**Glycobiology** vol. 23 no. 12 pp. 1422–1424, 2013
doi:10.1093/glycob/cwt084

# Glyco-Forum section

## The Fifth ACGG-DB Meeting Report: Towards an International Glycan Structure Repository

**Kiyoko F Aoki-Kinoshita[1], Hiromichi Sawaki[2], Hyun Joo An[3], Matthew Campbell[4], Qichen Cao[5], Richard Cummings[6], Daniel K Hsu[7], Masaki Kato[8], Toshisuke Kawasaki[9], Kay-Hooi Khoo[7], Jaehan Kim[3], Daniel Kolarich[10], Xianyu Li[5], Mingqi Liu[11], Masaaki Matsubara[12], Shujiro Okuda[9,13], Nicolle H Packer[4], René Ranzinger[14], Huali Shen[11], Toshihide Shikanai[2], Daisuke Shinmachi[2], Philip Toukach[15], Issaku Yamada[12], Yoshiki Yamaguchi[8], Pengyuan Yang[11], Wantao Ying[5], Jong Shin Yoo[16], Yan Zhang[17], Yang Zhang[11], and Hisashi Narimatsu[2]**

[1]Soka University, Tokyo, Japan; [2]National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan; [3]Chungnam National University, Daejeon, Korea; [4]Macquarie University, Sydney, NSW, Australia; [5]Beijing Institute of Radiation Medicine, Beijing, China; [6]Emory University, Atlanta, GA, USA; [7]Academia Sinica, Taipei, Taiwan; [8]RIKEN Global Research Cluster, Wako-shi, Saitama, Japan; [9]Ritsumeikan University, Kusatsu, Shiga, Japan; [10]Max Planck Institute of Colloids and Interfaces, Potsdam, Germany; [11]Fudan University, Shanghai, China; [12]The Noguchi Institute, Tokyo, Japan; [13]Niigata University, Niigata, Japan; [14]The University of Georgia, Athens, GA, USA; [15]Zelinsky Institute of Organic Chemistry, Moscow, Russia; [16]Korea Basic Science Institute, Daejeon, Korea; and [17]Shanghai Jiao Tong University, Shanghai, China

The Research Center for Medical Glycoscience at AIST (National Institute of Advanced Industrial Science and Technology) has been holding ACGG-DB (Asian consortium for glycobiology and glycotechnolgy—database) meetings since 2011, with the first international ACGG-DB meeting where participants from the USA, Australia, and Germany attended in addition with Asian representatives in Okinawa in 2012, as reported previously (Aoki-Kinoshita et al. 2013). As a follow-up to this meeting, the Fifth ACGG-DB meeting was held in Dalian, China, on 22 June 2013, where representatives from the USA, Australia, Russia, and Germany also attended (Group photo; Figure 1).

The purpose of this meeting was to gain a consensus on the framework for an international glycan structure repository. After presentations by representative participants, the main part of the meeting was the brainstorming session focusing on the glycan repository framework. It was first agreed upon that such a repository was needed, as has been recognized by many groups including the US National Academy of Sciences (National Academy of Sciences 2012) and as discussed at the 4th Charles Warren Workshop in 2012 in Athens, Georgia. The advantages of such a repository would be that: (a) it would function as a central location where glycan structures are registered, similar to DDBJ/EMBL/GenBank of the International Nucleotide Sequence Database Collaboration (INSDC) for nucleotide sequences, (b) it would provide unique identifiers for every glycan structure (including ambiguous structures and even monosaccharide compositions) such that glycan IDs can be uniquely identified from any resource by using these identifiers, and (c) researchers who publish newly identified structures can use the identifiers to link their structures with relevant information in their publications.

The next issue was to determine the data content, or scope, of the database: (a) whether to store glycan structures and/or glycoconjugates, and (b) whether to include additional information (called "metadata") to supplement the structure information, such as biological source, experimental procedures used to detect the structure, and so on. It was noted that this repository should be distinct from existing curated databases such as UniCarbKB (Campbell et al. 2011) and BCSDB (Toukach 2011) and that this repository was required as a resource for assigning unique identifiers to each structure. Thus, new structures that are found would be assigned a unique identifier and can be assured novelty when published. In other words, the main purpose of this repository would be to assure uniqueness of all registered structures guaranteeing that new structures at any level of detail or uncertainty will have an identifier assigned such that readers can retrieve the exact structure described in a paper. Therefore, at the minimum, it was agreed that glycan structures (with no aglycon information) would be the main content of this repository, and the user and time/date of registration would be attached with the structure information. Any additional metadata information and experimental data would be the responsibility of the author when publishing the experimental procedures used to define the structure (as would be covered by MIRAGE; Kolarich et al. 2013); that is, it would be expected that such relevant information should be submitted to a database able to store these data. This latter type of database, as well as existing carbohydrate structure databases and publications, will then be referenced by the registry allowing users of the registry not just to retrieve for any identifiers the structure itself but also links to complementary data in other resources.

It is expected that the assurance of uniqueness of all glycan structures registered can be resolved by using the WURCS (Web3.0 Unique Representation of Carbohydrate Structures) format for every carbohydrate sequence. WURCS was


**Fig. 1** Group photo of the participants of the 5th ACGG-DB Meeting.

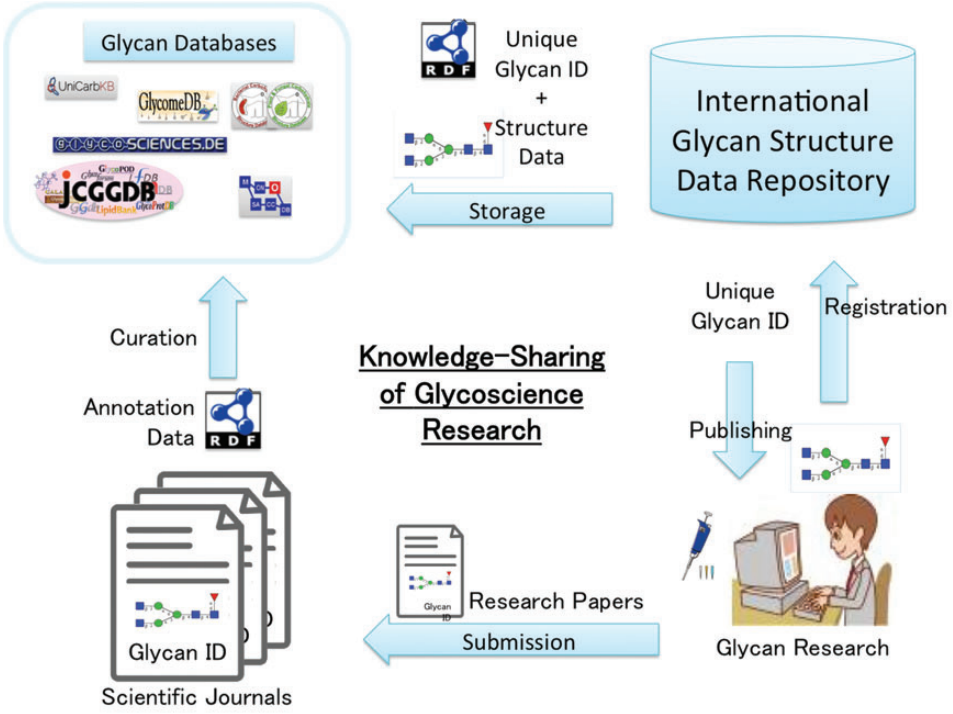


**Fig. 2** Overview of the flow of glycomics data with the development of an International Glycan Structure Data Repository.

proposed at this meeting as a format that can accommodate even stereoisomeric structures and glycan compositions uniquely as a text string. This format can serve as the foundation for the glycan repository to guarantee uniqueness of any glycan structure that may be published in the literature based on the sequence format. Although this format will soon be made public, because it was still under development at the time of this meeting, it was decided by the participants that the decision to support this format for the glycan repository would be put on hold until it became available.

It was pointed out that the identifiers used in the repository should somewhat reflect the structure to which it has been assigned. For example, using "N" or "O" in the identifier would be useful as it could indicate that the structure was an *N*-linked or *O*-linked glycan, respectively. However, determining the type of structure in itself would entail the addition of such metadata information, which is beyond the scope of the repository. Therefore, the possibility of building a semi-curated database based upon the repository data was proposed. Metadata information such as glycan type and biological source could be added (after publication), similar to UniProt entries with "gold stars" indicating curated protein entries (The UniProt Consortium 2012).

The next question discussed at the meeting concerned who would be responsible for developing and maintaining this repository. This topic also entailed the question of funding options for the glycoscience community. Although several glycoscience projects in the USA and Australia were underway; currently, it seemed that the JCGGDB project seemed to be most promising

to lead the development of this repository. It also seemed to fit in well with the goals of the Integrated Database Project being led by the Japanese government.

Next, technical details regarding the user interface for the system were discussed. First, it was agreed that it would be most appropriate to develop this repository as a part of the Semantic Web, as agreed upon in Okinawa (Aoki-Kinoshita et al. 2013). Therefore, the data would be stored in RDF (Resource Description Framework) format in a triplestore. The following user functionality was also deemed to be necessary:

- Identifier assignment
- Ability to add annotations
- Web services to enable software programs to access the data
- GUI for the display of structures and associated information
- Search functionality such as search by monosaccharide composition, mass, (sub) structure, (range of) date(s) registered, user or structure identifier
- Glycan structure drawing tool (such as GlycanBuilder; Damerell et al. 2012)

Figure 2 illustrates an overview of how the glycan structure repository would function. Scientists who wish to publish their newly discovered glycan structure(s) would register them into the repository such that their unique identifiers can be referenced in their publication. After publication, glycan databases can link the identifiers with relevant annotation information.

Finally, discussions on how and where to promote this repository was held. Collaborations with the MIRAGE working group would also be essential. In the end, it was agreed that once a prototype was developed and made available, more concrete discussions could be made in this regard.

## Funding

This meeting is a part of the JCGGDB project, which is supported by JST (Japan Science and Technology Agency) and NBDC (National Bioscience Database Center) Program for the Life Science Database Integration Project.

## Acknowledgements

All participants appreciate Madoka Ishizaki for her help to manage this meeting.

## References

Aoki-Kinoshita KF, Sawaki H, An HJ, Cho JW, Hsu D, Kato M, Kawano S, Kawasaki T, Khoo KH, Kim J, et al. 2013. The Third ACGG-DB Meeting Report: Towards an international collaborative infrastructure for glycobioinformatics. *Glycobiology*. 23:144–146.

Campbell MP, Hayes CA, Struwe WB, Wilkins MR, Aoki-Kinoshita KF, Harvey DJ, Rudd PM, Kolarich D, Lisacek F, Karlsson NG, et al. 2011. UniCarbKB: Putting the pieces together for glycomics research. *Proteomics*. 11:4117–4121.

Damerell D, Ceroni A, Maass K, Ranzinger R, Dell A, Haslam SM. 2012. The GlycanBuilder and GlycoWorkbench glycoinformatics tools: Updates and new developments. *Biological Chemistry*. 393:1357–1362.

Kolarich D, Rapp E, Struwe WB, Haslam SM, Zaia J, McBride R, Agravat S, Campbell MP, Kato M, Ranzinger R, et al. 2013. The minimum information required for a glycomics experiment (MIRAGE) project: Improving the standards for reporting mass-spectrometry-based glycoanalytic data. *Molecular and Cellular Proteomics*. 12:991–995.

National Research Council (US) Committee on Assessing the Importance and Impact of Glycomics and Glycosciences. 2012. *Transforming Glycoscience: A Roadmap for the Future*. The National Academies Press.

The UniProt Consortium. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*. 40:D71–D75.

Toukach PV. 2011. Bacterial carbohydrate structure database 3: principles and realization. *J Chem Inf Model*. 51:159–170.