

University of Mississippi

eGrove

Honors Theses

Honors College (Sally McDonnell Barksdale
Honors College)

Spring 5-8-2022

Lectin-Glycan Complexes: A Comprehensive Analysis of Docking Calculations

Ashton Custer

Follow this and additional works at: https://egrove.olemiss.edu/hon_thesis



Part of the [Molecular Biology Commons](#), and the [Other Chemicals and Drugs Commons](#)

Recommended Citation

Custer, Ashton, "Lectin-Glycan Complexes: A Comprehensive Analysis of Docking Calculations" (2022).
Honors Theses. 2661.

https://egrove.olemiss.edu/hon_thesis/2661

This Undergraduate Thesis is brought to you for free and open access by the Honors College (Sally McDonnell Barksdale Honors College) at eGrove. It has been accepted for inclusion in Honors Theses by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.

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

© 2022

Ashton Taylor Custer

ALL RIGHTS RESERVED

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	2
List of tables	3
1 Introduction	4
1.1 Terms	4
1.2 Biological roles	5
1.3 Computational study	6
2 Initial data collection and organization	7
3 Process, results, and discussion	12
3.1 AutodockVina and VinaCarb	12
3.2 Optimization	18
3.3 Random seed values	20
4 Conclusions	24
5 Further study	25
6 References	25
Appendix A	27
Appendix B	32

LIST OF FIGURES

Figure 1. Lectin-glycan complex PDB ID: 1FWU	5
Figure 2. L-fucose	8
Figure 3. Categorization pie charts	11
Figure 4. Lectin-glycan complex PDB ID: 1QOT	13
Figure 5. Average RMSD total bar graph	15
Figure 6. Average RMSD ions bar graph	17
Figure 7. Average RMSD heat maps	19
Figure 8. Random seeds graphs	23

LIST OF TABLES

Table 1. Excel sample	9
Table 2. Categorization of lectins	10
Table 3. Statistics	21

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 solanensium* is one such nonpolar example.¹⁴

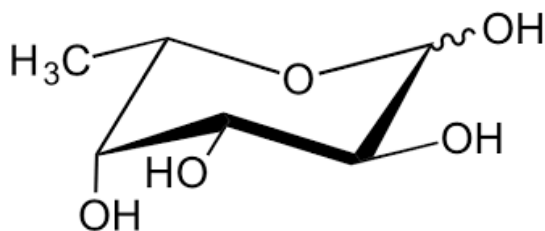


Figure 2. Monosaccharide L-fucose. The squiggled line shows the hydroxyl group that is attached to the anomeric carbon, denoting either alpha or beta orientation. This anomeric carbon denotes the L enantiomer. For the purposes of this study, L-fucose binding lectins were studied rather than D-fucose-binding 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 They 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, RSIIIL. 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, RSIIIL	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							
	<i>Ion</i>	<i>No Ion</i>	<i>Mono</i>	<i>Di</i>	<i>Tri</i>	<i>Tetra</i>	<i>Oligo</i>
<i>Animal</i>							
22	19	3	5	1	9	4	3
<i>Bacterial</i>							
89	59	30	47	1	9	14	18
<i>Fungal and Yeast</i>							
28	10	18	12	2	5	6	3
<i>Plant</i>							
10	10	0	2	0	4	2	2
<i>Virus</i>							
71	25	46	3	1	43	17	7
<i>Total</i>							
220	123	97	69	5	70	43	33

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

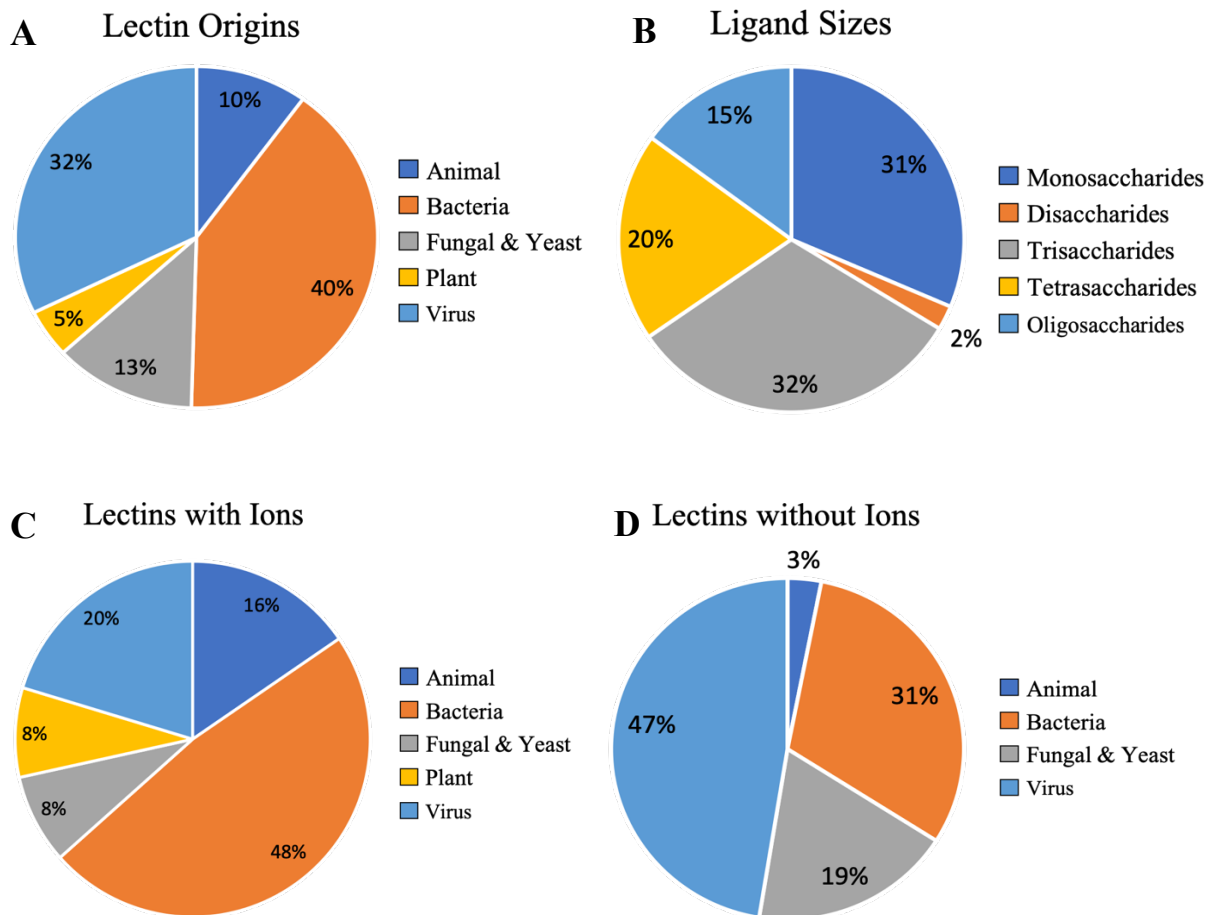


Figure 3. The pie charts show the breakdown of all the 220 complexes in different categories: origin, ligand size, and ion presence. Pie chart A shows the different origins of 220 complexes. Pie chart B shows the sizes of ligands bound to the lectins, from monosaccharides to oligosaccharides (greater 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.

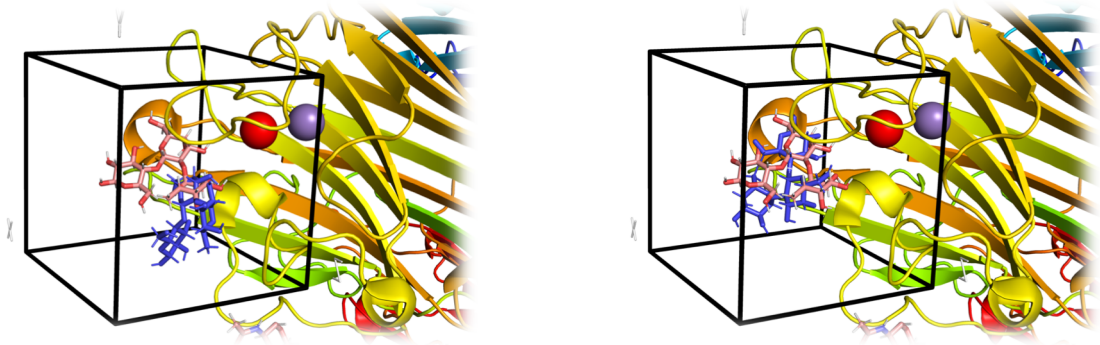


Figure 4. Lectin PDB ID: 1QOT (sourced from the RSCB PDB). (A) represents the very first pose AutodockVina calculated, while (B) represents the 16th and best pose the software calculated. This shows the software does not necessarily dock the lowest scoring, or best, pose with the first calculation. Pose 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 Autodock Vina 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.

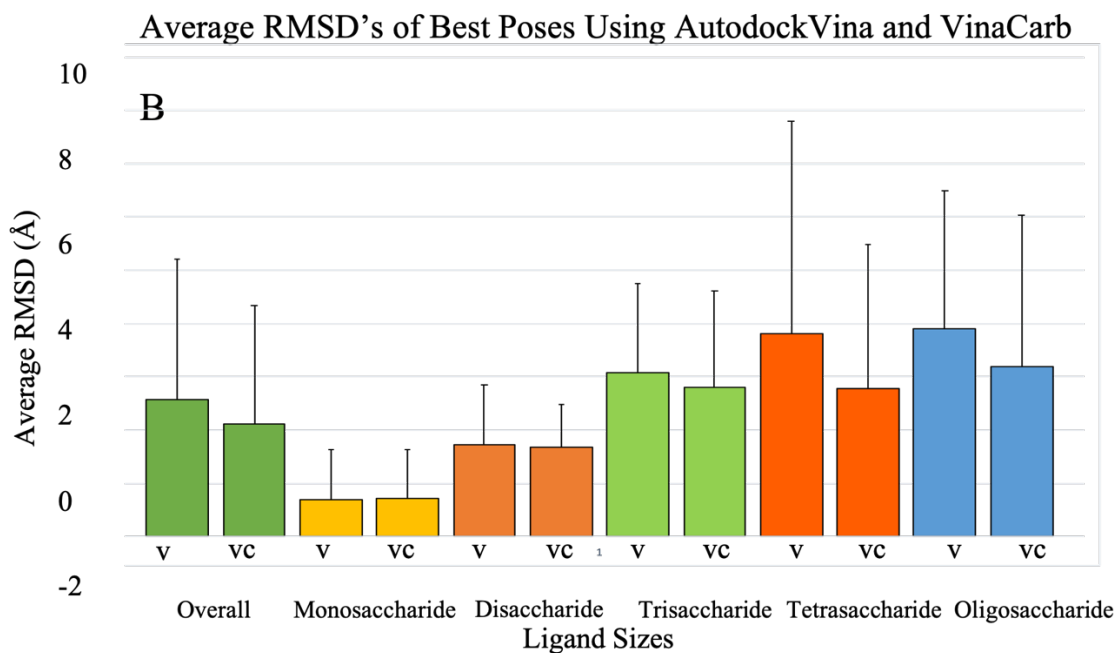
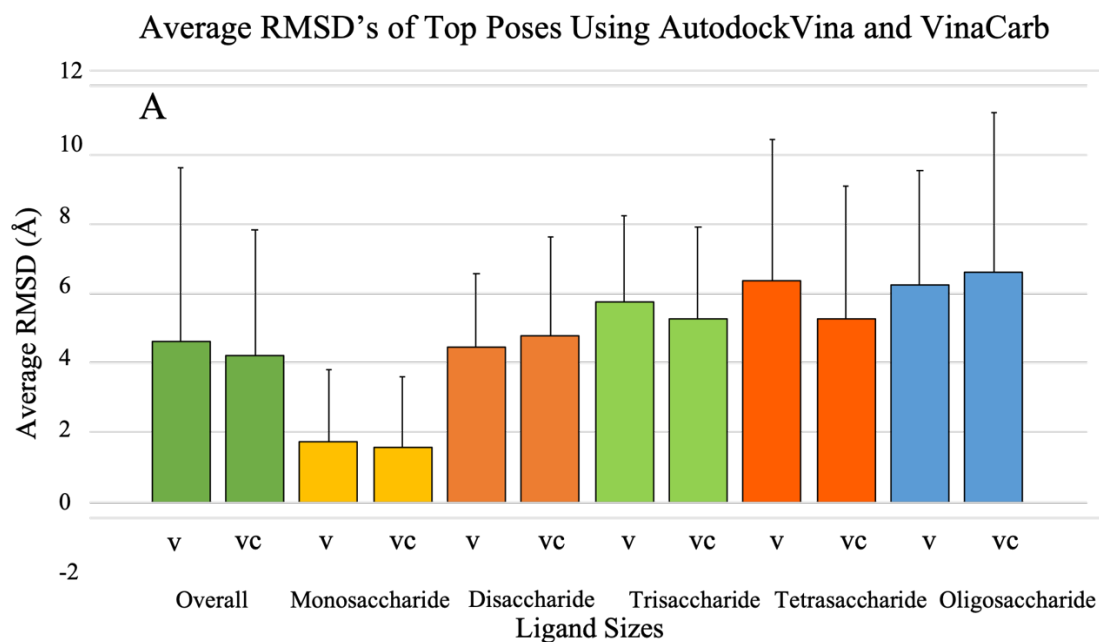


Figure 5. Average RMSD's of all of the different ligand sizes docked using parameters for Autodock Vina (v) and Vina-Carb (vc). For poses n=20, we took the average RMSD's of the first poses as well as the average RMSD's of the best, or lowest scoring, poses. Parameters representing vc generally resulted 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.

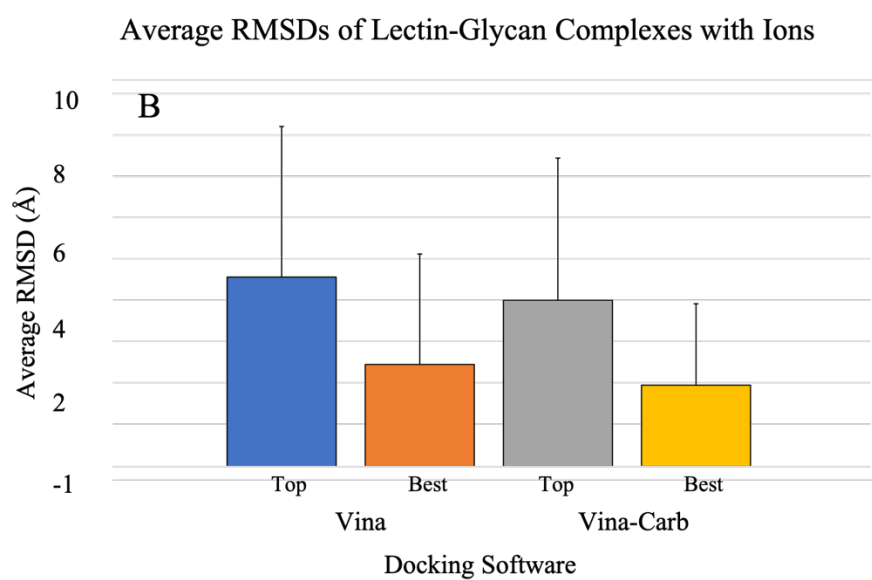
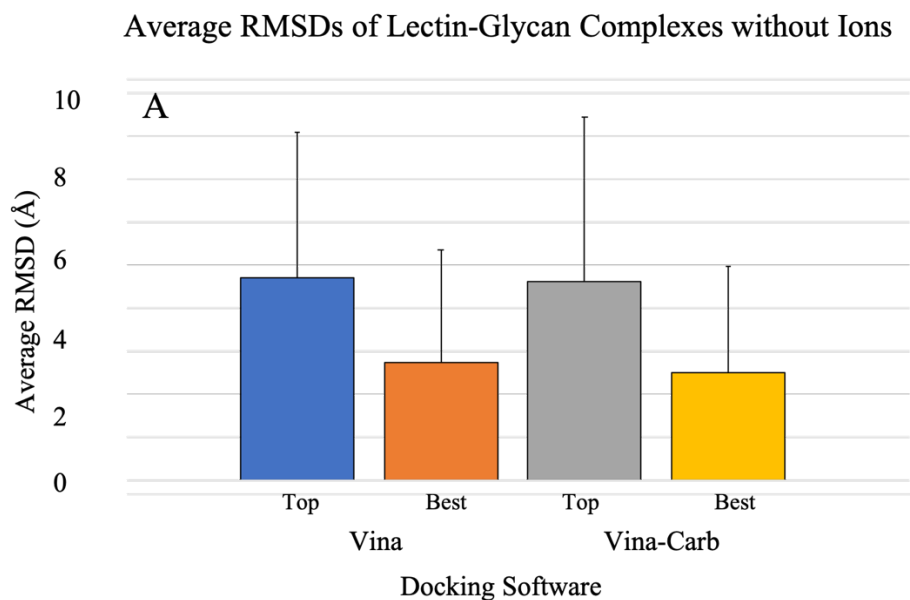


Figure 6. Average RMSD values of the lectin-glycan complexes without ions and with ions; (A) shows the complexes without ions and (B) shows the complexes with ions.

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.

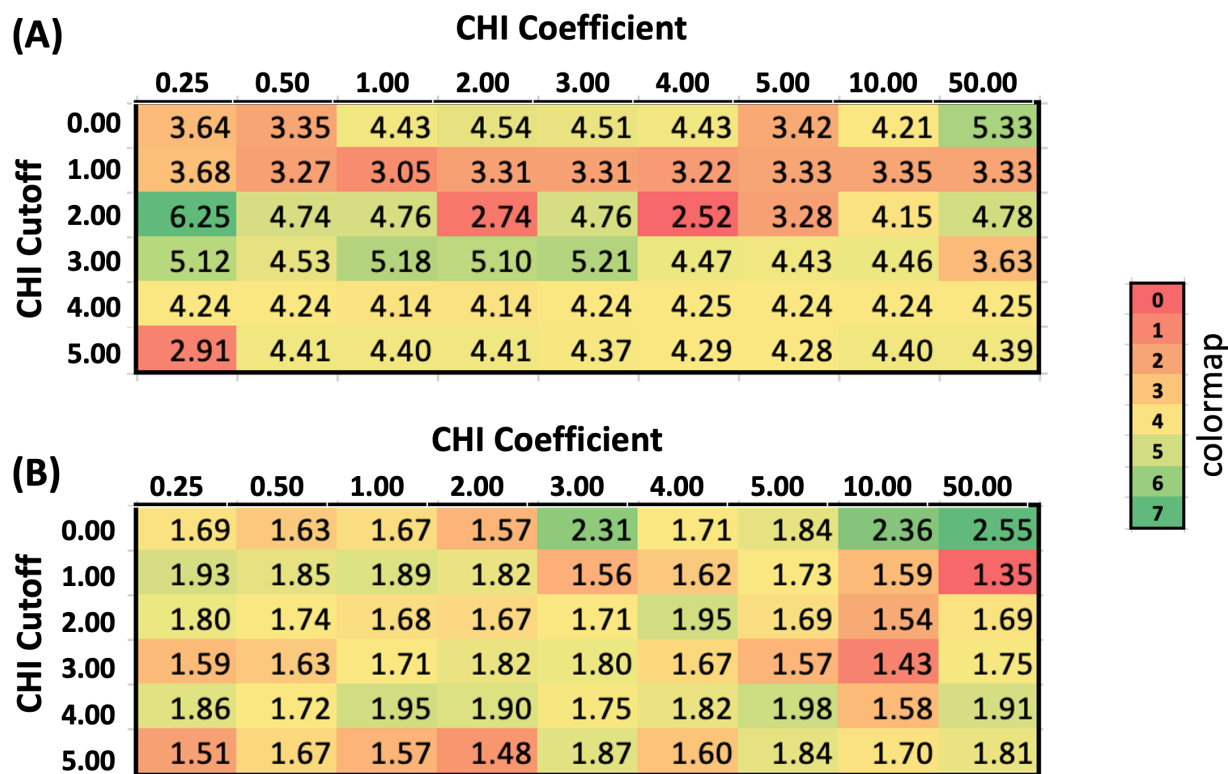


Figure 7. Heat maps showing the average RMSDs of optimized CHI values for five disaccharides. Heat map A represents the average RMSD values of the top scoring poses while heat map B represents the average RMSD values of the lowest scoring RMSD poses. There seems to be much variation within CHI value 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

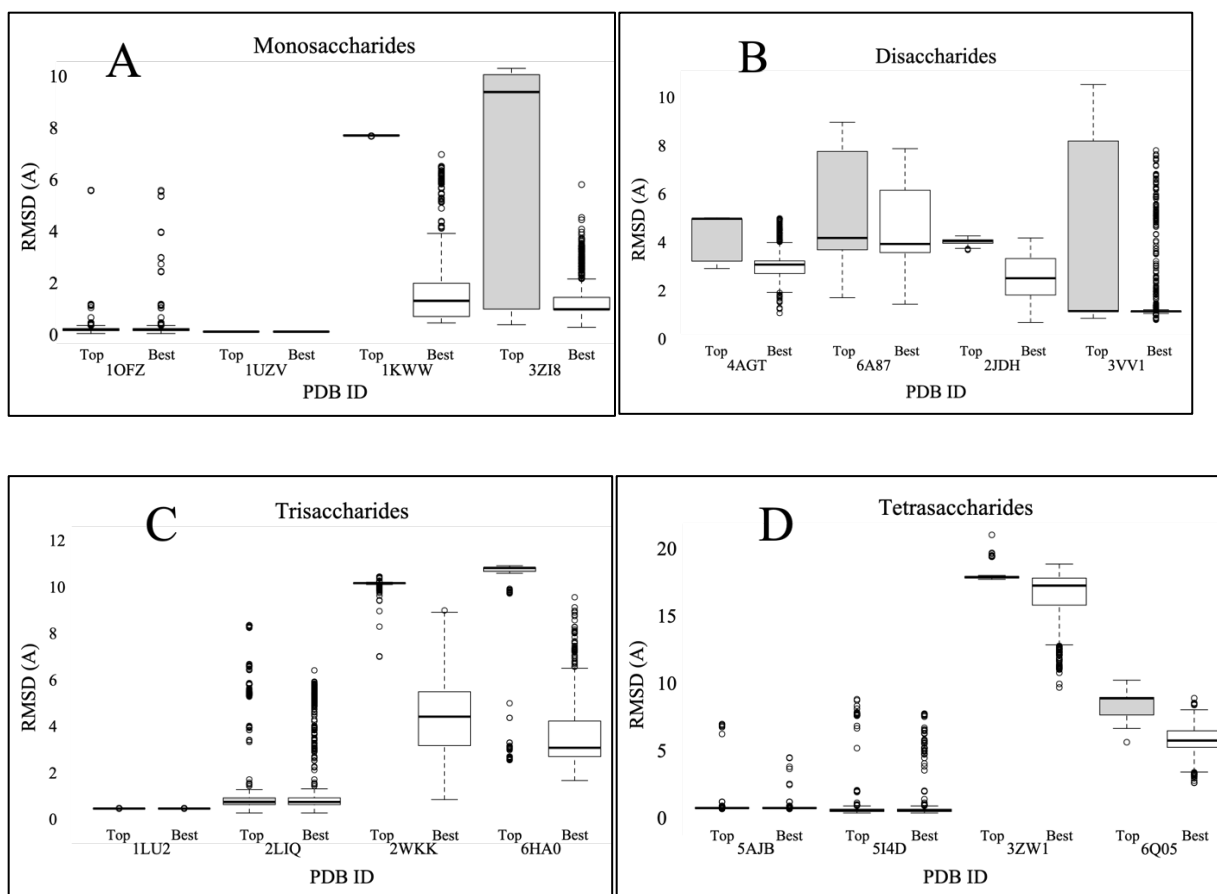


Figure 8. The RMSD (\AA) values of 1000 docking solutions for selected lectins binding to (A) monosaccharides, (B) disaccharides, (C) trisaccharides, and (D) tetrasaccharides. A different random seed was used in each docking run. In each graph, the two lectins on the left (first four bars) represent 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](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>.

6079	Bacterial	Leb / Pa-III, RSIII	2.01	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	10.94	34.24	-7.02	-0.8	1	0.735798546	1.086205	6	0.735798546	1.06936648
6085	Bacterial	Leb / Pa-III, RSIII	1.99	Fuc	Monosaccharide	BcA, 4NH2	None	None	None	None	-13.21	-51.1	-71.8	-0.8	1	1.17192763	1.17192763	1	1.17192763	1.150810087
6086	Bacterial	Leb / Pa-III, RSIII	2.01	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	-2.3	17.58	17.11	-0.8	1	1.180163642	1.180163642	1	1.180163642	1.158897691
6087	Bacterial	Leb / Pa-III, RSIII	2.54	Fuc	Monosaccharide	BcA, NH2	None	None	None	None	-10.96	49.99	69.8	-0.8	3	0.797640404	3.779318712	3	0.797640404	3.711217201
6088	Bacterial	Leb / Pa-III, RSIII	1.63	Fuc	Monosaccharide	BcA, NH2	None	None	None	None	38.15	16.15	15.84	-0.8	3	1.138709514	1.259982029	3	1.138709514	1.236394608
6089	Bacterial	Leb / Pa-III, RSIII	1.19	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	7.37	-17.11	21.22	-0.8	1	0.659768378	0.659768378	1	0.639041861	0.639041861
6089	Bacterial	Leb / Pa-III, RSIII	1.71	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	11.04	-50.52	20.33	-0.8	1	0.873920964	0.873920964	1	0.858173351	0.858173351
6089	Bacterial	Leb / Pa-III, RSIII	1.8	Gal(1-4)Fuc(1-3)GlcNAc(1-3)Gal	Trisaccharide	BcA, 2H2SO4(-1)	None	None	None	None	-4.55	0.65	35.07	-0.8	5	1.527271113	5.161165401	5	0.691099946	0.691099946
653P	Bacterial	Leb / Pa-III, RSIII	1.46	Fuc	Monosaccharide	BcA, NH2	None	None	None	None	-39.08	38.7	20.57	-0.8	2	0.585414444	1.050709778	2	0.585414444	1.031776519
653S	Bacterial	Leb / Pa-III, RSIII	2.08	Fuc	Monosaccharide	BcA, NH2	None	None	None	None	-39.67	7.81	32.82	-0.8	1	1.024726931	1.024726931	1	1.00627169	1.00627169
653S	Bacterial	Leb / Pa-III, RSIII	1.43	Fuc	Monosaccharide	BcA, NH2	None	None	None	None	-65.37	15.74	16.23	-0.8	1	0.793059056	0.793059056	1	0.78768833	0.78768833
653G	Bacterial	Leb / Pa-III, RSIII	1.84	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	-6.7	30.2	1.7	-0.8	2	0.738425787	1.566386563	2	0.738425787	1.921575358
653U	Bacterial	RSU, Baml	1.98	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	-17.4	27.74	14.47	-0.8	1	0.542793662	0.542793662	1	0.532252976	0.532252976
6799	Bacterial	RSU, Baml	2.7	Fuc	Monosaccharide	BcA, ZNH2	Mutant	None	None	None	5.73	-18.24	-71.21	-0.8	1	0.646095263	0.646095263	1	0.634646077	0.634646077
679A	Bacterial	RSU, Baml	2	Fuc	Monosaccharide	BcA, ZNH2	Mutant	W31WV6F	None	None	-19.3	-21.8	-0.8	-0.8	1	0.273462818	0.273462818	1	0.267709662	0.267709662
679B	Bacterial	RSU, Baml	1.46	Fuc	Monosaccharide	BcA, ZNH2	Mutant	W31WV6F	None	None	37.18	26.25	-63.78	-0.8	5	0.451675598	3.652529734	5	0.451675598	3.581596153
679D	Bacterial	Bc2L-C-Nter	1.61	Fuc(1-2)Gal(1-3)GlcNAc	Trisaccharide	(C2)HNO2	None	None	None	None	-8.37	11.75	41.59	-0.8	5	4.477659	8.494745454	5	4.421101535	7.462707632
679G	Bacterial	Bc2L-C-Nter	1.9	Fuc(1-2)Gal(1-3)GlcNAc(1-3)Gal(1-4)Gal	Oligosaccharide	None	None	None	None	None	-30.73	-15.12	6.42	-0.8	6	5.466468758	7.236833276	6	0.357716388	6.774412658
6Z7C	Bacterial	RSU, Baml	1.65	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	-26.28	13.72	53.34	-0.8	2	0.357716388	6.078716859	2	0.357716388	5.982400498
7C38	Fungal and yeast	AOI, AFI / Hea	1.2	Fuc	Monosaccharide	BcA, ZNH2	None	None	None	None	-11.54	22.43	52.06	-0.8	1	0.146803031	0.146803031	1	0.143712096	0.143712096
7C39	Fungal and yeast	AOI, AFI / Hea	1.85	Fuc	Monosaccharide	BcA, ZNH2	(GH507(-3))	None	None	None	-22.9	-24.35	1.59	-0.8	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