**Method**

We downloaded 562 UnicarbKB-currated(Campbell et al. 2014) human glycans observed on human proteins from GlyGen (GLYDS000040 v1.4.5)(Kahsay et al. 2020). Using glycompare v1.0, substructures were extracted from each glycan (glycompare.extract\_substructures.extract\_substructures\_pip) and then merged (glycompare.merge\_substructure\_vec.merge\_substructure\_dict\_pip) using structure information only (**Table S X A. Linkage Ambiguous**) and both structure and linakge information (**Table S X B. Linkage Specific**). Finally, 10,140 common linkage-ambiguous substructures and 23,093 linkage-specific substructures were merged (glycompare.merge\_substructure\_vec.substructure\_matching\_wrapper).

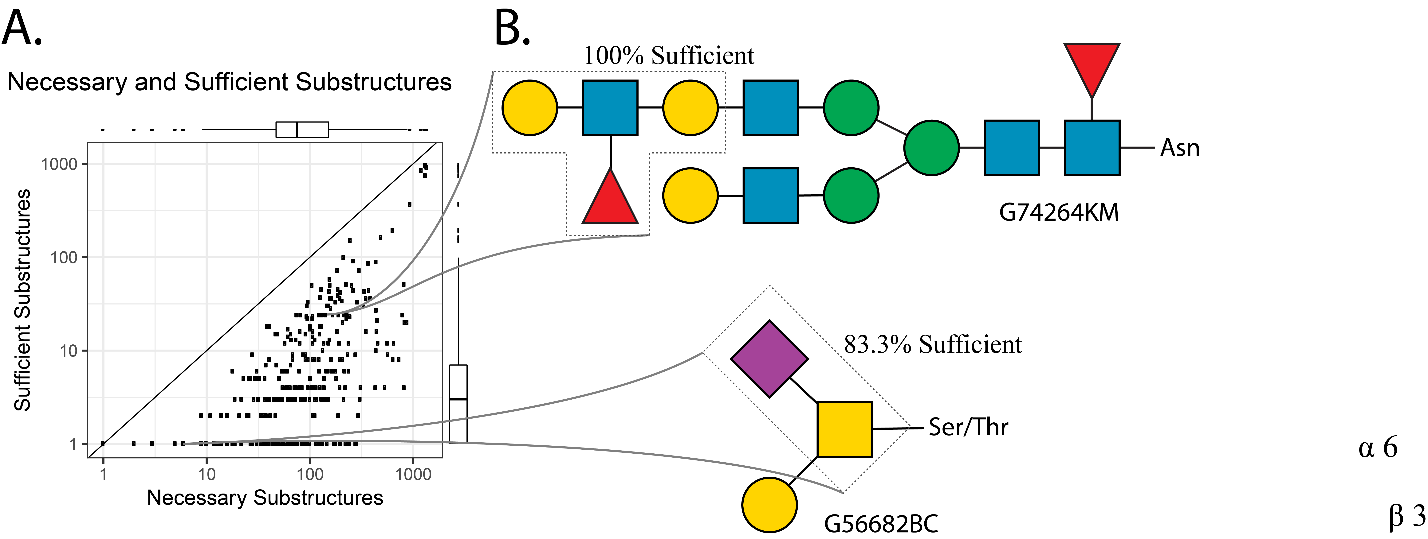


Figure – Necessary and sufficient substructures to specify a glycan superstructure. A. Scatter plot of glycan superstructures (e.g. G56682BC and G74264KM) indicating the number of necessary (substructures that must be present in the superstructure) and sufficient (substructures indicative of the superstructure). The median glycan has three 100% sufficient substructures (that specify it perfectly in the examined glycan set) and 75 necessary substructures. B. Two of over 500 glycan superstructures described in panel A with corresponding sufficient substructures (dotted line) highlighted.

**Results: Most glycans are sufficiently specified by a substructure**

To examine Glycompare’s broad flexibility, we extracted 10,140 linkage-ambiguous glycan substructures and 23,093 linkage-specific substructures from 562 human glycans in UnicarbKB. These include glycopeptide-conjugated N- and O- glycans, poly-lacNAc, and modified (e.g. sulfation) glycans. We used these glycan-substructure relations to demonstrate how substructures can provide necessary and/or sufficient specification of a glycan within a set of mixed glycan types **(Figure XA, Supplemental Table X)**. For example, within human peptide-conjugated glycans, 83.3% of glycans containing a galactose-sialic acid disaccharide (a-D-Neup5Ac-(2-6)-a-D-Galp2NAc) where built on a 6-sialyl core-1 O-glycan **(Figure XB)**. Therefore, the galactose-sialic acid disaccharide is necessary to build a 6-sialyl core-1 O-glycan and 83.3% sufficient to guarantee that the final structure will contain a 6-sialyl core-1 O-glycan. While these results are partially obvious, the necessary and sufficient glycan-substructure relations should mitigate uncertainty when constructing a glycome from substructures.

**Additional references**

Campbell, Matthew P., Robyn Peterson, Julien Mariethoz, Elisabeth Gasteiger, Yukie Akune, Kiyoko F. Aoki-Kinoshita, Frederique Lisacek, and Nicolle H. Packer. 2014. “UniCarbKB: Building a Knowledge Platform for Glycoproteomics.” *Nucleic Acids Research* 42 (Database issue): D215-21.

Kahsay, Robel, Jeet Vora, Rahi Navelkar, Reza Mousavi, Brian C. Fochtman, Xavier Holmes, Nagarajan Pattabiraman, et al. 2020. “GlyGen Data Model and Processing Workflow.” *Bioinformatics (Oxford, England)* 36 (12): 3941–43.