Integration of glycan-related disease information and ECM protein-GAG interaction data into the GlyCosmos Portal

Ran Fukazawa[1], Tamiko Ono[2], Kiyoko F. Aoki-Kinoshita[2]
[1]Life Sciences Concentration, Soka University of America, Aliso Viejo, CA
[2]Glycan & Life System Integration Center (GaLSIC), Soka University, Tokyo, JAPAN

ABSTRACT

Glycans have been found to play vital roles in all living organisms, and databases have been developed to collate glycan information for easier access to glycan information. However, much glycan-related information in the scientific literature is not collated in those databases. Moreover, databases with glycan-related information have been developed independently and are rarely cross-linked or cross-referenced, which makes it hard for the users to find only the glycan information.

The **GlyCosmos Portal**[1] aims to integrate all glycan-related information scattered across multiple databases and literature into one Web portal. Therefore, to enrich the glycan information in the GlyCosmos Portal, we integrated glycan-related disease data from the **Alliance of Genome Resources**[2] and protein-glycosaminoglycan (GAG) interaction data from **MatrixDB**[3]. We obtained these data and related resources from the GlyCosmos, FlyGlycoDB, KEGG, and the GlycoGene Database GGDB. Because the data in GlyCosmos is stored using Semantic Web technology, the curated data was converted to ***Resource Description Framework (RDF)** format based on the Human Disease Ontology[4], the OBO Relations Ontology[5], and the MatrixDB identifiers. The RDF format data was then integrated into and made browsable under the "Pathways/Interactions/Diseases/Organisms" page in GlyCosmos.

Consequently, glycan-related disease data stored in the Alliance of Genome Resources can now be searched in GlyCosmos with gene names, gene IDs, and species names responsible for the related disease names as well as the association types, evidence code names and references. Regarding carbohydrate-protein biomolecular interaction data from MatrixDB, protein-binding gene names, their IDs and species names can be found easily in GlyCosmos. Additionally, each gene in GlyCosmos has a detailed summary of gene- and glycan-related information.

*RDF is a machine readable format described in subject-predicate-object triple.

METHODS

This figure illustrates the flow of the Alliance of Genome Resources and MatrixDB data integrated into the GlyCosmos Portal.

1. Convert into RDF triples and upload → Virtuoso/Apache Jena Fuseki RDFdata Store
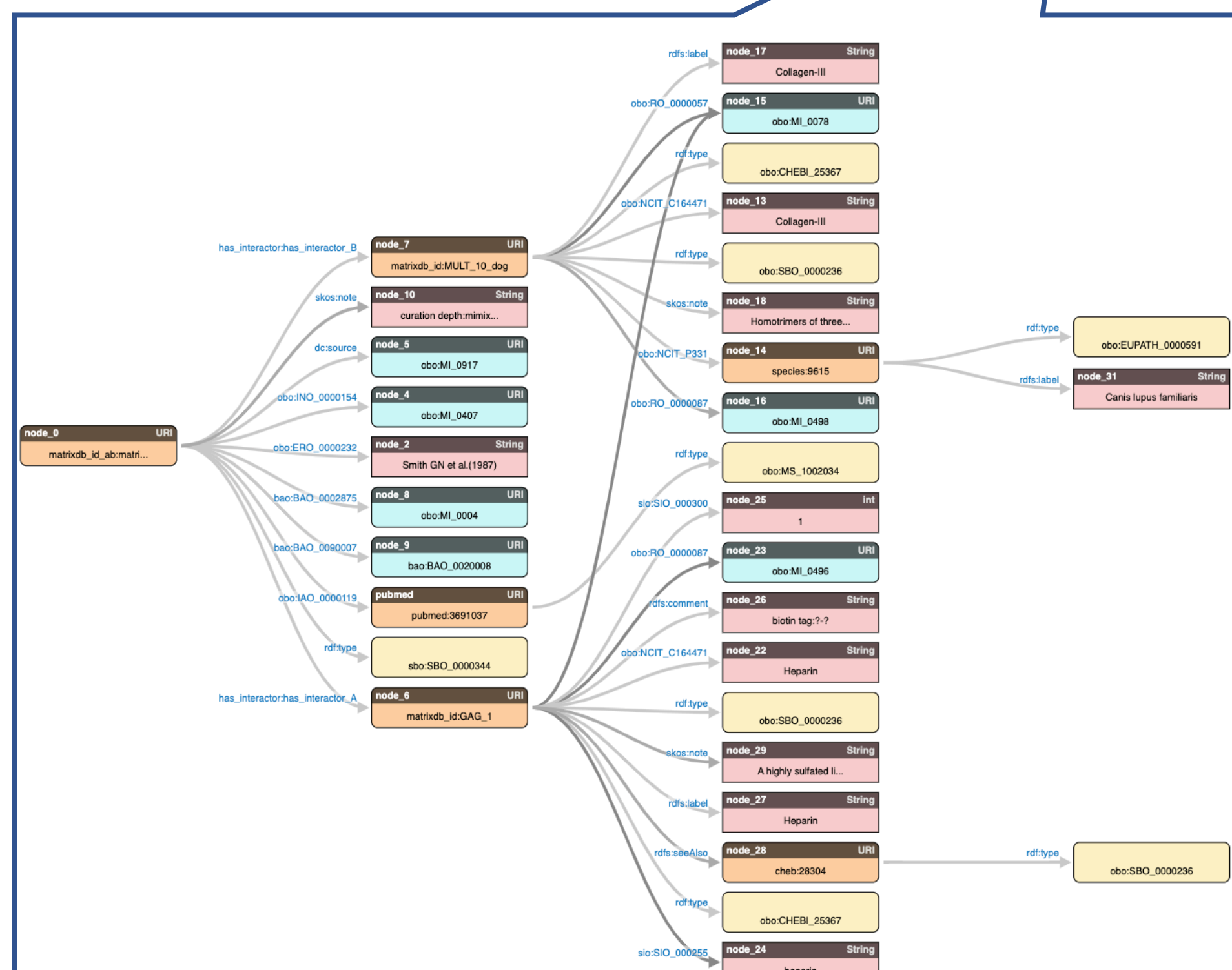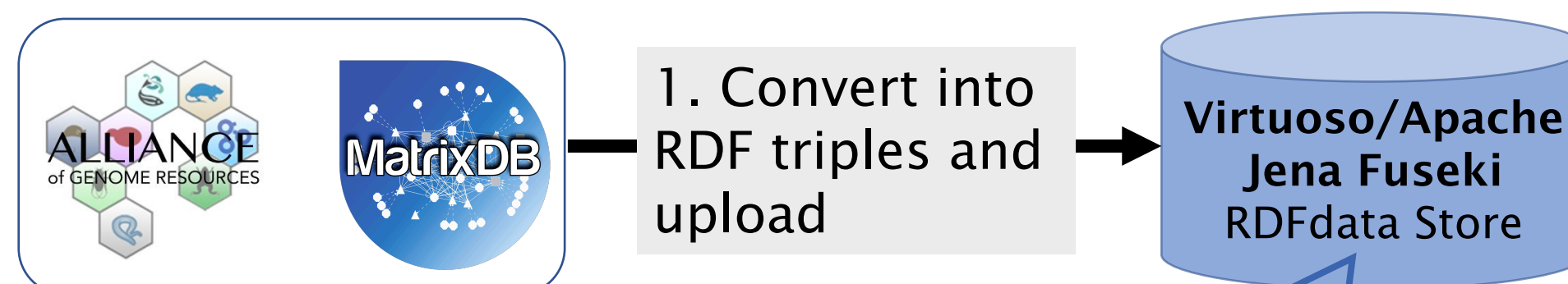
Ex: Schema of the interaction between Heparin (GAG_1) and Collagen-III (MULT_10_dog) from MatrixDB.
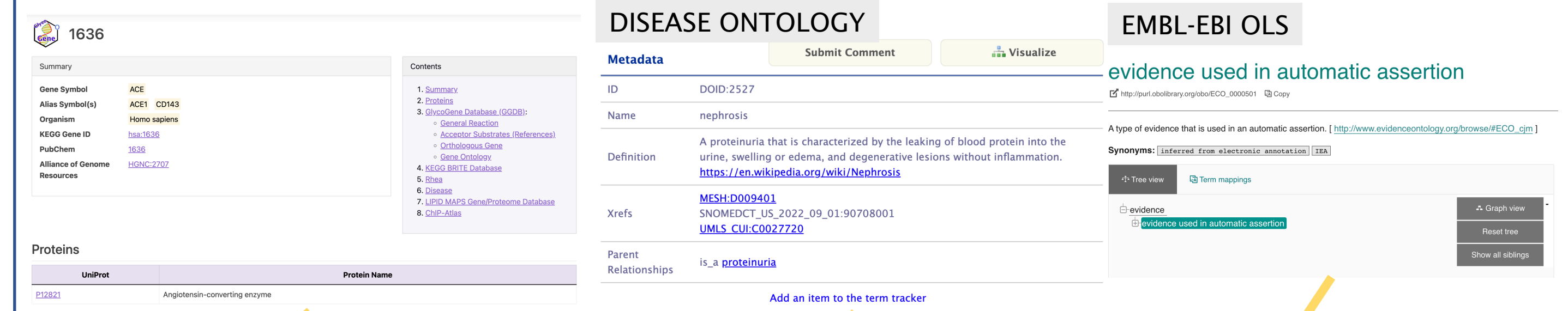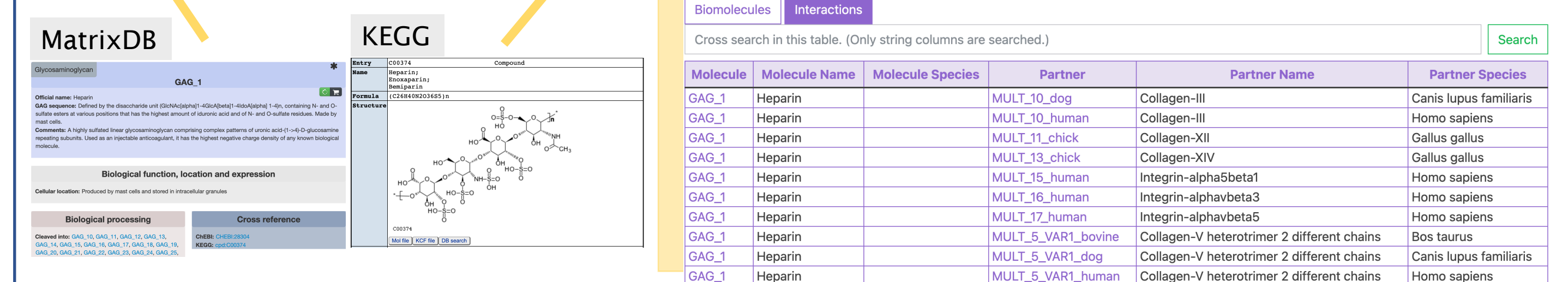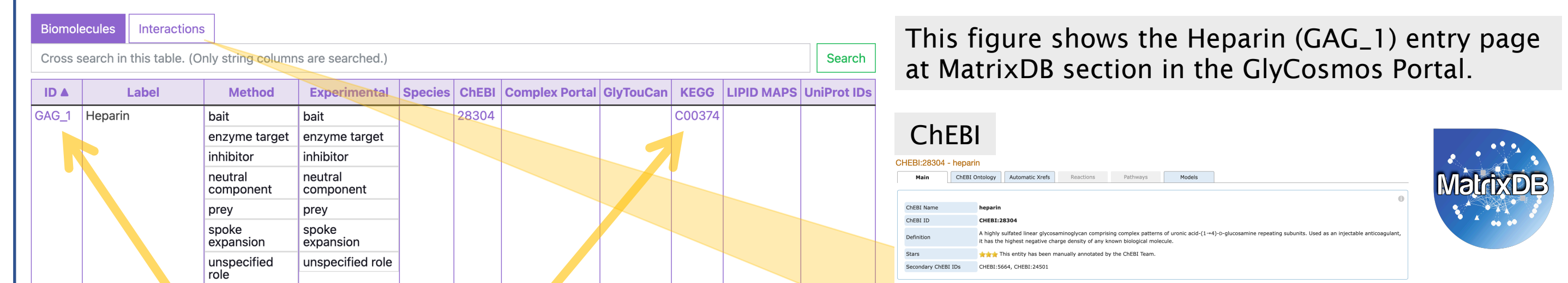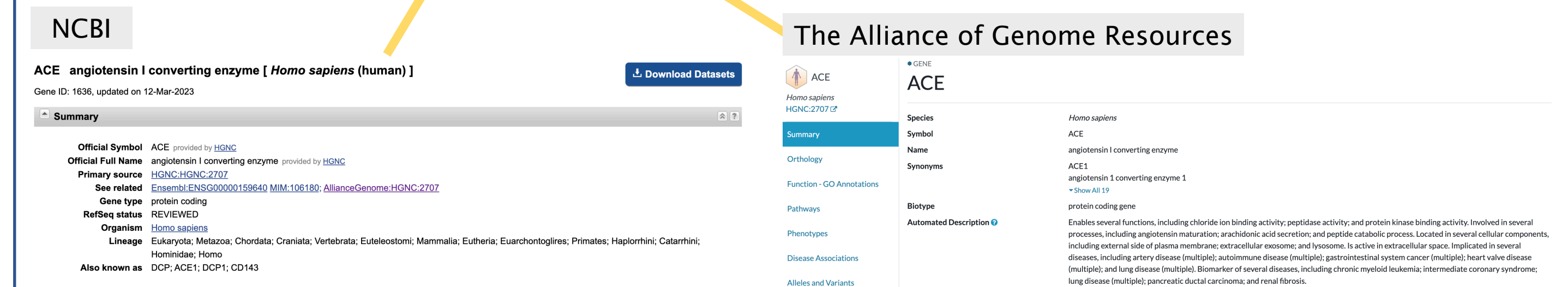
2. Query, retrieve and return RDF data using SPARQL

3. Integrate ← SPARQLlist

RESULTS

This figure shows the ACE entry page at the Alliance of Genome Resources section in the GlyCosmos Portal.

The Alliance of Genome Resources

This figure shows the Heparin (GAG_1) entry page at MatrixDB section in the GlyCosmos Portal.

FUTURE WORK

Populate additional data from other databases into the GlyCosmos Portal to further enrich glycan information with related omics data.

REFERENCES

[1] Yamada I, Shiota M, Shinmachi D, Ono T, Tsuchiya S, Hosoda M, Fujita M, Aoki NP, Watanabe Y, Fujita N, et al. 2020. The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences. Nature Methods. 17(7):649–650. doi:https://doi.org/10.1038/s41592-020-0879-8.
[2] Agapite J, Albou L-P, Aleksander S, Argasinska J, Arnaboldi V, Attrill H, Bello SM, Blake JA, Blodgett O, Bradford YM, et al. 2019. Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Research. 48(D1):D650–D658. doi:https://doi.org/10.1093/nar/gkz813.
[3] Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N, Ricard-Blum S. 2018. MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Research. 47(D1):D376–D381. doi:https://doi.org/10.1093/nar/gky1035.
[4] Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, Baron J, Jackson R, Bello S, Bearer C, et al. 2021. The Human Disease Ontology 2022 update. Nucleic Acids Research. 50(D1):D1255–D1261. doi:https://doi.org/10.1093/nar/gkab1063.
[5] OBO Relations Ontology < Ontology Lookup Service < EMBL-EBI. https://www.ebi.ac.uk/ols/ontologies/ro.