Collaborative Development of GlyTouCan: an International Glycan Structure Repository

国際糖鎖構造リポジトリの開発

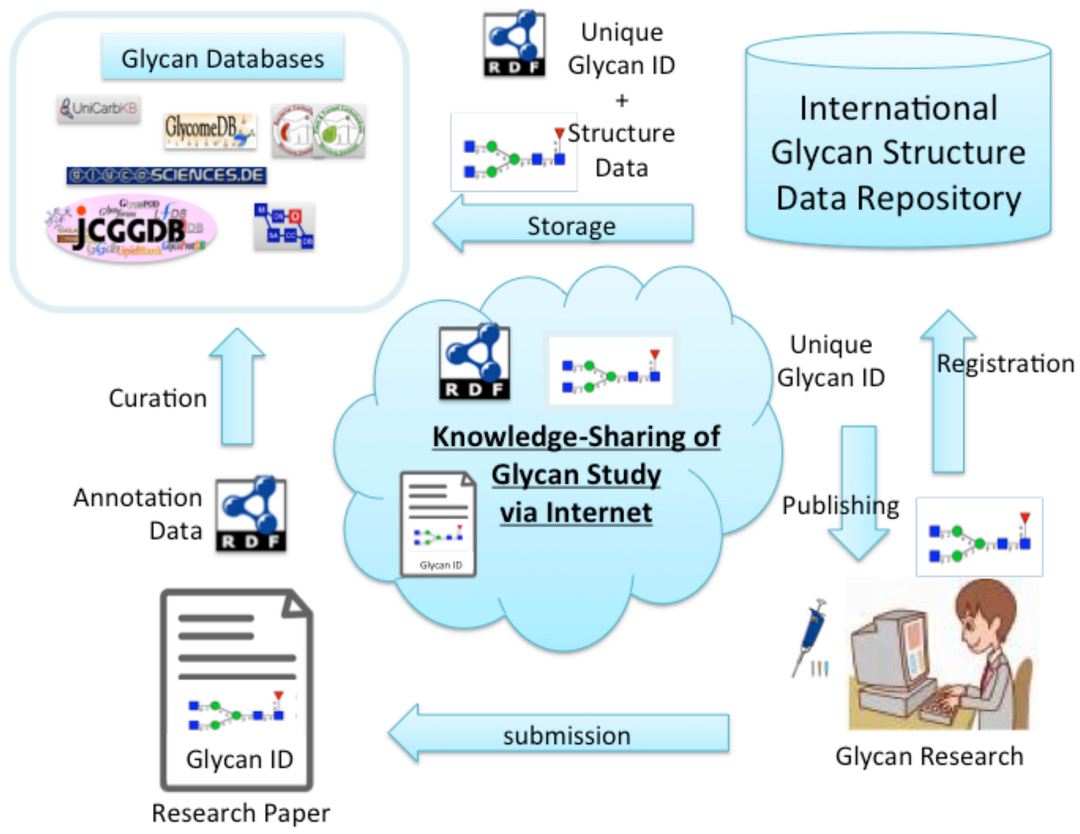
Kiyoko Aoki-Kinoshita[1,2], Nobuyuki Aoki[1], Daisuke Shinmachi[1], Akihiro Fujita[1], Shinichiro Tsuchiya[1], Masaaki Matsubara[3], Issaku Yamada[3], Shujiro Okuda[4], Toshihide Shikanai[2], Noriaki Fujita[2], Elena Solovieva[2], Yoshinori Suzuki[2], Toshisuke Kawasaki[5] and Hisashi Narimatsu[2]

[1]Soka Univ.; [2]AIST; [3]The Noguchi Institute; [4]Niigata Univ.; [5]Ritsumeikan Univ

Researchers throughout the world now use international repositories such as GenBank and PDB to obtain unique IDs for gene sequences and protein structures, respectively, registered into their databases. These IDs can then be referenced in journal manuscripts. However, no such international repository exists for glycan structures (sequences), and because many glycan-related databases have been developed in recent years, the need for such a repository has become urgent.

Thus, we have started a collaboration with glycoscientists in the United States to develop an international glycan structure repository. Using the glycan structure registration system developed by the U.S. GlySpace project, our project focuses on developing a user-friendly web interface called GlyTouCan. When a glycan is being registered, the system checks to see if it already exists, and if it is new, a new ID along with the registering user is assigned to the structure and stored. Moreover, by providing an easy-to-use search interface, any user can easily find their glycan of interest based on mass, motif, monosaccharide composition, etc.

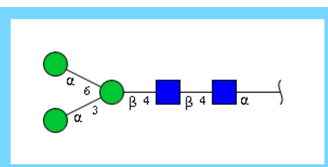
Furthermore, by utilizing Semantic Web technologies, we aim to be able to integrate the data in GlyTouCan with glycan-related data such as proteins from UniProt's RDF data and lipids. Therefore, we hope to be able to provide a central resource that serves as a knowledgebase for glycan structures and related information. GlyTouCan can be accessed from http://www.glytoucan.org/.



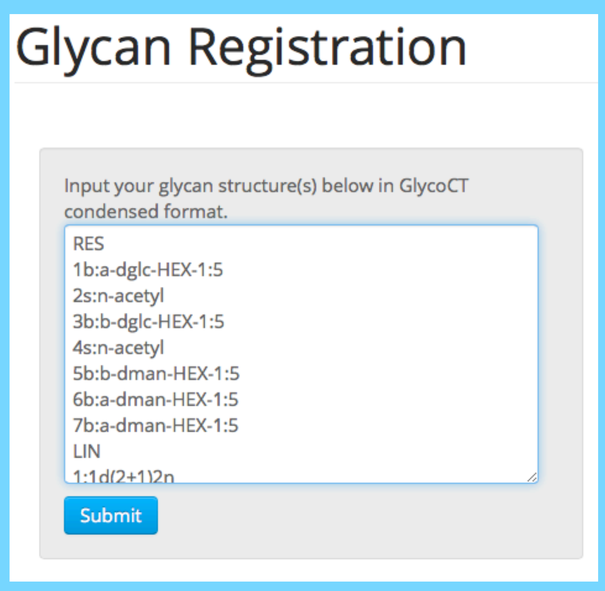
Glycan Registration

GlyTouCan uses GlycoCT format as the main glycan structure representation for its data. GlycoCT basically handles glycans as graphs, where the RES section specifies monosaccharides and substituents, and the LIN section describes linkages.

GlycoCT is designed to generate unique strings for each glycan structure by defining rules by which to order residues and linkages, etc.



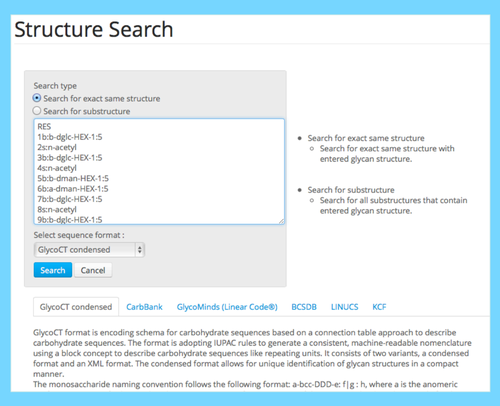
```
RES
1b:a-dglc-HEX-1:5
2s:n-acetyl
3b:b-dglc-HEX-1:5
4s:n-acetyl
5b:b-dman-HEX-1:5
6b:a-dman-HEX-1:5
7b:a-dman-HEX-1:5
LIN
1:1d(2+1)2n
2:1o(4+1)3d
3:3d(2+1)4n
4:3o(4+1)5d
5:5o(3+1)6d
6:5o(6+1)7d
```



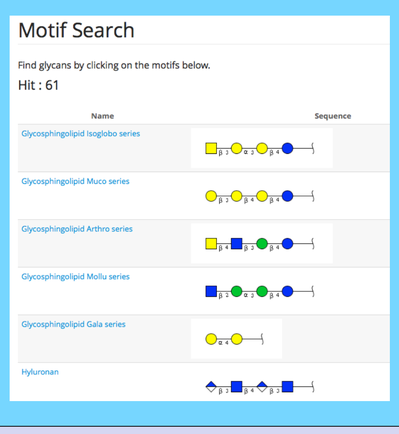
Glycan registration starts with the input of a single glycan structure in GlycoCT format. When the Submit button is pressed, the structure is compared against the currently registered glycans and checked to see if it has already been registered. If so, a message indicating the accession number (formatted as "G" + five random digits + two random alphabetical characters) will be displayed along with the structure drawn as an image. If not, a confirmation displaying the structure as an image is displayed so that the user can confirm the structure they wish to register.

Search

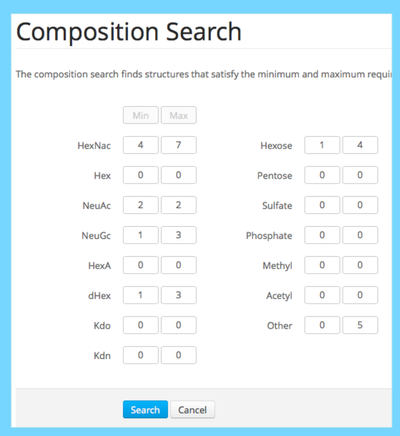
Structure Search



Motif Search

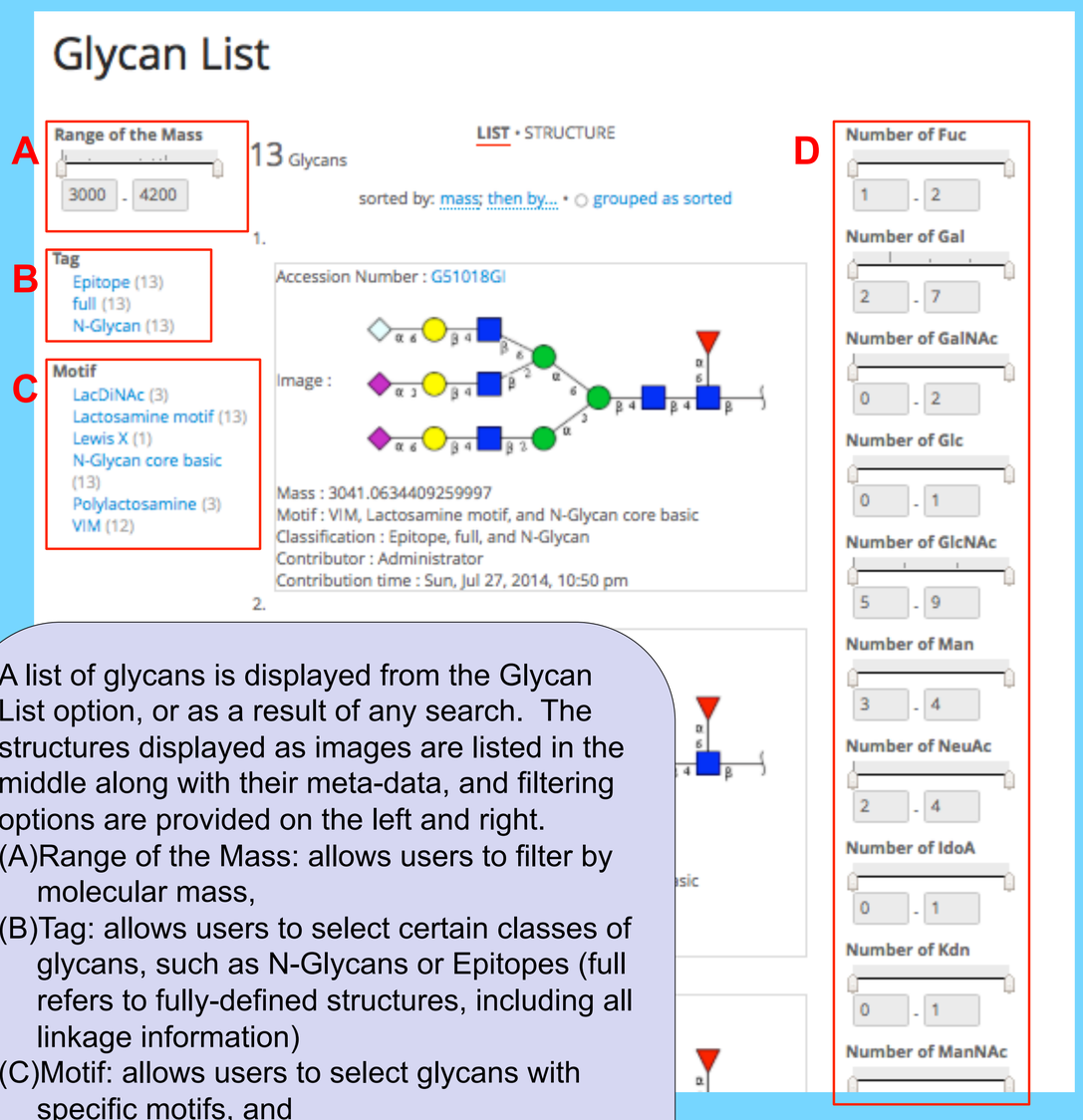


Composition Search



Three ways for search structures are currently provided:

1. Structure search allows users to enter structures in any major glycan representation, including GlycoCT, KCF, Linear Code, etc.
2. Motif search allows users to start from well-known motifs in glycoscience such as Lewis and lactose structures. By selecting a motif from the list of available motifs, users can access the list of glycans that contain it.
3. Composition search is most often used by glycomics analysts who wish to find glycans having a certain monosaccharide composition. Ranges for specific monosaccharides can be specified.



A list of glycans is displayed from the Glycan List option, or as a result of any search. The structures displayed as images are listed in the middle along with their meta-data, and filtering options are provided on the left and right.

(A)Range of the Mass: allows users to filter by molecular mass,

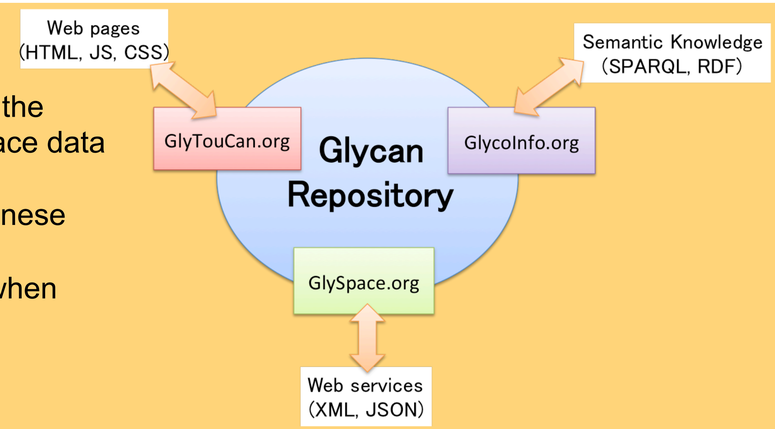
(B)Tag: allows users to select certain classes of glycans, such as N-Glycans or Epitopes (full refers to fully-defined structures, including all linkage information)

(C)Motif: allows users to select glycans with specific motifs, and

(D)allows users to filter by monosaccharide composition.

Results and Future work

- GlyTouCan (version 0.1) has been released as of August, 2014. Since its release, we have been working on the RDFization of the data by developing an ontology and RDFizing its data. We are also working on developing the automatic syncing of the GlySpace data with the triplestore in GlycoInfo.
- New features such as language preferences are currently be developed, including Chinese traditional, Chinese simplified, Japanese and Russian.
- By next year, we should be able to contact journals to consider using this system to require authors to register their structures when publishing manuscripts related to glycan structures.



Acknowledgements

This research was supported by National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST), and National Institute of Advanced Industrial Science and Technology (AIST).

Licensed under a Creative Commons 表示2.1日本 license
(c)2014 新町大輔(創価大学)