

Supplementary Data

GS-align for Glycan Structure Alignment and Similarity Measurement

Hui Sun Lee, Sunhwan Jo, Srayanta Mukherjee, Sang-Jun Park, Jeffrey Skolnick,
Jooyoung Lee and Wonpil Im

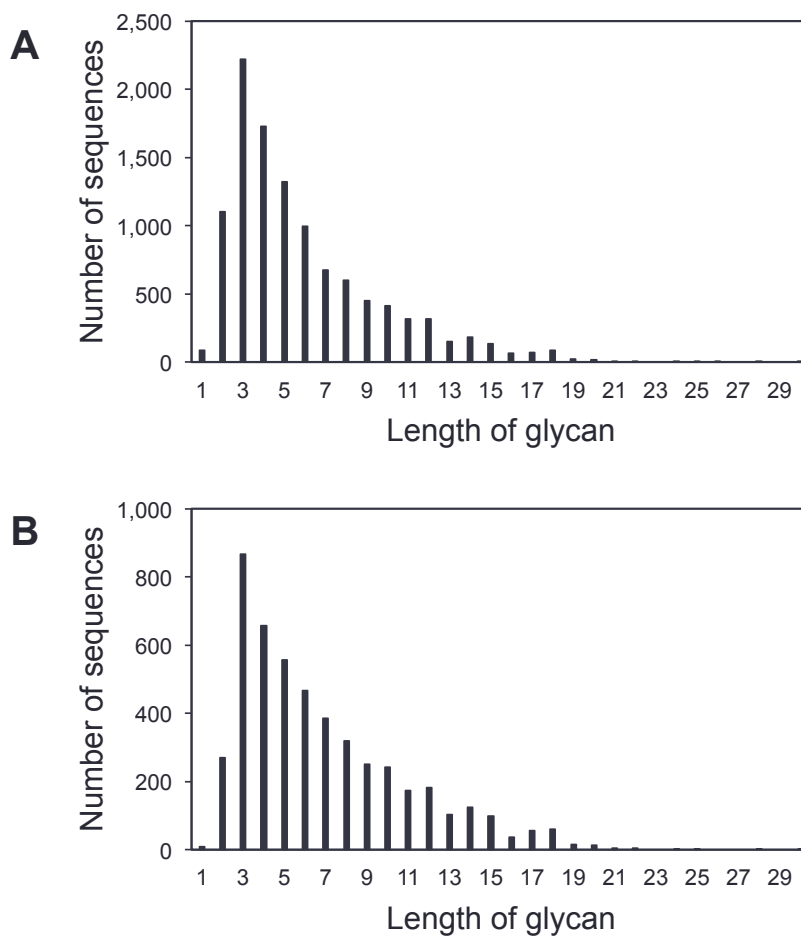


Fig. S1. Distributions of unique sequences in the KEGG database as a function of glycan length for (A) all KEGG glycans and (B) the glycans containing pyranose monosaccharides available in the CHARMM carbohydrate force field.

Table S1. The number of the random glycan structures in terms of glycan length.

Length of glycan	Number of structures
4	40
5	70
6	70
7	100
8	80
9	90
10	90
11	100
12	60
Total	700

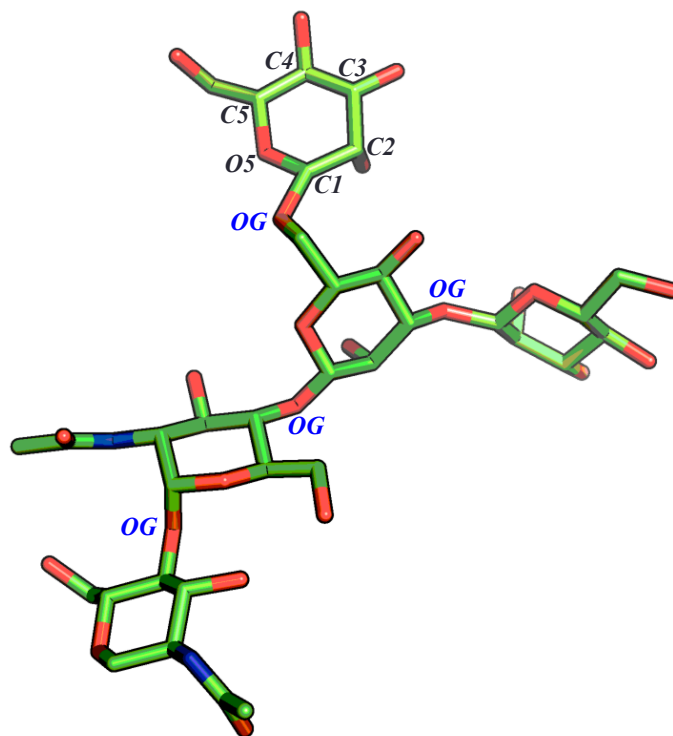


Fig. S2. An example glycan structure to show the atom names used for the definition of GS-score.

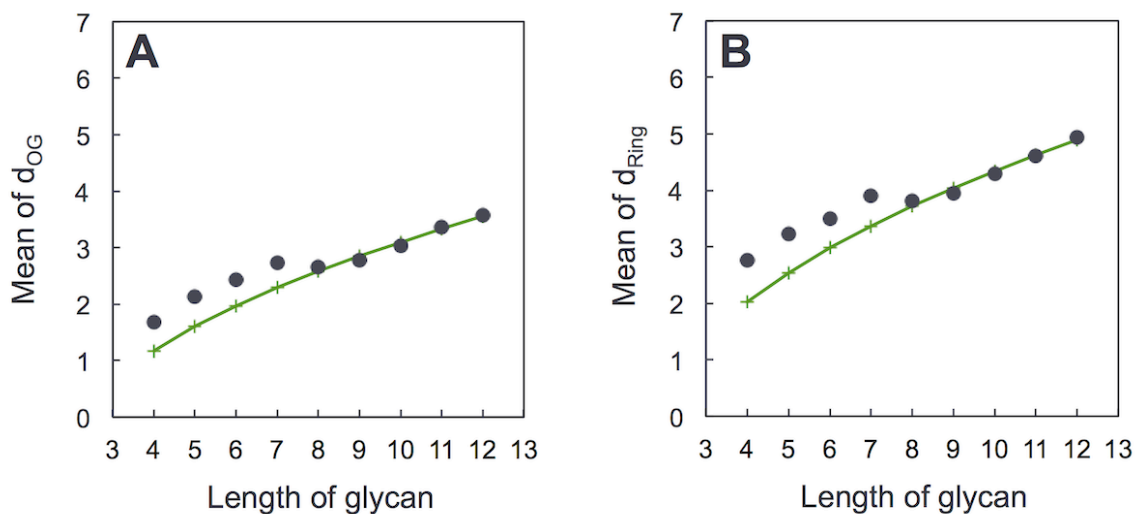


Fig. S3. (A) The average distance between aligned glycosidic oxygen atoms (d_{OG}) and (B) the average RMSD deviation between aligned sugar ring atoms (d_{Ring}) as a function of glycan length. The green lines in each plot are the scaling factors obtained from curve fitting to the average distance or RMSD deviation so that they provide the best normalization of GS-scores.

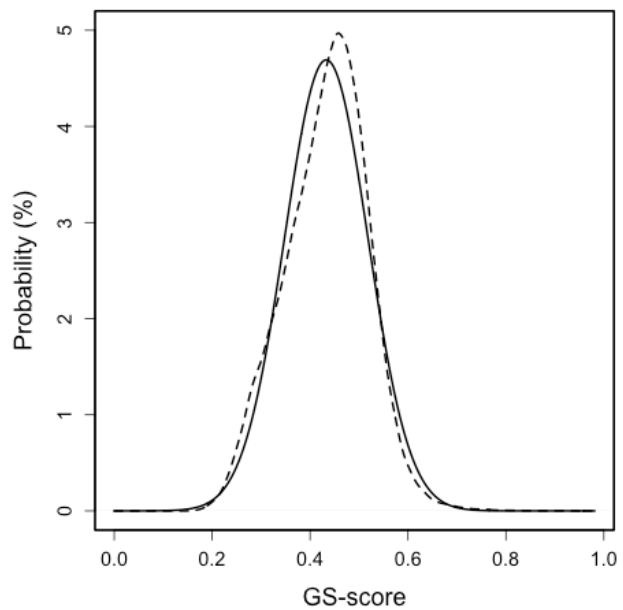


Fig. S4. GS-score distributions from random glycan structures. The dashed line is the observed probability density and the solid line is a direct fit using the normal distribution.

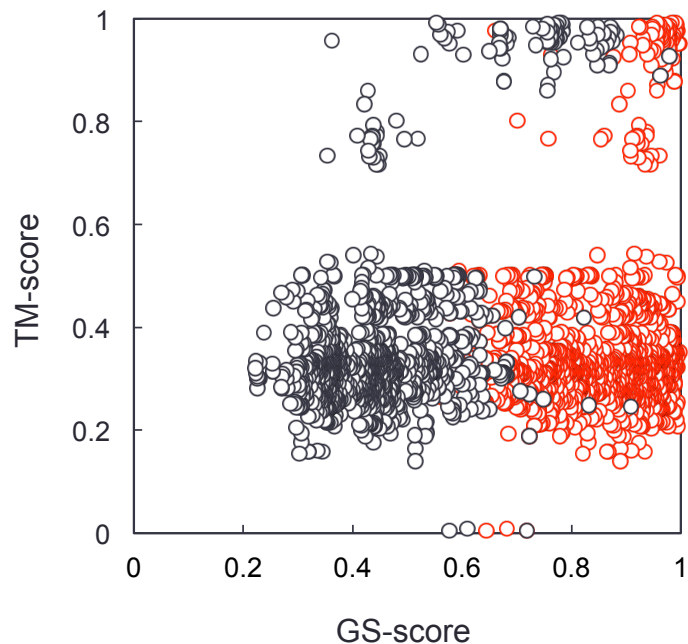


Fig. S5. Comparison of glycan similarity (GS-score) with glycoprotein similarity (TM-score) through a PDB *N*-glycan library search. All PDB *N*-glycans and their parent proteins were structurally compared with the target glycan and its parent glycoprotein (PDB:1L6X_A). In the plot, the black and red circles represent the GS-scores normalized by larger and smaller glycan structures among each pair, respectively.