# Development of an algorithm to analyze atomic structures of glycan

**Akihiro Fujita**
e13d5605@soka.ac.jp

**Kiyoko F.Aoki-Kinoshita**
kkiyoko@soka.ac.jp

Division of Bioengineering, Graduate School of Engineering, Soka University
1-236 Tangi-machi,Hachioji, Tokyo 192-8577, Japan

**Keywords**: glycomics, glycans, bioinformatics, cheminformatics

## 1 Introduction

In the structural analysis of glycans, it is often insufficient to consider glycans at the monosaccharide level, although bioinformatic analysis at the monosaccharide level is rather mainstream. For example, two completely different glycan structures may both equally bind to the same binding site of a lectin[1], as illustrated in Figure 1. Here, one can see that only some atoms of the monosaccharides on these glycans are directly involved in binding. Therefore, the need for glycan binding analysis at the atomic level has been suggested. The problem, however, is that it is difficult to obtain experimental three-dimensional structures of glycans because they are very flexible, and it is not possible to crystallize them. Therefore, molecular modeling and bioinformatics methods are necessary for the 3D analysis of glycan structures.

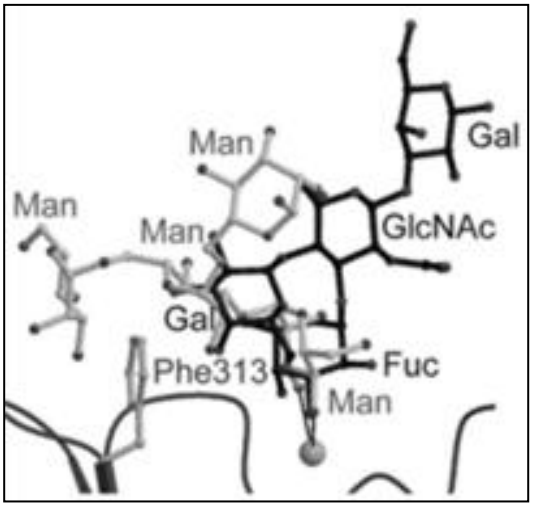


Figure 1. An example of two completely different glycans that bind to the same lectin. This fact suggests that glycans have a common structure in the composition of different monosaccharide.

In this work, we focused on glycan array data for finding the key structures of glycan binding analytes such as lectins and viruses. We can obtain glycan array data from the CFG (Consortium for Functional Glycomics) database [2]. We then developed a novel algorithm for predicting the key glycan binding factors at the atomic level to better understand glycan function.

## 2 Methods

In order to extract the key atomic features of glycans that are important for glycan recognition and binding, we developed the following algorithm consisting of four steps.

1. Predict the three dimensional glycan structure conformations.
   We use the Shape software[3] which is a fully-automated conformation prediction software developed specifically for glycans. Shape is based on molecular mechanics and uses a genetic algorithm and mm3 force field to predict the possible atomic conformations of a single glycan structure. We adjusted the parameters for this software to generate approximately 10,000 conformations per glycan.

2. For each glycan: Narrow down the number of conformations.
   We narrowed down the conformations generated in step 1 down to about 1,500 conformations using a hierarchical clustering method to obtain sufficiently distinct

conformations. This operation reduces computation time and allows us to obtain the major conformations predicted for a glycan.

3. For all glycans under analysis: Align conformations.
   In order to align all glycan conformations obtained in the previous step, we developed an in-house alignment program based on the comparison of triangulated atoms in the conformations. This program was developed based on 3-Points Pharmacophoric Fingerprinting [4].

4. Extract motif structures.
   We extract atomic patterns from the aligned substructures obtained in the previous step by comparing the aligned structures of highly binding glycans against those of non-binding glycans. As a result, we can obtain pairs of common atomic patterns of glycans.

## 3 Results and Discussion

We tested our algorithm on the data from glycan array data of DC-SIGN. We selected two binding and two non-binding glycan structures, and Figure 2 shows one of the 10 pairs obtained in our results compared with a pair of non-binding glycans. The atoms labeled A, B, C and D indicate the aligned atom pairs that are putative atoms important for binding. This result confirms the fact that DC-SIGN recognizes atoms on mannose or fucose, which was confirmed in the literature[1]. Moreover, our new approach took only two days of computation time compared with months of work using traditional experimental procedures.

We are currently implementing this algorithm in Java, making sure to write it efficiently because of the large number of conformation data being produced. We expect that in the future, our work will enable functional research of glycan structures using comprehensive analysis at the atomic level.

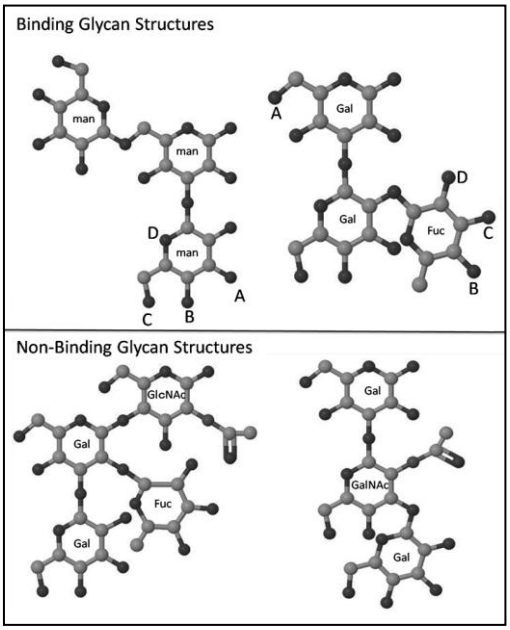


Figure 2. Two different glycan structures that bind to DC-SIGN. A, B, C and D indicate the aligned atoms in our results, confirmed in the literature.

## References


[1] Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nature Structural & Molecular Biology.*11:591-598, 2004.

[2] Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong CH, Paulson JC. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins.*Proceedings of the National Academy of Sciences of the United States of America.* 101(49):17033-17038, 2004.

[3] Rosen J, Miguet L, Perez S. Shape: automatic conformation prediction of carbohydrates using a genetic algorithm. *Journal of Cheminformatics.* 1:1-7, 2009.

[4] Nicolas B, Fareed A,E, XavierB, Ben D, Martin D, Brian D, Harry F, Christophe F, Christine R, Heather S, Roderick E.H. Design and Characterization of Libraries of Molecular Fragments for Use in NMR Screening against Protein Targets. *J. Chem. Inf. Comput. Sci*., 44(6):2157 - 2166, 2004.