breakdown of lectins with ions based on origin. There were 123 complexes out of 220 that contained ions. Pie chart D shows the breakdown of lectins without ions based on origin. There were 97 complexes that did not contain ions. It should be noted that all plant complexes in this study contained ions.

Although we did not analyze the results based on origin, we did focus on ligand size and ion presence to draw some of our conclusions, as these characteristics affected binding affinity

and docking the most. The purpose of Table 2 and Figure 4 is to simply display the types of lectins we tested, and we analyzed a large variety of lectins in this study. A more specialized study looking at one type of lectin, such as virus lectins, may produce slightly different results than what is shown in this report.

3 Process, results, and discussion

3.1 AutodockVina and Vina-Carb. Vina is the latest docking software from AutoDock Suite that is the computationally fastest, most accurate, and well-cited software employed in protein-docking.[17] However, the scoring function of this program has been trained on mainly non-glycan like ligands; therefore, it is not proficient in docking glycans.[18] Because of its limitations in predicting a correct glycosidic linkage conformation of oligosaccharides, the Woods group (University of Georgia) introduced a CHI energy function to the Vina scoring function and released a program called Vina-Carb. CHI-energy functions assign relative energies to the dihedral angles of the glycosidic linkages based on the distribution of glycosidic dihedral angles found in the glycans that participate in protein-glycan complexes. Thus, two new parameters called CHI-energy weight term (*chi_coeff*) and CHI-cutoff term (*chi_cutoff*) were introduced. The *chi_coeff* affects the CHI-energy penalties provided the magnitude of the CHI energies is more than the *chi_cutoff* value. Therefore, CHI-energy penalties are ignored below the specified value of the cutoff (*chi_coeff*).[3]

The first part of our study consisted of docking the specific ligands to all 220 complexes obtained from the UniLectin database and the RCSB PDB.[13,16] The ligands docked are not rigid but instead change into their conformers to fit into the docking region with the lowest amount of energy.

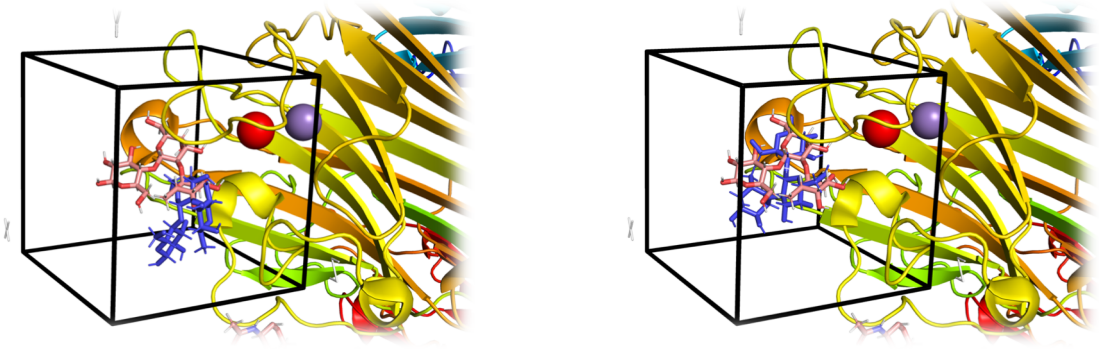

Figure 4. Lectin PDB ID: 1QOT (sourced from the RSCB PDB). (A) represents the very first pose AutodockVina calculated, while (B) represents the 16th and best pose the software calculated. This shows the software does not necessarily dock the lowest scoring, or best, pose with the first calculation. Pose 16 shown in (B) may not even be the very best pose Vina-Carb could have produced, but pose 16 is the best it produced out of n=20 poses.

As noted in Figure 4, Autodock Vina can produce a very good docking pose, as measured by root mean square deviation (RMSD) relative to the experimental lectin-glycan structure, but it may not be the first pose (as top-ranked by the scoring function) it presents. Therefore, several calculations need to be run in order to find a low RMSD pose. But how many calculations would that take – 20, 50, 100, 1000? Even with 1000 calculations, Autodock Vina may not give the best pose it can possibly produce. The program may not be optimal for the subject under study. Vina is generally a very good software used for molecular docking, but the mathematical functional form of the program is limited and may not apply well to lectin-glycan complexes. There could also be issues with the physical effects, such as hydrogen bonding and covalent bonding, as well as the training set data used to guide the choice of parameters. How do we fix this then? We will talk about optimizing Vina-Carb later in this report. For now, it is good to recognize this challenge and know that computational study is not perfect and can always be improved.

First, we wanted to compare the docking results obtained with Autodock Vina and Vina-Carb. It should be noted that we did not actually use Autodock Vina. Instead, we changed the carbohydrate intrinsic (CHI) values in Vina-Carb to match Autodock Vina where AutodockVina has a *chi_coeff* of 0 and a *chi_cutoff* of 12 and where Vina-Carb has a *chi_coeff* of 1 and a *chi_cutoff* of 2. Our hope was that Vina-Carb would dock glycans better than Autodock Vina. The comparison of the two is presented in Figure 6 below.

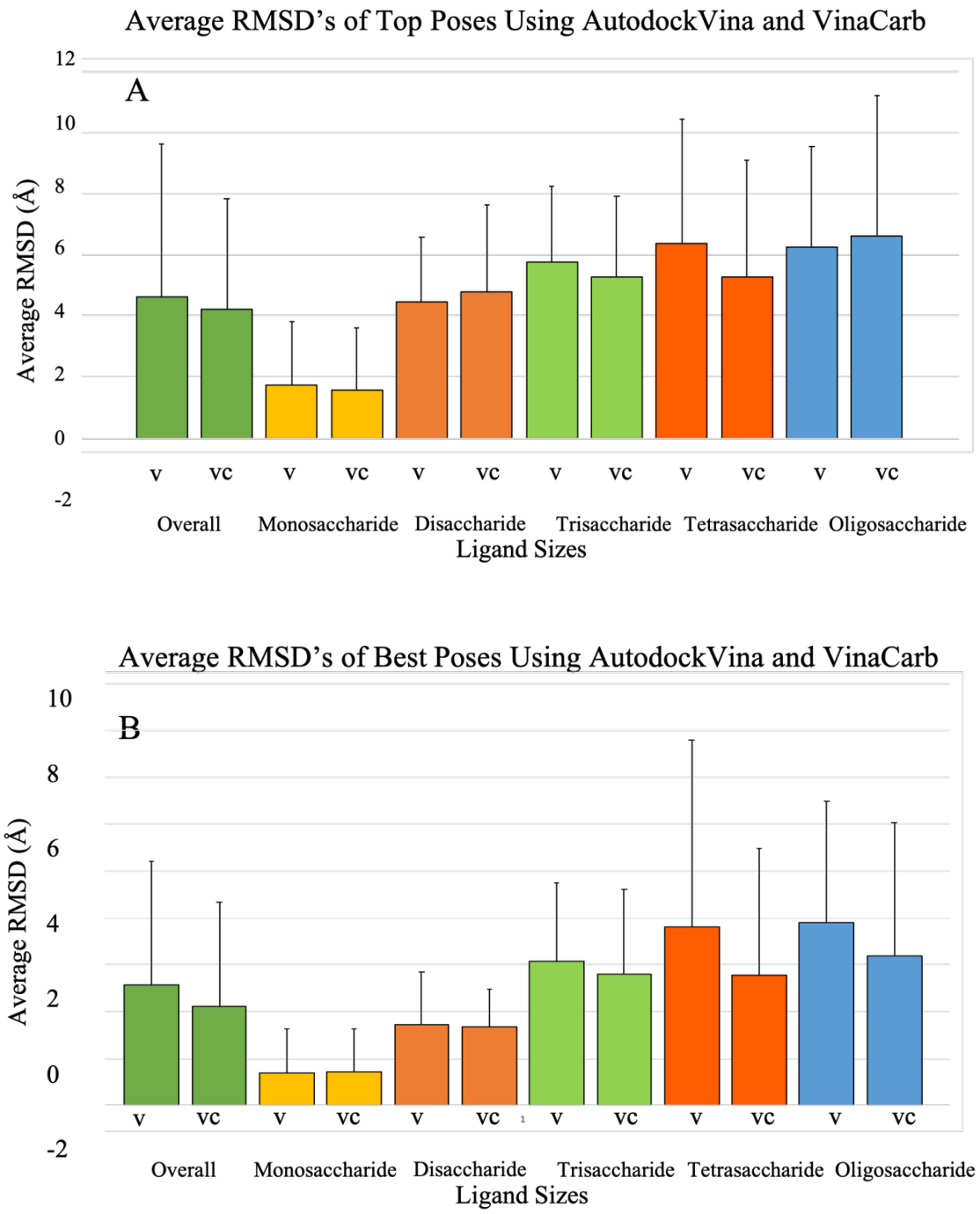

Figure 5. Average RMSD's of all of the different ligand sizes docked using parameters for Autodock Vina (v) and Vina-Carb (vc). For poses n=20, we took the average RMSD's of the first poses as well as the average RMSD's of the best, or lowest scoring, poses. Parameters representing vc generally resulted in better docking but not always. We expected vc to dock much better than v every time, but this was not the case.

As can be noted in the graphs, we docked each ligand with each software – Autodock Vina and Vina-Carb – and produced n=20 poses for each. We took the RMSD of all of 20 docking conformations from their crystal structure binding mode and reported the RMSD of the top scoring docking pose (docked first with the lowest amount of energy) and the best scoring docking pose (lowest RMSD) for each type of ligand. Ideally, the top scoring docking solution should have the lowest RMSD from the crystal structure, but in reality, scoring functions within the docking programs often fail to rank the poses in their correct order of accuracy. We observed that both Vina and Vina-Carb docked monosaccharides very well, showing an average RMSD under 2 Å, but show poor performance for larger glycans – di-, tri-, tetra-, and oligosaccharides. This is likely due to the programs' inability to predict correct torsional angles of the glycosidic bonds between monosaccharides. It should also be noted that the standard deviation (shown as error bars) of the RMSD values for all the lectin-glycan complexes are quite large in some instances. This is because in some cases, both software docked a ligand very well and in others very poorly. Vina-Carb is expected to predict the correct torsional angles along the glycosidic linkages due to CHI-energy terms, but interestingly, this was not always the case. The improvement in prediction accuracy compared to Vina was suboptimal. Vina-Carb (vc) did dock better in some instances, but Autodock Vina (v) docked better or about the same in more cases. We are not sure why this occurred, so we decided to run more calculations, starting with a random seed, since the default values of the *chi_coeff* and *chi_cutoff* did not show any significant improvement. In the performance of the docking software, these two parameters were optimized on selected protein-glycan complexes.

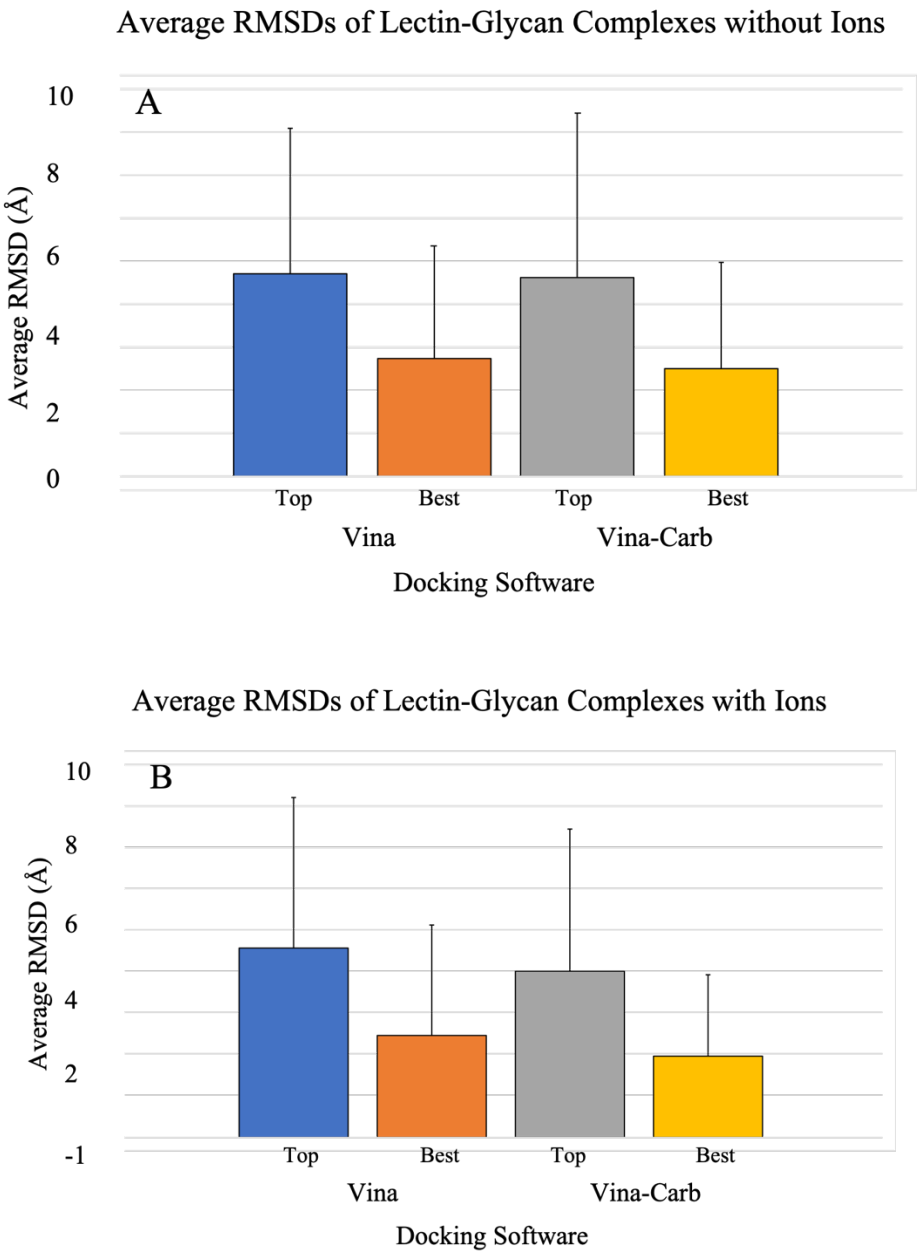

Figure 6. Average RMSD values of the lectin-glycan complexes without ions and with ions; (A) shows the complexes without ions and (B) shows the complexes with ions.

Looking at Figure 6, it appears that for each average RMSD, glycans docked better in complexes with ions than without ions. The difference is marginal but it is there nonetheless. It is especially present in Vina-Carb, suggesting that Vina-Carb docks glycans better in lectins both

with and without ions than Autodock Vina. However, both Autodock Vina and Vina-Carb dock glycans better in lectins with ions in or near the docking sites. This suggests that ions may have a stabilizing effect on the complex and may make it easier for these software to do the docking well. See Appendix B for further details.

3.2 Optimization. There was little difference between Autodock Vina and Vina-Carb from the previous set of calculations. Next, we looked to the CHI values of the docking software.[3,19] We decided to change these in small and large increments to see if there was a significant difference in docking. For all five disaccharides in this study, we used settings of 0.25, 0.5. 1, 2, 3, 4, 5, 10, and 50 for *chi_coeff*; and we used values of 0, 1, 2, 3, 4, 5 for *chi_cutoff*.[19] For example, we paired up the values as (0, 0.25), (0, 0.5), (0, 1), (0, 2), (0, 3), (0, 4), (0, 5), (0, 10), and (0, 50), for a total of 9 calculations in that one sequence per complex. There were 54 calculations in total per complex. The heat maps in Figure 8 show the combinations that were tested.

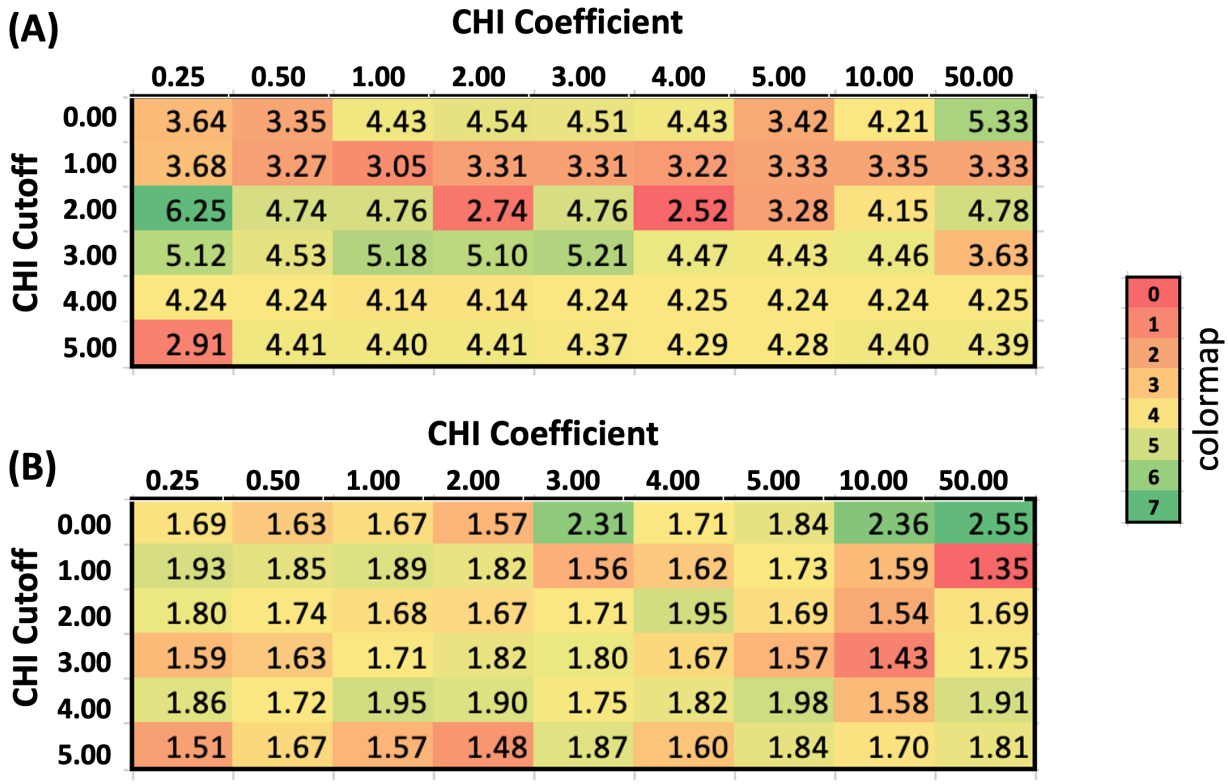

Figure 7. Heat maps showing the average RMSDs of optimized CHI values for five disaccharides. Heat map A represents the average RMSD values of the top scoring poses while heat map B represents the average RMSD values of the lowest scoring RMSD poses. There seems to be much variation within CHI value sets, but much consistency. Clearer patterns may be recognized with a larger data set. More variation was present in the first or top poses while the best poses had more consistent RMSD values across the entire map.

The red backgrounds indicate the lowest RMSD values while the green backgrounds indicate the largest RMSD values. The first heat map (Fig. 7A) represents the RMSD values of the first poses of all five disaccharides while the second heat map (Fig. 7B) represents the RMSD values of the lowest scoring poses of all five disaccharides. There was much variation with CHI Coefficient, but there was a good amount of consistency with a CHI Cutoff of 1 and 4, as seen in (A). All of the CHI Coefficient and Cutoff values did reasonably well in (B). It should be noted

that these calculations were run on only five disaccharides. Clearer and more reliable patterns may arise with calculations including more complexes as well as with a variety of kinds of complexes.

3.3 Random seed values. Instead of starting from the same starting point each time, next we decided to begin each calculation with a random seed value. Random seeds are essentially a pseudo-random number used by Vina for generating docking results.[20] To reproduce the same results in multiple independent runs, it is required to have the same random seed value. Therefore, we looked into the effect of the random seed in the docking solution by repeating a docking calculation 1000 times per complex with a different seed value while keeping all other parameters constant. For each size of ligand from previous calculations – monosaccharides, disaccharides, trisaccharides, and tetrasaccharides – we picked two lectin-glycan complexes for which the resulting docked pose was accurate and two complexes for which the docking program could not dock the glycan correctly. It should be noted that there were only five disaccharides out of 220 lectins to choose from, so this graph appears the most different. It would be useful to have a larger pool of disaccharides, but such a pool was not available in the RCSB PDB and UniLectin.[13,16] For each complex, we ran 1000 calculations, with each calculation beginning with a random seed value. We calculated the RMSD of the docking poses from each of the 1000 calculations and created a box and whisker plot to see the outliers. See Table 3 and Figure 9 below.

Table 3. Statistics of the random seed calculations for all lectin-glycan complexes tested in monosaccharide, disaccharide, trisaccharide, and tetrasaccharide pools. The information is further visualized in Figure 8 below.

Statistics of the Random Seed Calculations						
			Min	Max	Mean	Standard Deviation
Monosaccharides	1OFZ	Top	0.12	5.69	0.35	0.58
		Best	0.12	5.69	0.33	0.44
	1UZV	Top	0.19	0.19	0.19	0
		Best	0.19	0.19	0.19	0
	1KWW	Top	7.79	7.84	7.82	0.0091
		Best	0.53	7.08	1.71	1.31
	3ZI8	Top	0.46	10.43	7.34	3.89
		Best	0.36	5.91	1.38	0.75
Disaccharides	4AGT	Top	2.58	4.77	3.96	0.93
		Best	0.65	4.75	2.79	0.69
	6A87	Top	1.32	8.90	5.23	2.22
		Best	1.04	7.76	4.44	1.67
	2JDH	Top	3.38	3.99	3.75	0.09
		Best	0.25	3.90	2.17	0.96
	3VV1	Top	0.43	10.53	4.06	3.72
		Best	0.36	7.69	1.21	1.41
Trisaccharides	1LU2	Top	0.52	0.55	0.53	0.0034

		Best	0.52	0.55	0.53	0.0034
	2LIQ	Top	0.34	8.33	1.88	2.29
		Best	0.34	6.41	1.60	1.70
	2WKK	Top	6.99	10.4	10.11	0.23
		Best	0.91	8.96	4.43	1.53
	6HA0	Top	2.59	10.86	10.12	2.07
		Best	1.72	9.52	3.65	1.37
Tetrasaccharides	5AJB	Top	0.64	7.06	0.75	0.48
		Best	0.64	4.53	0.73	0.23
	5I4D	Top	0.34	8.94	0.83	1.37
		Best	0.34	7.85	0.74	1.01
	3ZW1	Top	18.01	21.37	18.21	0.29
		Best	9.83	19.16	16.88	1.64
	6Q05	Top	5.69	10.38	8.69	0.84
		Best	2.61	9.02	5.86	1.01

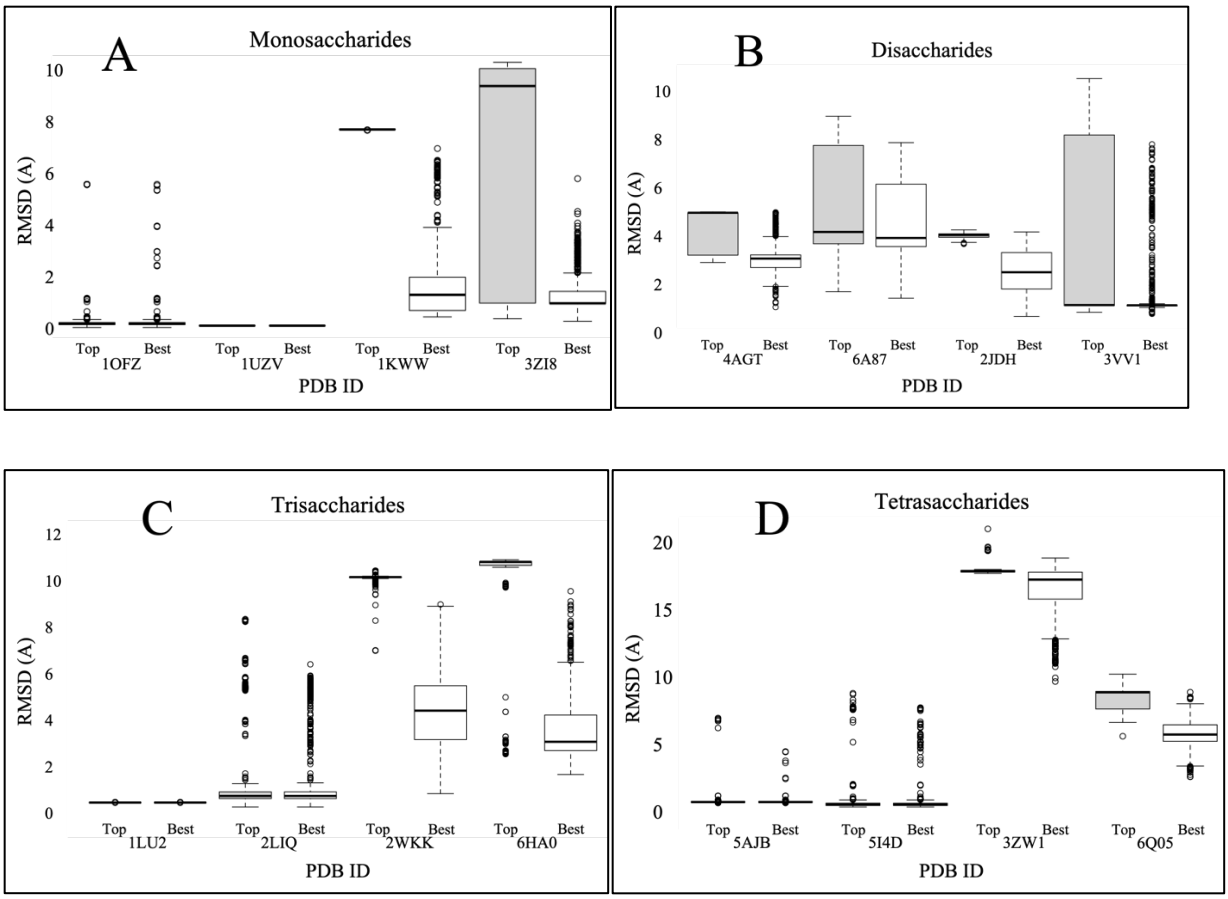

Figure 8. The RMSD (Å) values of 1000 docking solutions for selected lectins binding to (A) monosaccharides, (B) disaccharides, (C) trisaccharides, and (D) tetrasaccharides. A different random seed was used in each docking run. In each graph, the two lectins on the left (first four bars) represent cases that exhibited very good docking (low RMSD) on the first pose (best-scored with lowest energy) from previous calculations. The two lectins on the right (last four bars) represent cases that exhibited very poor docking on the first pose from previous calculations. For each lectin, the first bar represents the RMSDs of the top scoring poses from all 1000 calculations, and the second bar represents the RMSDs of the best docked pose from the top 20 docking solutions. The tops of the bars represent the highest RMSD value; the bottoms of the bars represent the lowest RMSD value; and the lines in the middle of the bars represent the average RMSD value.

As shown in the graphs and table, the ligands that docked well had a very small range of error and low RMSDs while the ligands that did not dock well in previous calculations typically had a larger range of error and higher RMSDs. These results did not differ much from the calculations originally made without a random seed value. Therefore, we could not find a significant difference in using a random seed value for the calculations versus allowing the ligands to begin at the original starting point from the PDB download. There also needs to be a normalization of the data because the sizes of the pools of ligands vary – 69 monosaccharides, 5 disaccharides, 70 trisaccharides, and 43 tetrasaccharides. We have not figured out a way to normalize this data yet to present it so that it reflects the pool size.

4 Conclusions

Docking softwares are great tools to use in the computational study of lectin-glycan complexes. However, they are not perfect. In this study, we have tried to gain a better understanding of the Autodock Vina software. First, we attempted to change the CHI values so that Vina-Carb mimicked Autodock Vina. In our comparison of the two, Vina-Carb did not do as well as we had hoped. Then, we tried optimizing the software by changing the CHI values. Some patterns did emerge. More patterns would likely develop with a larger sample size. Then, we tried using a random seed in our calculations to give the ligands a random starting point. These results were more unclear due to the pool size of the ligands. After all of this, we know that docking software is good, but it can be better. However, finding ways to optimize it successfully has proven to be quite a challenge. We set out analyzing the docking software Vina-Carb in an attempt to make it better. We have found some parameters that need further exploration. Optimizing docking software like Vina-Carb will be a great aid in computational docking and will broaden the prospects of computational research

5 Further study

There are more ways to try to optimize Autodock Vina. We could continue to change the CHI values with a larger sample size to see if patterns emerge. We could change the random seed value with a larger sample size. We could even look at different docking software to see how they compare to Autodock Vina and Vina-Carb.

6 References

(1) Santos, A. F. S. Lectins: Function, Structure, Biological Properties and Potential Applications. 22.

(2) Seeliger, D.; de Groot, B. L. Ligand Docking and Binding Site Analysis with PyMOL and Autodock/Vina. *Journal of Computer - Aided Molecular Design* **2010**, *24* (5), 417–422. http://dx.doi.org/10.1007/s10822-010-9352-6.

(3) Nivedha, A. K.; Makeneni, S.; Foley, B. L.; Tessier, M. B.; Woods, R. J. The Importance of Ligand Conformational Energies in Carbohydrate Docking: Sorting the Wheat from the Chaff. *J Comput Chem* **2014**, *35* (7), 526–539. https://doi.org/10.1002/jcc.23517.

(4) Protein Data Bank, R. P. D. RCSB PDB: Homepage https://www.rcsb.org/ (accessed 2021 -04 -08).

(5) Sharon, N. History of Lectins: From Hemagglutinins to Biological Recognition Molecules. *Glycobiology* **2004**, *14* (11), 53R-62R. https://doi.org/10.1093/glycob/cwh122.

(6) Van Breedam, W.; Pöhlmann, S.; Favoreel, H. W.; de Groot, R. J.; Nauwynck, H. J. Bitter-Sweet Symphony: Glycan–Lectin Interactions in Virus Biology. *FEMS Microbiol Rev* **2014**, *38* (4), 598–632. https://doi.org/10.1111/1574-6976.12052.

(7) Chrispeels, M. J.; Raikhel, N. V. Lectins, Lectin Genes, and Their Role in Plant Defense. *Plant Cell* **1991**, *3* (1), 1–9.

(8) Khan, F.; Khan, M. Fungal Lectins: Current Molecular and Biochemical Perspectives. *International Journal of Biological Chemistry* **2011**, *5*, 1–20. https://doi.org/10.3923/ijbc.2011.1.20.

(9) Sharon, N. Bacterial lectins, cell-cell recognition and infectious diseases. *FEBS Letters* 1987, 217, 1873-3468. https://doi.org/10.1016/0014-5793(87)80654-3.

(10) Purchase, S.; Denize, S. Using Docking/Replication to Verify and Validate Computational.

(11) Neumann, D.; Lehr, C.-M.; Lenhof, H.-P.; Kohlbacher, O. Computational Modeling of the Sugar–Lectin Interaction. *Advanced Drug Delivery Reviews* **2004**, *56* (4), 437–457. https://doi.org/10.1016/j.addr.2003.10.019.

(12) Rentzsch, R.; Renard, B. Y. Docking Small Peptides Remains a Great Challenge: An Assessment Using AutoDock Vina. *Briefings in Bioinformatics* **2015**, *16* (6), 1045–1056. https://doi.org/10.1093/bib/bbv008.

(13) Unilectin. UniLectin interactive lectin database https://www.unilectin.eu/ (accessed 2021 -04 -08).

(14) Mishra, S. K.; Adam, J.; Wimmerová, M.; Koča, J. In Silico Mutagenesis and Docking Study of *Ralstonia Solanacearum* RSL Lectin: Performance of Docking Software To

Predict Saccharide Binding. *J. Chem. Inf. Model.* **2012**, *52* (5), 1250–1261. https://doi.org/10.1021/ci200529n.
(15) Garber, N.; Guempel, U.; Gilboa-Garber, N.; Doyle, R. J. Specificity of the Fucose-Binding Lectin of Pseudomonas Aeruginosa. *FEMS Microbiology Letters* **1987**, *48* (3), 331–334. https://doi.org/10.1111/j.1574-6968.1987.tb02619.x.
(3) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(17) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(18) Tanchuk, V. Y.; Tanin, V. O.; Vovk, A. I.; Poda, G. A New, Improved Hybrid Scoring Function for Molecular Docking and Scoring Based on AutoDock and AutoDock Vina. *Chemical Biology & Drug Design* **2016**, *87* (4), 618–625. https://doi.org/10.1111/cbdd.12697.
(19) Nivedha, A. K.; Thieker, D. F.; Makeneni, S.; Hu, H.; Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.* **2016**, *12* (2), 892–901. https://doi.org/10.1021/acs.jctc.5b00834.
(20) Mitchell, J.; Rosen, J.; Phillips, A.; Eyck, L. *Coupled Optimization in Protein Docking*; 1999; p 284. https://doi.org/10.1145/299432.299502.

Appendix A

The 220 lectin-glycan complexes in this study. An example of the layout is seen in Table 1 in the report.

PBD ID	Origin	Family	Resolution (Å)	UPAC	Ligand Size	Ion	Mutation	Other	X	Y	Z	Vina(1,2) Best Pose	Vina Best RMSD	Vina First RMSD	VC(0.12) Best Pose	VC Best RMSD	VC First RMSD
1FWU	Animal	Cys-rich domain man-receptor	1.9	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	None	None	None	-1.32	5.07	10.18	7	4.502987285	9.186597204	10	3.642305316	7.194524236
1G1S	Animal	P-selectin	1.9	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-6)[Gal(b1-3)]GalNAc	Oligosaccharide	2Na(+1), 2Sr(2+)	None	Carbonyl Sulfide (2/2O5)	62.64	112.99	29.85	20	4.25239358	5.681831409	14	4.59286716	5.499077964
1G1T	Animal	E-selectin	1.5	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	CaCl2(+)	None	None	34.42	117.57	38.83	6	4.945491839	5.201885858	6	4.664055914	5.273228199
1G5L	Plant	GSL-IV	2	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	CaMn	None	None	19.33	36.16	32.35	14	7.514498793	10.5001047	19	6.809435112	7.535582645
1G79	Plant	Ecorl and ECL	1.7	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	2(PO4(3-))	None	Carbonyl Sulfide (2/2O5)	46.83	27	-6.51	17	4.157034209	7.15498697	17	4.157034209	7.048981303
1GZT	Bacterial	LecB / PA-IIL, RSIIL	1.3	Fuc	Monosaccharide	8Ca (2+)	None	None	-2.6	16.59	36.06	1	0.725611426	0.725611426	1	0.745334805	0.745334805
1IUC	Animal	E-selectin	2.24	Fuc	Monosaccharide	H2SO4(-1)	None	None	25.88	8.69	38.18	1	0.395493748	0.395493748	1	0.387166631	0.387166631
1IXN	Plant	UEA-1	2.3	Fuc	Monosaccharide	4CaMn	None	None	34.7	48.94	-0.6	1	0.360196946	0.360196946	1	0.353202165	0.353202165
1K12	Animal	AAA	1.9	Fuc	Monosaccharide	HCl, Ca	None	None	-9.22	26.15	0.19	2	0.633861052	0.633861052	2	0.456800532	0.456800532
1KWW	Animal	MBP-A / MBL1	1.9	Fuc	Monosaccharide	10Ca, 3HCl	None	None	45.23	48.76	23.52	7	0.759307117	7.98553079	7	0.759307117	7.983457183
1LED	Plant	GSL-IV	2	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc	Tetrasaccharide	CaMn, H2SO4(-1)	None	None	31.06	14.44	31.23	1	0.51516965	0.51516965	1	0.529474405	0.529474405
1LGB	Plant	favin, LOL pea, lens	3.3	Gal(b1-4)GlcNAc(b1-2)Man(a1-3)[GlcNAc(b1-2)Man(a1-6)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)]GlcNAc	Oligosaccharide	CaMn	None	None	-15.1	28.82	76.11	19	6.737998919	10.17577596	18	4.916488145	9.434329872
1LGC	Plant	favin, LOL pea, lens	3.2	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-2)Man(a1-3)[GlcNAc(b1-2)Man(a1-6)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)]GlcNAc	Oligosaccharide	3CaMn	None	2-methyl-2,4-pentanediol (C6H14O2)	48.06	90.91	77.63	15	7.562827495	12.93439141	5	5.847012383	14.72345248
1LOZ	Plant	Dolichos biflorus DBL and DBS8	2.8	Gal/NAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	2CaMn	None	None	-5.78	32.54	27.71	1	0.538924437	0.538924437	1	0.530436907	0.530436907
1OFZ	Fungal and yeast	AAL	1.5	Fuc	Monosaccharide	None	None	None	42.69	61.64	46.54	1	0.270088713	0.270088713	1	0.235686939	0.235686939
1QXC	Bacterial	LecB / PA-IIL, RSIIL	1.2	Fuc	Monosaccharide	8Ca, H2SO4(-1)	None	None	39.33	52.92	-22.01	1	0.394321626	0.394321626	1	0.297914138	0.297914138
1QOT	Plant	UEA-2	3	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	4CaMn	None	None	78.47	77.76	29.91	16	1.827760322	5.862591066	8	1.620778656	6.962393334
1RDI	Animal	MBP-C / MBL2	1.8	Fuc	Monosaccharide	4Ca, HCl	None	None	3.47	-9.36	6.3	6	0.488206057	2.876528741	1	0.488206057	2.820668407
1RDJ	Animal	MBP-C / MBL2	1.8	Fuc	Monosaccharide	4Ca, HCl	None	None	3.33	46.4	0.94	9	0.654374128	2.483278617	9	0.654374128	2.380096882
1S15	Animal	DC-SIGN / DC-SIGNR	1.8	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-2)Man(a1-3)[Gal(b1-4)Glc(a1-3)]Man(b1-4)Glc	Oligosaccharide	2Ca, Mg	None	None	29.6	-7.61	-3.89	10	5.751893953	7.223600551	7	7.115029165	9.800316054
1S16	Animal	DC-SIGN / DC-SIGNR	2.25	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	2Ca	None	None	99.02	54.37	22.76	10	4.46832662	6.117024383	17	4.215836562	5.043406346
1UID	Fungal and yeast	coprinus galectin-2	2.1	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	None	None	None	-7.59	34.67	14.95	20	6.488366076	8.897022346	10	5.750486029	8.375000163
1ULF	Fungal and yeast	coprinus galectin-2	2.36	Gal/NAc(a1-3)[Fuc(a1-2)]Gal(b1-4)Glc	Tetrasaccharide	None	None	None	56.48	3.84	6.02	9	4.436090278	6.72300477	2	4.891430975	6.822774983
1UZV	Bacterial	LecB / PA-IIL, RSIIL	1	Fuc	Monosaccharide	8Ca, 2H2SO4(-1)	None	2 Glycerol (C3H8O3)	7.82	54.47	42.9	1	0.193346726	0.193346726	1	0.189275833	0.189275833
1W8F	Bacterial	LecB / PA-IIL, RSIIL	1.05	Gal(b1-4)GlcNAc(b1-3)Gal(b1-4)[Fuc(a1-3)]Glc	Oligosaccharide	8Ca, 4H2SO4(-1)	None	2 Glycerol (C3H8O3)	-1.97	26.65	-10.32	13	5.654593437	5.758231585	20	6.331686433	11.87592183
1WBH	Bacterial	LecB / PA-IIL, RSIIL	1.75	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	8Ca, 4H2SO4(-1)	None	3 Glycerol (C3H8O3)	36.63	13	6.12	6	0.869267871	4.57155861	6	0.869267871	4.549525134
2BOI	Bacterial	CV-IIL	1.1	Fuc	Monosaccharide	4Ca	None	None	34.82	20.5	21.65	1	0.411229182	0.411229182	1	0.403434389	0.403434389
2BS5	Bacterial	RSL BamDL	2.1	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	HCl	None	6 Glycerol (C3H8O3)	17.79	-6.39	39.72	9	3.909586701	7.214277486	16	2.292913946	8.00937932
2BS6	Bacterial	RSL BamDL	1.8	Fuc(a1-2)Gal(b1-2)Xyl	Trisaccharide	None	None	6 Glycerol (C3H8O3)	21.23	14.91	-3.87	1	1.806634537	1.806634537	1	2.004099057	2.004099057
2BT9	Bacterial	RSL BamDL	0.94	Fuc	Monosaccharide	None	None	None	-24.41	9.65	-8.38	1	0.278122994	0.278122994	1	0.272722034	0.272722034
2JDH	Bacterial	LecB / PA-IIL, RSIIL	1.1	Fuc(a1-4)Glc/NAc	Disaccharide	8Ca, H2SO4(-1)	None	None	-3.22	-17.72	9.61	4	0.555000411	3.832754591	3	1.281799173	1.469824643
2JDM	Bacterial	LecB / PA-IIL, RSIIL	1.7	Fuc	Monosaccharide	8Ca, 2H2SO4(-1)	Mutant S23A	None	4.43	-12.44	32.48	1	0.360213463	0.360213463	1	0.352629173	0.352629173
2JDP	Bacterial	LecB / PA-IIL, RSIIL	1.3	Fuc	Monosaccharide	8Ca	Mutant S23A	None	-32.6	-19.77	-13.62	1	0.405453635	0.405453635	1	0.539757446	0.539757446
2JDU	Bacterial	LecB / PA-IIL, RSIIL	1.5	Fuc	Monosaccharide	4Ca, 6H2SO4(-1)	Mutant G24N	2 Glycerol (C3H8O3)	-28.74	-18.08	17.21	1	0.430937026	0.430937026	1	0.42256852	0.42256852
2KMB	Animal	MBP-A / MBL1	2	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	10Ca, 3HCl	None	None	44.28	50.76	20.28	20	6.038056489	11.28177035	7	4.139288819	5.080765971
2LRI	Fungal and yeast	CCL2	0	Gal/NAc(b1-4)[Fuc(a1-3)]Gal/NAc	Trisaccharide	None	None	None	58.53	-10.9	20.65	1	0.948210142	0.948210142	1	0.835472754	0.835472754
2W7B	Animal	Sig-lec-8	0	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	4.47	-17.52	-22.68	2	2.049232764	2.105124793	1	2.329643595	2.329643595
2ZDL	Bacterial	LTB, CTB, CFXB	2.53	Gal/NAc(a1-3)[Fuc(a1-2)]Gal(b1-3)[Gal(b1-4)[Fuc(a1-3)]Glc	Oligosaccharide	None	None	None	15.23	-55.46	-15.55	1	1.709424551	1.709424551	1	1.055218918	1.055218918
2ZDS	Virus	Norovirus GII.4	2	Gal/NAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	19.56	9.59	-7.33	19	4.271578042	7.923306343	11	3.978613149	5.517484918
2ZDT	Virus	Norovirus GII.4	2	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	-18.71	-8.89	-8.08	14	6.339341875	8.257687832	6	4.184845612	8.057673862
2ZX9	Animal	SRCL / COLEC12	1.95	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	16Ca	None	None	79.7	81.38	44.99	8	4.565683986	5.981724468	3	4.002424532	4.548228711
2R61	Bacterial	SSL5	2.75	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	HCl	None	2 Glycerol (C3H8O3)	23.02	-4.49	37.31	15	1.881877748	8.314671778	2	1.833885786	2.282923933
2RDG	Bacterial	SSL11	1.6	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	2K, Potassium 2-Oxido-1,2,3-propanetricarboxylate (KC6H4O7)	None	None	25.84	6.12	4.14	6	3.976934871	8.404737973	7	0.987229697	10.59016079
2VUD	Bacterial	LecB / PA-IIL, RSIIL	1.7	Fuc	Monosaccharide	8Ca, 2H2SO4(-1)	None	None	31.37	-12.4	32.87	1	0.375315404	0.375315404	1	0.367413142	0.367413142
2WKK	Fungal and yeast	coprinus galectin-2	1.5	Gal(b1-4)Fuc(a1-6)GlcNAc	Trisaccharide	Mg	None	Glycerol (C3H8O3)	-30.83	0.1	20	1	5.471070105	10.47132735	7	6.064429904	10.08557626
2ZL7	Virus	Norovirus GI.1	1.35	Gal/NAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	Oligosaccharide 2Mg	None	Acetate (C2H3O2(-1)), Ca, 2Mg, 2C2H8O2(-1)	None	None	-5.4	4	5.176450286	6.852639441	9	4.300806998	6.767734803
2Z16	Virus	Norovirus GI.1	1.43	Fuc(a1-2)Gal(b1-3)[Gal/NAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	Acetate (C2H3O2(-1)), 2Mg	None	None	25.74	39.51	-19.13	10	8.747780158	10.21156086	4	8.533063701	10.79453363

3AP9	Animal	galectin-8	1.33	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	3HCl	None	2-methyl-2,4-pentanediol (C6H14O2)	21.48	21.59	41.15	5	0.936822044	8.621221315	1	0.939383173	0.939383173
3ASP	Virus	Norovirus GI 2	1.6	GalNAc(a1-3)[Fuc(a1-2)Gal(b1-3)GlcNAc	Tetrasaccharide	2Na	None	None	28.74	7.64	-25.91	14	5.417688891	7.527397435	14	6.006613352	7.972942446
3ASQ	Virus	Norovirus GI 2	1.6	Fuc(a1-2)Gal(b1-3)GlcNAc	Trisaccharide	2Na	None	None	31.28	5.49	-32.42	3	3.152572208	7.413023787	1	3.150730925	3.150730925
3ASR	Virus	Norovirus GI 2	1.6	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	2Na	None	None	-8.82	-30.8	-31.61	5	4.490965638	6.427450457	9	0.611032823	6.321658599
3ASS	Virus	Norovirus GI 2	1.6	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc	Tetrasaccharide	2Na	None	None	-10.72	29.21	22.75	12	19.96943465	20.05343459	3	0.542885948	7.797695981
3AST	Virus	Norovirus GI 2	1.4	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc	Tetrasaccharide	2Na	None	None	26.59	7.95	22.7	11	0.507477693	5.43598572	2	0.528996815	8.069069506
3CAH	Plant	SNA-II	1.55	Fuc	Monosaccharide	3C2H3O2(-1), 7H2SO4(-1)	None	None	27.2	7.83	1.02	2	0.560608192	3.285546617	2	0.560608192	3.210006712
3CQO	Animal	bass lectin	2.32	Fuc	Monosaccharide	6Ca, 3HCl	None	None	15.18	-1.76	38.83	3	0.234603744	3.174378734	3	0.234603744	3.107542223
3D26	Virus	Norovirus GI 1	2.3	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	102.68	96.45	-156.79	7	3.944378172	5.969122438	10	5.361906887	6.314771069
3DCQ	Bacterial	LecB / PA-IIL, RSIIL	2.13	Fuc	Monosaccharide	8Ca	None	None	43.16	53.81	-22.87	1	0.308289249	0.308289249	1	0.301798222	0.301798222
3E6J	Animal	Lamprey VLR	1.67	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	5H2SO4(-1)	None	Glycerol (C3H8O3)	2.53	18.46	5.64	3	0.516869815	5.612079352	3	0.502435232	5.593708043
3EF2	Fungal and yeast	MOA	1.8	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	16Ca, 4C2H3O2	None	None	-13.84	55.12	-73.6	2	1.407646952	1.673891881	2	0.779024767	1.553451963
3EFX	Bacterial	LTB, CTB, CFXB	1.93	GalNAc(a1-3)[Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]Glc	Oligosaccharide	None	None	None	16.09	-23.4	10.71	2	1.764037674	1.783418154	1	1.06573355	1.06573355
3KMB	Animal	MBP-A / MBL1	1.95	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	10Ca, 3HCl	None	None	47.02	49.58	21.7	14	6.139405724	6.826399436	20	1.553824459	7.789667365
3LEG	Bacterial	Lectinolysin	2.01	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	Ca, Ni	None	None	-3.95	-3.45	-19.52	10	3.063147713	3.402019773	9	1.808244511	3.384987881
3LEI	Bacterial	Lectinolysin	1.9	Fuc	Monosaccharide	Ca, Ni	None	None	4.5	7.28	-16.85	3	0.369488272	1.353596209	3	0.369488272	1.325096254
3ONY	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.85	Fuc	Monosaccharide	None	None	10 Ethylene Glycol (C2H6O2)	-6.99	7.77	-13.78	14	1.148562652	3.143819555	14	1.148562652	3.077626466
3P5G	Animal	langerin	1.6	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	4Ca	None	None	32.94	-7.09	39.46	18	2.968604767	7.564890127	12	2.972934948	7.071653744
3PA1	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.48	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	2 Imidazole (C3H4N2), 16 Ethylene Glycol (C2H6O2)	0.82	4.46	41.09	15	2.883698329	5.08158769	6	0.942574795	5.751547317
3PA2	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.48	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	1 Imidazole (C3H4N2), 22 Ethylene Glycol (C2H6O2)	2.67	2.64	40.4	1	0.777873998	0.777873998	3	4.14920893	6.258436852
3PUN	Virus	Norovirus GII 9	2.05	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	-36.41	-4.09	-16.51	5	0.903477525	5.328488868	1	0.874443568	0.874443568
3PVD	Virus	Norovirus GII 9	1.9	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	-37.55	-4.45	-14.72	20	6.758115086	9.0234044	5	3.01003866	7.980126151
3Q38	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.28	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	6 Imidazole (C3H4N2), 20 Ethylene Glycol (C2H6O2)	1.72	5.14	41.21	6	0.958989986	6.2976494	5	0.877306221	6.248047728
3Q39	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.25	Fuc(a1-2)Gal	Disaccharide	None	None	4 Imidazole (C3H4N2), 24 Ethylene Glycol (C2H6O2)	3.11	4.83	41.65	7	3.110178696	3.82757758	20	2.883018525	6.968164455
3Q3A	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.4	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	None	None	5 Imidazole (C3H4N2), 22 Ethylene Glycol (C2H6O2)	1.45	4.9	39.96	1	0.984083316	0.984083316	1	0.958386722	0.958386722
3R6K	Virus	Norovirus GII 1, GII 10, GII 12, GII 19	1.6	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	2 Ethylene Glycol (C2H6O2)	1.43	-53.59	-33.5	12	4.583757632	5.756755108	19	2.136363293	5.716461645
3SEJ	Virus	Norovirus GII 4	3.04	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	84.77	-23.15	25.14	20	5.393665718	6.844529733	4	5.356005872	5.925098084
3SLD	Virus	Norovirus GII 4	2.68	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	116.88	21.63	-84.18	4	1.922082865	2.253213978	1	2.23711408	2.23711408
3SLN	Virus	Norovirus GII 4	2.84	Fuc(a1-2)Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	-5.3	20.32	-89.44	3	6.121679832	6.978257033	2	6.123609173	6.938655201
3VV1	Animal	Nematode galectin-6	1.5	Gal(b1-4)Fuc	Disaccharide	3Mg	None	None	19.39	17.82	49.42	2	0.732356738	8.170539104	2	0.743073289	8.114225348
3WG3	Fungal and yeast	agrocybe ACG	1.35	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)GlcNAc	Tetrasaccharide	None	None	Dihydroxyethylether (C4H10O3), Ethylene Glycol (C2H6O2)	-2.71	-48.14	-3.08	8	0.898234106	3.728405345	2	0.919334777	5.895975028
3WG4	Fungal and yeast	agrocybe ACG	1.6	GalNAc(a1-3)[Fuc(a1-2)Gal(b1-4)GlcNAc	Tetrasaccharide	None	Mutant N46A	Dihydroxyethylether (C4H10O3), Ethylene Glycol (C2H6O2)	-3.17	-48.42	-3.61	13	1.816288426	8.743736292	2	1.135741064	8.236841478
3ZI8	Bacterial	RSL, BambL	1.5	Fuc	Monosaccharide	Mg	None	2-methyl-2,4-pentanediol (C6H14O2)	21.74	11.54	-18.39	4	1.033154393	9.742646248	4	1.033154393	9.518647426
3ZNM	Virus	influenza virus	2.4	NeuAc(a2-3)Gal(a1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	54.96	62.52	99.83	6	3.644276084	5.937591376	8	5.103641155	5.476431299
3ZW0	Bacterial	RSL, BambL	1.6	Fuc	Monosaccharide	None	None	None	-4.78	-1.82	3.95	1	0.246036759	0.246036759	1	0.240856458	0.240856458
3ZW1	Bacterial	RSL, BambL	1.6	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-3)Gal(b1-4)Glc	Tetrasaccharide	None	None	None	-3.61	-2.85	16.48	16	17.69896267	18.2252133	16	15.31924172	19.0614795
3ZW2	Bacterial	RSL, BambL	1.6	Fuc(a1-2)Gal(b1-3)GlcNAc(b1-3)Gal	Tetrasaccharide	None	None	None	0.29	42.75	37.67	2	1.776362096	3.910061265	16	2.263499784	3.865487963
3ZWE	Bacterial	RSL, BambL	1.75	Gal(a1-3)[Fuc(a1-2)]Gal(b1-4)Glc	Tetrasaccharide	None	None	None	1.45	-11.11	23.71	19	1.439976775	2.983952706	1	2.697448716	2.697448716
3ZZV	Bacterial	RSL, BambL	1.68	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	None	None	None	-11.78	-2.1	-0.15	17	0.844382023	2.994146367	18	5.555923513	6.788981703
4AGI	Fungal and yeast	AOL, AFL / FleA	1.6	Fuc	Monosaccharide	None	None	None	34.25	-3.54	95.88	2	0.641104576	1.731291287	2	0.641104576	1.69767078
4AGT	Fungal and yeast	AOL, AFL / FleA	2	Fuc(a1-6)GlcNAc	Disaccharide	Na	None	None	3.06	19.8	9.23	9	2.043145633	2.913799884	3	2.003242909	4.971369081
4AH4	Fungal and yeast	AOL, AFL / FleA	1.75	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	HCl	None	None	0.45	15.61	49.55	14	1.267922423	5.743104134	12	1.26903906	5.976828179
4AHA	Fungal and yeast	AOL, AFL / FleA	2.2	Fuc	Monosaccharide	None	None	None	-1.65	26.39	42.17	1	0.254172998	0.254172998	1	0.248821388	0.248821388
4B4Q	Bacterial	FedF / F18	2	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	7.93	-29.29	12.7	18	8.150064182	8.557619854	3	1.796185696	11.57628448
4B4R	Bacterial	FedF / F18	1.8	Gal(a1-3)[Fuc(a1-2)]Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	8.28	-29.12	12.17	9	4.68064471	8.342602097	8	1.302455483	11.51644441
4C1Y	Fungal and yeast	AOL, AFL / FleA	2	Fuc	Monosaccharide	Ca	None	9 Glycerol (C3H8O3), Ethylene Glycol (C2H6O2)	1.71	37.22	-25.22	1	0.494585887	0.494585887	1	0.484981364	0.484981364
4CE8	Bacterial	LecB / PA-IIL, RSIIL	0.9	Fuc	Monosaccharide	11H2SO4(-1), 8Ca	None	Urea (CH4N2O)	29.37	18.83	59.93	7	7.372355079	9.10545321	7	7.372355079	9.10545321
4CSD	Bacterial	RSL, BambL	1.35	Fuc	Monosaccharide	None	None	None	-12	-19.89	56.94	1	0.163236516	0.163236516	1	0.160066573	0.160066573
4CSY	Animal	E-selectin	2.41	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	2Ca	None	None	-48.97	-5.91	-3.56	20	2.128194903	5.207567517	15	3.601385999	5.235723428
4D4U	Fungal and yeast	AOL, AFL / FleA	1.99	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	3 Glycerol (C3H8O3)	-2.5	-6.27	9.09	5	0.783373949	7.478014104	1	2.479087053	2.479087053

4DRV	Virus	Human rotavirus P[14], P[9]	1.56	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	31.12	-27.36	29.38	3	1.804502776	4.836127356	2	2.131795665	7.041255032
4DS0	Virus	Human rotavirus P[14], P[9]	1.56	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)GlcNAc	Tetrasaccharide	None	None	None	10.5	-10.39	13.28	1	4.552137625	5.576978278	8	4.536863677	5.733829682
4I6S	Bacterial	RSL, BambL	1.54	Fuc	Monosaccharide	None	None	Ethylene Glycol (C2H6O2), Diethylene Glycol (C4H10O3), Triethylene Glycol (C6H14O4)	-17.08	-9.71	17.01	1	0.27712656	0.27712656	1	0.271291663	0.271291663
4KMB	Animal	MBP-A / MBL1	2	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	11Ca, 3HCl, Zn	None	None	30.5	54.97	33	7	5.110541207	7.802542964	11	4.459987759	7.219490989
4OP7	Virus	Norovirus GII 4	1.92	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	2 Imidazole (C3H4N2)	27.87	-5.93	-29.71	2	2.11194417	2.881242523	2	2.104652479	2.845363421
4OPO	Virus	Norovirus GII 4	1.4	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc	Tetrasaccharide	2 Acetate (C2H3O2(-1))	None	5 Ethylene Glycol (C2H6O2)	122.9	3.93	30.11	1	5.180056624	5.180056624	10	0.463203508	5.205468438
4P1V	Virus	Norovirus GI 7	1.55	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	None	None	None	58.48	60.1	122.46	7	2.561537777	6.70376734	16	2.717957222	6.661980987
4P25	Virus	Norovirus GI 7	1.5	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	106.23	35.84	27.91	11	4.82549053	6.915911414	1	1.245844446	1.245844446
4P26	Virus	Norovirus GI 7	1.9	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)GlcNAc	Tetrasaccharide	None	None	None	13.76	25.19	117.8	4	5.303383204	7.234371195	8	5.311850252	7.645069974
4P2N	Virus	Norovirus GI 7	1.7	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	None	None	None	17.37	-35.08	-31.93	14	2.211306621	7.228491872	16	3.970906153	8.048737541
4P3I	Virus	Norovirus GI 7	1.6	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	None	None	None	98.31	57.27	29.33	11	4.199149982	4.731771121	6	3.202791949	4.630559782
4RDK	Virus	Norovirus GI 8	1.63	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc	Tetrasaccharide	None	None	None	-19.92	-69.25	25.34	2	2.54383383	5.361047715	1	0.58720443	0.58720443
4RDL	Virus	Norovirus GI 8	1.45	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	66.18	36.63	1.41	7	4.561024598	8.552315776	1	0.675570744	0.675570744
4RM0	Virus	Norovirus GII 13, GII 21	2	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	None	None	None	-11.24	-13.82	47.72	11	1.636836154	6.084013138	12	0.600459306	5.369464342
4USO	Fungal and yeast	CCL2	1.95	NeuAc(a2-3)GlcNAc(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	None	-37.49	122.39	19.43	1	2.274301457	2.274301457	10	1.293053553	2.257965355
4UT5	Bacterial	LecB / PA-IIL, RSIIL	1.75	Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal	Tetrasaccharide	8Ca	None	None	-10.33	9.36	7.7	5	2.362563813	9.520656048	1	1.398156375	1.398156375
4WZE	Virus	Norovirus GII 4	1.46	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	2 Acetate (C2H3O2(-1))	None	None	-190.21	-8.02	37.28	2	2.13529557	4.738278416	1	1.99434381	1.99434381
4WZK	Virus	Norovirus GII 4	1.47	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	2 Acetate (C2H3O2(-1))	None	None	119.64	4.76	109.83	12	2.841973268	6.092216125	6	1.783958156	5.232125998
4WZL	Virus	Norovirus GII 4	1.57	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	4 Acetate (C2H3O2(-1))	None	9 Ethylene Glycol (C2H6O2)	189.94	2.75	36.37	4	4.303738868	5.863855929	7	2.782545383	5.844365443
4WZT	Virus	Norovirus GII 4	1.91	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	4 Ethylene Glycol (C2H6O2)	28.31	10.89	-47.27	15	2.47055787	2.712267418	15	2.34814751	2.926353302
4X06	Virus	Norovirus GII 4	1.22	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	2 Acetate (C2H3O2(-1))	None	None	259.65	9.37	17.34	2	1.524183841	2.42730081	1	1.519651992	1.519651992
4X07	Virus	Norovirus GII 4	1.28	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	2 Acetate (C2H3O2(-1))	None	None	-200.71	-5.2	130.54	2	2.682345373	3.07962882	1	3.057540921	3.057540921
4X0C	Virus	Norovirus GII 4	1.7	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	None	None	4 Ethylene Glycol (C2H6O2), 6 Methanediol (CH4O2)	-23.93	39.58	0.84	4	0.990795026	6.103265623	1	1.001836206	1.001836206
4YM1	Animal	galectin-4	2	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	3H2SO4(-1)	None	None	-50.08	27.73	38.84	15	5.929427713	7.715833974	13	5.134317078	7.675390389
4Z4R	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.8	Fuc	Monosaccharide	None	None	5 Ethylene Glycol (C2H6O2)	4.43	7.09	40.56	1	0.353348578	0.353348578	1	0.345908827	0.345908827
4Z4Y	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.8	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	9 Ethylene Glycol (C2H6O2)	1.35	5.29	41.26	8	2.8235438	6.166038778	20	1.091188357	6.120154016
4ZH7	Virus	BabA	1.91	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal	Oligosaccharide	None	None	None	-34.24	-19.61	51.31	19	1.831208128	9.281116362	8	1.731248792	9.41994556
5A6X	Bacterial	LecB / PA-IIL, RSIIL	1.55	Fuc	Monosaccharide	8Ca	None	None	-4.43	7.82	8.77	1	0.208883508	0.208883508	1	0.204827132	0.204827132
5A6Z	Bacterial	LecB / PA-IIL, RSIIL	1.5	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	8Ca, 3H2SO4(-1)	None	None	36.16	66.99	24.03	12	1.915344187	3.956551364	7	0.878984677	3.295800266
5AJB	Bacterial	RSL, BambL	1.8	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-4)Glc	Tetrasaccharide	None	None	None	8.1	-14.19	24.39	1	0.712999241	0.712999241	1	1.017643462	1.017643462
5AJC	Bacterial	RSL, BambL	1.7	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	Tris(hydroxyethyl)aminomethane (C4H11NO3)	42.4	67.41	10.33	2	0.731686132	2.901611713	5	5.868639048	10.922860l
5C9P	Bacterial	PLL	1.75	Fuc	Monosaccharide	None	None	10 Glycerol (C3H8O3)	-0.43	17.89	43.98	2	0.392348803	1.038723757	2	0.392348803	1.016853439
5CB7	Virus	Human rotavirus P[14], P[9]	1.9	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	2HCl	None	2 Glycerol (C3H8O3)	2.29	4.67	20.09	5	4.70513644	6.970639076	2	2.73525842	6.92582312
5D2A	Bacterial	LecB / PA-IIL, RSIIL	1.8	Fuc	Monosaccharide	4Ca	None	None	-67.06	16.87	15.39	2	1.148621814	2.692600134	2	1.148621814	2.644080842
5DUX	Animal	galectin-4	1.85	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	None	None	2 Glycerol (C3H8O3), Methanediol (CH4O2)	80.53	-19.96	-44.67	18	1.871383029	7.314703434	6	4.501842704	6.04756983
5ELB	Bacterial	LTB, CTB, CFXB	1.08	Fuc(a1-2)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	10Ca, 10 [Bis(2-hydroxyethyl)amino]acetate(-1)	None	Diethylene Glycol (C4H10O3)	-45.42	-18.69	16.03	2	1.204437637	2.121956436	1	1.205005597	1.205005597
5ELD	Bacterial	LTB, CTB, CFXB	1.4	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)[Fuc(a1-3)]GlcNAc	Oligosaccharide	Na, 2-(4-Morpholinyl)ethanesulfonate (C6H12NSO4(-1))	None	3 Triethylene Glycol (C6H12O4), 2 Diethylene Glycol (C4H8O3)	-27.44	-30.63	-33.9	2	2.503884137	8.844354206	5	0.855281557	7.801295623
5ELF	Bacterial	LTB, CTB, CFXB	1.55	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)[Fuc(a1-3)]Glc	Oligosaccharide	10Ca, 10 [Bis(2-hydroxyethyl)amino]acetate(-1)	None	Bis[2-(2-hydroxyethoxy)ethyl]ether (C8H18O5)	-39.75	-7.81	-47.7	14	0.83328311	6.186532	12	2.274983107	7.654038677
5EO7	Fungal and yeast	AOL, AFL / FleA	2.3	Fuc	Monosaccharide	4H2SO4(-1)	None	None	100.23	95.58	71.96	1	1.022167853	1.022167853	1	1.002318044	1.002318044
5F7M	Bacterial	BabA	2.72	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	11.26	-8.73	58.63	13	2.721563375	4.884747477	1	0.662013518	0.662013518
5F7N	Bacterial	BabA	2.72	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-3)[Fuc(a1-4)]GlcNAc	Oligosaccharide	None	None	None	-12.3	11.85	-64.86	1	1.047098612	1.047098612	3	0.518946	1.095119763
5F7W	Bacterial	BabA	2.81	Gal(a1-3)[Fuc(a1-2)]Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	9.96	14.32	-1.76	1	1.169292142	1.169292142	1	1.013933849	1.013933849
5F7Y	Bacterial	BabA	2.44	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	-10.25	8.49	1.47	1	0.958914433	0.958914433	18	0.911464845	1.030784952
5F8R	Bacterial	BabA	2.44	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	20.31	-74.52	-123.86	5	1.949880398	4.131099756	1	1.272341947	1.272341947
5F93	Bacterial	BabA	2.99	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	143.65	-9.87	122.26	1	3.712515259	3.712515259	11	1.597605532	3.216091899
5F97	Bacterial	BabA	2.62	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	20.5	26.06	182.56	2	2.387477977	2.387477977	12	1.007529197	3.464620138
5F9A	Bacterial	BabA	2.44	Fuc(a1-2)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	17.56	127.35	280.63	2	2.895640289	3.899131328	10	2.275637573	11.07888239
5F9D	Bacterial	BabA	2.59	Gal(a1-3)[Fuc(a1-2)]Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	-18.25	9.86	32.05	16	1.789266117	3.245167973	2	1.28272958	1.400802632
5F5D	Fungal and yeast	tectonin-2	1.65	Fuc	Monosaccharide	6H2SO4(-1), 2Ni	None	16 Ethylene Glycol (C2H6O2)	41.49	93.25	89.21	1	2.971943304	3.03079938	1	2.971943304	2.971943304
5H47	Fungal and yeast	AOL, AFL / FleA	2.3	Fuc	Monosaccharide	None	None	None	-52.81	178.05	108.54	1	1.728563394	1.728563394	1	1.697415556	1.697415556

5HCH	Bacterial	LecB / PA-IIL, RSIIL	0.9	Fuc	Monosaccharide	2Ca, Na	None	None	-22.83	71.07	45.91	4	1.363762119	5.872030628	4	1.363762119	5.785032772
5HZA	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.48	Gal(b1-4)[Fuc(a1-3)]Glc	Trisaccharide	None	None	12 Ethylene Glycol (C2H6O2)	2.25	4.94	43.3	7	5.50538578	6.611236884	13	4.937401484	6.798646332
5HZB	Virus	Norovirus GII 1, GII 10, GII 12, GII 18	1.55	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	None	None	4 Ethylene Glycol (C2H6O2)	2.32	3.02	41.99	13	2.543211882	5.47325079	9	2.457300354	6.572188408
5I4D	Bacterial	SSL3	1.75	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	3 Bis[2-(2-hydroxyethoxy)ethyl]ether (C8H18O4)	27.11	220.67	2.9	1	0.440527761	0.440527761	1	0.447017916	0.447017916
5IYP	Virus	Norovirus GII 4	1.27	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	Citrate Trianion (C6H5O7(-3))	None	7 Ethylene Glycol (C2H6O2)	-29.02	-5.8	1.5	1	1.959293282	1.959293282	1	1.913797901	1.913797901
5IYQ	Virus	Norovirus GII 4	1.41	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	Citrate Trianion (C6H5O7(-3))	None	18 Ethylene Glycol (C2H6O2)	-29.3	-5.96	1.92	1	1.718635062	1.718635062	1	1.737009254	1.737009254
5IYR	Virus	Norovirus GII 4	1.54	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	Citrate Trianion (C6H5O7(-3))	None	None	-13.34	-47.83	2.67	9	3.063030804	7.164985161	2	2.830969619	7.401495197
5IYW	Virus	Norovirus GII 4	1.41	Gal(b1-4)[Fuc(a1-3)]Glc	Trisaccharide	Citrate Trianion (C6H5O7(-3))	None	2 Ethylene Glycol (C2H6O2)	-13.87	-47.68	3.03	6	3.810886349	4.853630717	10	1.754032824	3.768172466
5J35	Virus	Norovirus GII 4	1.47	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	None	None	4 Ethylene Glycol (C2H6O2)	69.69	-17.75	148.56	6	2.664561972	2.711843042	11	2.191569542	2.665077617
5J3O	Virus	Norovirus GII 4	1.47	Fuc	Monosaccharide	None	None	6 Ethylene Glycol (C2H6O2)	192.22	-14.55	130.24	1	0.519908939	0.519908939	1	0.508962261	0.508962261
5KIQ	Bacterial	SrpA	1.64	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	7Ca, 6 Acetate (C2H3O2(-1))	None	None	12.31	11.98	7.69	17	6.826647799	9.868464477	14	1.005014489	10.01895975
5LFE	Virus	Norovirus GII 4	2.3	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	63.45	11.06	94.72	12	3.005652865	3.101937483	1	3.075616457	3.075616457
5LKC	Virus	Norovirus GII 1, GII 10, GII 12, GII 20	1.81	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	2Mg	None	4 Ethylene Glycol (C2H6O2)	-39.98	-12.59	-29.82	18	4.041710048	8.697068174	10	8.027812529	8.664717698
5LKG	Virus	Norovirus GII 1, GII 10, GII 12, GII 20	1.51	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	2Mg	None	None	32.29	-16.92	-28.29	14	5.258892547	8.039971488	15	4.79576413	9.173635228
5LKK	Virus	Norovirus GII 1, GII 10, GII 12, GII 20	1.49	Gal(b1-4)[Fuc(a1-3)]Glc	Trisaccharide	2Mg	None	7 Ethylene Glycol (C2H6O2)	-31.39	15.57	-28.38	18	6.05097479	8.973249625	17	7.284398269	8.076028614
5MAZ	Bacterial	LecB / PA-IIL, RSIIL	1.45	Fuc	Monosaccharide	8Ca	None	Ethylene Glycol (C2H6O2), Glycerol (C3H8O3)	23.03	41.68	18.5	1	0.251563681	0.251563681	1	0.245779835	0.245779835
5MB1	Bacterial	LecB / PA-IIL, RSIIL	1.65	Fuc	Monosaccharide	7Ca, Citratie Trianion (C6H5O7(-3))	None	None	32.28	-7.73	6.68	1	0.284820578	0.284820578	1	0.278272104	0.278272104
5MKC	Fungal and yeast	AAL	1.14	Fuc	Monosaccharide	None	Mutant N224Q	3 Glycerol (C3H8O3)	39.99	-37.05	-14.95	1	0.244538259	0.244538259	1	0.239789491	0.239789491
5MKF	Bacterial	PhL, PLL-like	1.9	Fuc	Monosaccharide	8HCl, 4Na	None	None	-3.42	16.14	-9.23	1	0.250370925	0.250370925	1	0.245508891	0.245508891
5MXG	Bacterial	PhL, PLL-like	2.1	Fuc(a1-2)Gal(b1-4)GlcNAc	Trisaccharide	Na	None	None	2.95	13.14	26.89	4	0.786375624	6.836273827	1	0.849366823	0.849366823
5N7M	Virus	Norovirus GI 1	1.73	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	2Na	None	9 Ethylene Glycol (C2H6O2)	-22.09	34.7	33.9	6	2.260892972	7.992964046	1	2.25124625	2.25124625
5NF0	Bacterial	LecB / PA-IIL, RSIIL	1.27	Fuc	Monosaccharide	3NH2, 8Ca	None	3 o-Xylene (C8H10)	-8.3	-15.86	-6.78	1	0.980192838	0.980192838	1	0.962530259	0.962530259
5O7U	Bacterial	RSL, BambL	1.15	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-3)Gal	Tetrasaccharide	Ca, 2Mg	None	None	68.08	42.9	21.6	12	7.612563326	8.163738368	16	4.447510738	7.541851207
5T54	Plant	BFL	1.65	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	4Ca	None	4 Ethylene Glycol (C2H6O2)	45.72	14.12	18.29	1	1.335938406	1.335938406	1	1.343657063	1.343657063
5VK5	Virus	Human rotavirus P[6], P[19]	1.94	Fuc(a1-2)Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	2H2SO4(-1)	None	4 Glycerol (C3H8O3)	5.26	-18.56	15.95	10	2.729748321	5.876542263	8	2.508704994	10.83570329
5VX5	Virus	Human rotavirus P[8], P[4]	1.28	Fuc(a1-2)Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	1.99	25.7	6.72	19	2.117630127	6.675256659	14	1.087528281	6.907577333
5XFD	Fungal and yeast	agrocybe ACG	1.5	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)GlcNAc	Tetrasaccharide	None	None	None	-63.52	-26.95	8.69	7	1.038315302	3.005565807	3	1.048985588	2.312542752
5ZHO	Virus	Human rotavirus P[2]	1.4	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	88.87	78.32	236.94	2	1.630744524	4.44921822	1	1.638715554	1.638715554
5ZV5	Virus	Norovirus GII 1, GII 10, GII 12, GII 20	2.1	GalNAc(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	-39.62	-12.5	-29.71	15	4.145060294	9.228479063	15	5.893396452	9.250605924
5ZV7	Virus	Norovirus GII 1, GII 10, GII 12, GII 20	1.95	Gal(a1-3)[Fuc(a1-2)]Gal	Trisaccharide	None	None	None	-35.97	12.63	-29.18	20	6.187104845	9.21470254	19	6.752533943	9.144090594
5ZVC	Virus	Norovirus GII 13, GII 21	1.8	Gal(b1-3)[Fuc(a1-4)]GlcNAc	Trisaccharide	None	None	2 Glycerol (C3H8O3)	59.01	94.85	157.26	16	1.199412322	6.473581829	6	1.45279053	4.708657834
6A87	Fungal and yeast	PhoSL	2.41	Fuc(a1-6)GlcNAc	Disaccharide	None	None	None	-20.6	40.63	4.86	2	2.413789925	3.504553839	2	1.484276255	2.295440838
6F5W	Bacterial	PhL, PLL-like	1.91	Fuc	Monosaccharide	9Na, 11HCl	None	5 Dimethyl sulfoxide (C2H6SOH)	46	-25.73	30.4	5	0.538200706	4.056797352	5	0.538200706	3.988611028
6FLU	Bacterial	PhL, PLL-like	1.78	Fuc	Monosaccharide	3Na, 6HCl	None	None	46.99	-1.96	15.43	1	0.564461564	0.564461564	1	0.551483701	0.551483701
6FX1	Fungal and yeast	PhoSL	2.1	GlcNAc(b1-2)Man(a1-6)[GlcNAc(b1-2)Man(a1-3)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)]GlcNAc	Oligosaccharide	12 Malonate dianion (C3H2O4(-2))	None	None	50.66	-2.09	-23.27	16	1.585726297	5.330377773	1	1.175865171	1.175865171
6FX2	Fungal and yeast	PhoSL	1.7	Gal(b1-4)GlcNAc(b1-2)Man(a1-6)[GlcNAc(b1-2)Man(a1-3)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)]GlcNAc	Oligosaccharide	None	None	None	-9.02	-7.18	12.18	2	10.60579238	11.49898899	14	11.13557235	15.21406493
6FX3	Fungal and yeast	PhoSL	1.7	NeuAc(a2-3)Gal(b1-4)GlcNAc(b1-2)Man(a1-6)[GlcNAc(b1-2)Man(a1-3)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)]GlcNAc	Oligosaccharide	None	None	None	-13.66	-27.38	-32.88	5	6.286093763	11.19994494	19	8.122511581	9.772112167
6H70	Virus	Norovirus GI 1	1.83	Fuc(a1-2)Gal(b1-4)Glc	Trisaccharide	2Na	None	4 Ethylene Glycol (C2H6O2)	161.64	28.1	63.68	9	2.236247559	7.576115109	2	2.250937151	7.384100003
6HA0	Virus	Human rotavirus P[8], P[4]	1.85	Fuc(a1-2)Gal(b1-3)GlcNAc	Trisaccharide	None	None	None	30.58	6.99	70.33	16	2.430975202	10.85221271	8	2.686798648	10.73036641
6HID	Bacterial	LTB, CTB, CFXB	1.54	Gal(b1-4)[Fuc(a1-3)]GlcNAc	Trisaccharide	10Ca, 10 [Bis(2-hydroxyethyl)amino]acetate(-1)	None	None	-33.47	-49.36	-1.54	15	4.79373688	6.039882898	4	0.748377452	3.99679227
6HMW	Bacterial	LTB, CTB, CFXB	1.95	Fuc	Monosaccharide	10Ca, 10 [Bis(2-hydroxyethyl)amino]acetate(-1)	None	None	30.65	2.39	38.42	1	0.544307937	0.544307937	1	0.532847539	0.532847539
6HMY	Bacterial	LTB, CTB, CFXB	1.6	Fuc(a1-2)Gal(b1-3)GalNAc(b1-4)[NeuAc(a2-3)]Gal(b1-4)Glc	Oligosaccharide	10Ca, 10 [Bis(2-hydroxyethyl)amino]acetate(-1)	None	None	-36.51	-3.72	64.27	13	4.920197419	10.0736266	15	4.791983159	7.800966646
6HTN	Fungal and yeast	KozL	1.55	Fuc	Monosaccharide	None	None	6 2-methyl-2,4-pentanediol (C6H14O2), 3 Ethylene Glycol (C2H6O2), Octaethylene Glycol (C16H34O8)	23.65	96.85	-16.12	1	0.22224239	0.22224239	1	0.217926593	0.217926593
6I55	Virus	Norovirus GII 3	2.5	GalNAc(a1-3)[Fuc(a1-2)]Gal(b1-4)Glc	Tetrasaccharide	None	None	None	130.14	118.47	230.73	12	2.312875704	6.779874719	3	1.673656482	7.293040778
6K2O	Virus	Human rotavirus P[8], P[4]	2.3	Fuc(a1-2)Gal(b1-3)GlcNAc(b1-3)Gal	Tetrasaccharide	Na	None	None	-28.78	47.38	6.48	16	2.612837847	6.531002092	6	2.433633212	6.089249369
6OAI	Virus	Human rotavirus P[6], P[19]	1.9	Fuc(a1-2)Gal(b1-3)GlcNAc(b1-3)Gal(b1-4)Glc	Oligosaccharide	None	None	None	36.32	11.85	41.45	1	3.644936014	3.644936014	19	5.586638808	9.376123218
6Q05	Virus	MERS-CoV	2.8	NeuAc(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc	Tetrasaccharide	None	None	2 Sialyl-LewisX	231.75	283.26	228.11	17	6.620373472	9.038370471	18	5.812790021	9.357415443
6Q6W	Bacterial	LecB / PA-IIL, RSIIL	1.44	Fuc	Monosaccharide	2Ca, NH2	None	None	-19.7	-2.36	-12.36	1	0.932289218	0.932289218	1	0.915489838	0.915489838
6Q6X	Bacterial	LecB / PA-IIL, RSIIL	1.52	Fuc	Monosaccharide	8Ca, 4NH2	None	None	-32.73	-29.4	-16.42	2	0.484918182	1.229814771	2	0.484918182	1.207654131
6Q77	Bacterial	LecB / PA-IIL, RSIIL	2	Fuc	Monosaccharide	2Ca	None	None	61.16	63.97	15.76	5	1.189701523	1.296585459	5	1.189701523	1.273221645

6Q79	Bacterial	LecB / PA-IIL, RSIIL	2.01	Fuc	Monosaccharide	8Ca, 2NH2	None	None	10.94	34.24	-7.02	6	0.735798546	1.080405	6	0.735798546	1.060936648
6Q85	Bacterial	LecB / PA-IIL, RSIIL	1.99	Fuc	Monosaccharide	8Ca, 4NH2	None	None	-13.21	-51.1	-21.8	1	1.17192763	1.17192763	1	1.150810087	1.150810087
6Q86	Bacterial	LecB / PA-IIL, RSIIL	2.01	Fuc	Monosaccharide	4Ca, 2NH2	None	None	-2.3	17.58	17.11	1	1.180163642	1.180163642	1	1.158897691	1.158897691
6Q87	Bacterial	LecB / PA-IIL, RSIIL	2.54	Fuc	Monosaccharide	2Ca, NH2	None	None	-10.96	49.99	69.8	3	0.797640404	3.779318712	3	0.797640404	3.711217301
6Q8D	Bacterial	LecB / PA-IIL, RSIIL	1.63	Fuc	Monosaccharide	2Ca	None	None	38.15	16.15	15.84	3	1.138709514	1.259082029	3	1.138709514	1.236394008
6Q8G	Bacterial	LecB / PA-IIL, RSIIL	1.19	Fuc	Monosaccharide	8Ca, 2NH2	None	None	7.37	-17.11	21.22	1	0.650768378	0.650768378	1	0.639041861	0.639041861
6Q8H	Bacterial	LecB / PA-IIL, RSIIL	1.71	Fuc	Monosaccharide	2Ca, 2NH2	None	None	11.04	-50.52	20.33	1	0.873920964	0.873920964	1	0.858173351	0.858173351
6R35	Bacterial	LecB / PA-IIL, RSIIL	1.8	Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-3)Gal	Trisaccharide	8Ca, 2H2SO4(-1)	None	8 Ethylene Glycol (C2H6O2)	-4.55	0.65	35.07	5	1.527271113	5.161165401	1	0.691099946	0.691099946
6S5P	Bacterial	LecB / PA-IIL, RSIIL	1.46	Fuc	Monosaccharide	8Ca, NH2	None	None	-39.08	38.7	-20.57	2	0.585414444	1.050709778	2	0.585414444	1.031776519
6S5R	Bacterial	LecB / PA-IIL, RSIIL	2.08	Fuc	Monosaccharide	8Ca	None	None	-39.67	7.81	32.82	1	1.024736931	1.024736931	1	1.00627169	1.00627169
6S5S	Bacterial	LecB / PA-IIL, RSIIL	1.43	Fuc	Monosaccharide	2Ca, NH2	None	None	-65.37	15.74	16.23	1	0.793059056	0.793059056	1	0.778768533	0.778768533
6S7G	Bacterial	LecB / PA-IIL, RSIIL	1.84	Fuc	Monosaccharide	8Ca, 2NH2	None	None	-6.7	30.2	1.7	2	0.738425787	1.956836563	2	0.738425787	1.921575358
6SU0	Bacterial	RSL, BambL	1.98	Fuc	Monosaccharide	6Na	None	12 Cucurbit[7]uril	-17.4	27.74	14.47	1	0.542793662	0.542793662	1	0.532252976	0.532252976
6T99	Bacterial	RSL, BambL	2.7	Fuc	Monosaccharide	2Na	Mutant W31YW76Y	None	5.73	-18.24	-71.21	1	0.646005263	0.646005263	1	0.633460277	0.633460277
6T9A	Bacterial	RSL, BambL	2	Fuc	Monosaccharide	None	Mutant W31FW76F	3 Ethylene Glycol (C2H6O2)	-19.3	-21.8	-0.8	1	0.273462818	0.273462818	1	0.267705062	0.267705062
6T9B	Bacterial	RSL, BambL	1.46	Fuc	Monosaccharide	8 Glycine Anions (C2H4NO2)	Mutant W31A	None	37.18	26.25	-63.78	5	0.451675598	3.652525734	5	0.451675598	3.581596153
6TID	Bacterial	Bc2L-C-Nter	1.61	Fuc(a1-2)Gal(b1-3)GlcNAc	Trisaccharide	None	None	None	-8.37	11.75	41.59	5	4.477689	8.494745454	16	4.421101535	7.462707832
6TIG	Bacterial	Bc2L-C-Nter	1.9	Fuc(a1-2)Gal(b1-3)GalNAc(b1-3)Gal(a1-4)Gal	Oligosaccharide	None	None	None	-30.73	-15.12	6.42	6	5.464688758	7.236833276	12	0.941529733	6.774412658
6ZFC	Bacterial	RSL, BambL	1.65	Fuc	Monosaccharide	None	None	None	-26.28	13.72	53.34	2	0.357716368	6.078716859	2	0.357716368	5.998200498
7C38	Fungal and yeast	AOL, AFL / FleA	1.2	Fuc	Monosaccharide	None	None	5 Glycerol (C3H8O3)	-11.54	22.43	52.08	1	0.146803033	0.146803033	1	0.143712096	0.143712096
7C39	Fungal and yeast	AOL, AFL / FleA	1.85	Fuc	Monosaccharide	2 Citrate Trianions (C6H5O7(-3))	None	Glycerol (C3H8O3)	-22.9	-24.35	1.59	1	0.306176028	0.306176028	1	0.300230296	0.300230296

Appendix B

The values of the graphs in Figure 6. This table more clearly shows the differences between ion presence in the docking region of the lectin-glycan complexes.

No Ions				Ions			
Vina		Vina-Carb		Vina		Vina-Carb	
Top	Best	Top	Best	Top	Best	Top	Best
4.69	2.73	4.60	2.47	4.56	2.45	4.00	1.94