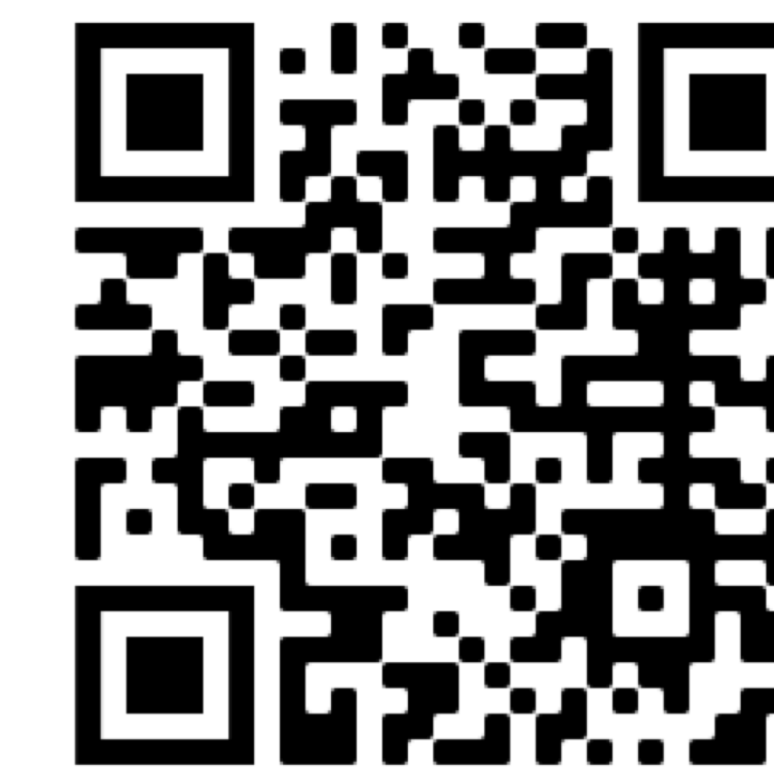


GlyGen collaborations with NCBI to enhance glycoscientist interactions with PubChem and RefSeq

Rahi Navelkar¹, Jeet Vora¹, Nathan Edwards¹, Paul Thiessen², Terence Murphy², Evan Bolton², Tiejun Cheng², Raja Mazumder¹

¹The GlyGen Consortium; ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.



Abstract

Growing recognition of the importance of altered protein glycosylation for the pathophysiology of human diseases is driving the development of accessible bioinformatics resources in glycosciences. As a result, PubChem and RefSeq are collaborating with glycoinformatics resources such as GlyGen to enrich NCBI's data collection and link existing genes and proteins with glycan structures and functions. This poster outlines on-going and future plans co-developed by the GlyGen consortium and NCBI resources designed to highlight and identify glycan records within PubChem, facilitating access by the biomedical research community, and improving linkage to proteins within RefSeq.

Objectives

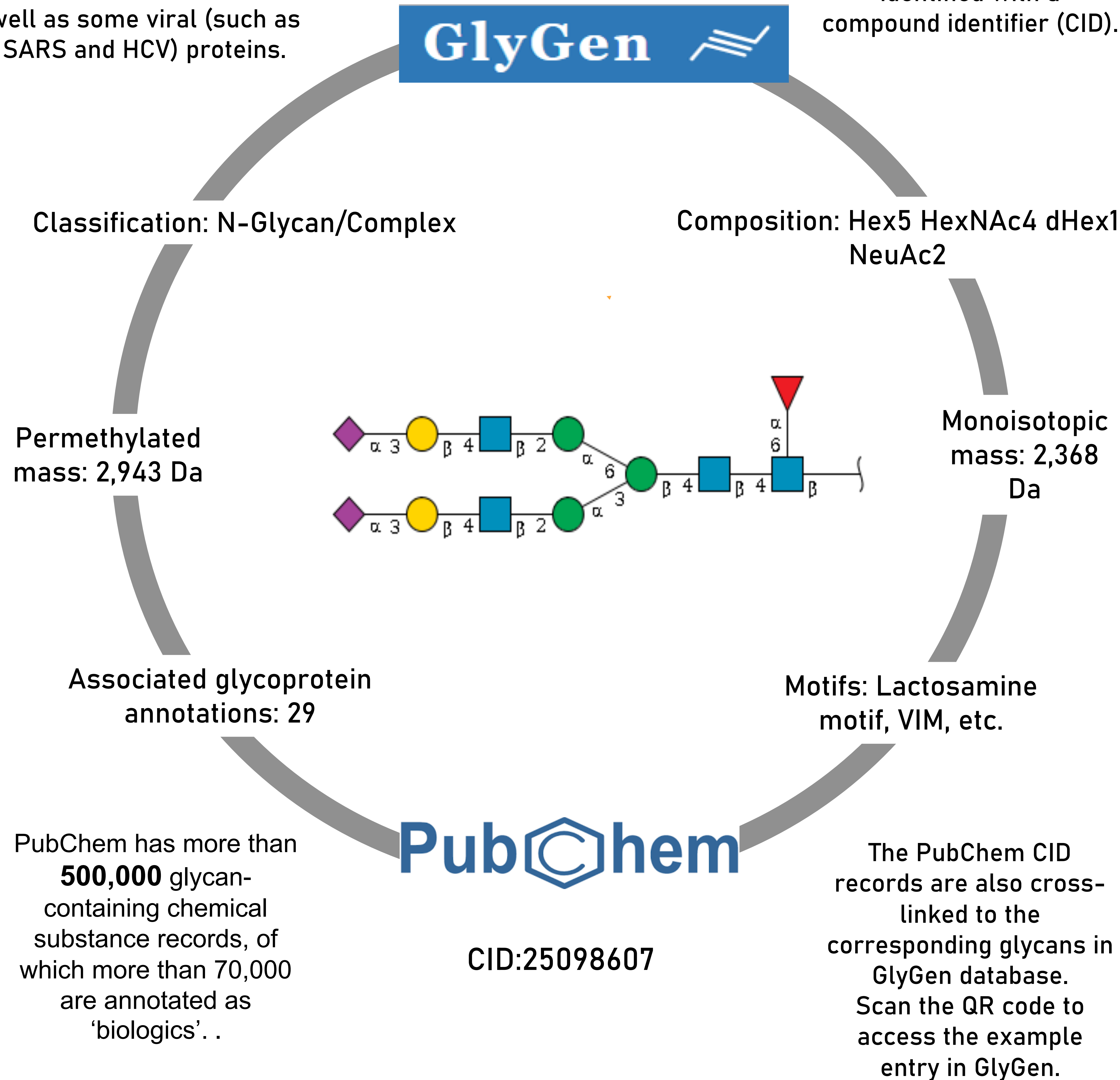
- Connect GlyGen and PubChem resources at the glycan level.
- Highlight and identify glycan records within PubChem.
- Submit glycan annotations such as classification, motif, etc.
- Submit glycosylation information to PubChem Protein pages.
- Make glycan-centric searches more prominent in PubChem database.

Results

Currently, GlyGen has a set of ~30,000 glycans (a subset of GlyTouCan) which includes glycans reported to be associated on mammalian as well as some viral (such as SARS and HCV) proteins.

GlyTouCan: G17689DH

Of the total set of GlyGen glycans, ~10,000 are already present in PubChem and each is identified with a compound identifier (CID).



Conclusions

- In addition to the glycan annotations submitted to PubChem compound pages, GlyGen is also planning to submit glycoprotein annotations such as glycosylation sites, glycan with a GlyTouCan accession or PubChem CID (wherever available) on PubChem protein pages.
- All the submitted annotations are linked back to GlyGen where users can access additional detail about the glycan and protein and also track the source for submitted annotation through the provided PubMed ID.
- By submitting such annotations as well as cross-linking both resources at the compound level, it is expected that returned results will be improved and the barrier for both glycobiology experts and non-experts in using these resources will be significantly lowered.

Acknowledgements

- This work was supported in part by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.
- The GlyGen project is supported by an NIH Common Fund grant (U01 GM125267-01).