OXFORD

Glyco-Forum

Meeting Report of the International Life Science Integration Workshop 2018

The first International Life Science Integration Workshop was held on 5–9 March 2018 in Tokyo, Japan. This workshop was sponsored by the Glycoinformatics Consortium (GLIC), Japan Science and Technology Agency (JST) and the National Bioscience Database Center of Japan (NBDC). The first 2 days of this workshop consisted of a symposium with invited talks from experimentalists and bioinformatics experts in the life sciences. The last 3 days were devoted to a Hackathon to further develop communication between the different software projects and support the integration of different life science domains. This was one of the first meetings where glycomics software developers met with leading proteomics and lipidomics resource experimentalists, and the hackathon sessions provided a forum to discuss strategies for data integration and improved data interoperability in the life science domain. The results of the workshop are presented here.

The workshop web page can be found at http://glic.glycoinfo. org/meetings/LSworkshop2018/. The initial 2-day symposium included international talks introducing the GlyCosmos Project developed in Japan (https://glycosmos.org), the GlyGen Project developed in the USA (http://glygen.org), the UniCarbKB (Campbell et al. 2014) and GlycoStore projects from Australia (Zhao et al. 2018), and the GlyConnect Project developed in Switzerland (Alocci et al. in preparation) and related work at ExPASy (Mariethoz et al. submitted). There were also talks from developers of the Carbohydrate-Active Enzymes Database (CAZy) (Lombard et al. 2014), the Carbohydrate Structure Database (CSDB) (Toukach and Egorova 2016), the GRITS Toolbox (http:// www.grits-toolbox.org/), as well as glycoinformatics resources available at VirtualGlycome.org (Cheng et al. 2017; Liu et al. 2017), ACGG-DB (http://acgg.asia/db/), UniProt (UniProt Consortium 2018), LipidBank (Yasugi and Seyama 2007), PDBj (Protein Data Bank Japan) (Kinjo et al. 2018), services and tools provided by DBCLS (Database Center for Life Science) and jPOST (Japan Proteome Standard Repository/Database) (Okuda et al. 2017). The audience was also introduced to the Minimum Information Required for A Glycomics Experiment (MIRAGE) initiative and their objectives (York et al. 2014). Hisashi Narimatsu presented the technologies used at AIST to produce such glycomics data, and Toshisuke Kawasaki closed the meeting with his experiences in determining glycans' functions using antibodies.

On the second day, a discussion session was held to identify the immediate requirements of the community and the necessary steps to work toward their implementation. One of the needs that arose was access to a single and central location listing existing glycomics and related web resources that are available to the community. As was evident from the many talks of the previous days, currently many databases and web resources are available, but these are often not easy to find. Table I lists the relevant resources that were introduced during the workshop as a reference.

Another need expressed by the meeting participants was the accessibility of published glycomics data. Currently, there is no requirement by journals to register glycomics data into any public repository. Glycomics and glycoproteomics mass spectrometry (MS) analysis generates vast amounts of data that should be stored and made publicly available. However, the necessary infrastructure and technology need to be in place to handle such data. During the hackathon, members of the jPOST and GlyCosmos development team agreed to initiate an extension of jPOST for glycomics and glycoproteomics MS data, called GlycoPOST. Planning and implementation of this repository to accept metadata conforming to the MIRAGE MS guidelines (Kolarich et al. 2013) also made progress during the hackathon. As a part of this activity, discussions took place regarding the requirements for glycomics data in GlycoPOST with Niclas Karlsson (University of Gothenburg, Sweden), who initiated the systematic collection of LC/MS-MS glycan data in a repository called UniCarb-DR (Karlsson et l., manuscript in preparation).

Another topic that arose during the discussion of this workshop was the need to easily understand glycan function from existing data. However, glycan function can be described in a variety of forms; they are recognized by antibodies, lectins bind to them to initiate signaling events, they are used as signals to chaperones to indicate the fitness of protein folding, etc. Oftentimes, glycan recognition may only involve the nonreducing end, meaning the same glycan substructure (or motif or epitope) may be found on different carriers, including *N*-glycans, *O*-glycans and lipids. Glycans are also synthesized by glycosyltransferases, which may have a variety of substrate specificities. All of these information are currently not integrated in any web resource, but the talks at the symposium and the discussion session at the end indicated that several projects are aiming to do so, e.g. GlyConnect, GlyCosmos and GlyGen.

During the hackathon, several activities ensued including the following topics:

- CSDB integration with GlyTouCan, and beta-testing and incorporation of a new glycan-drawing tool called SugarSketcher, developed by the Swiss Institute of Bioinformatics as a replacement of GlycanBuilder.
- Development of Web3 Unique Representation for Carbohydrate Structures (WURCS) to handle modifications and a converter from WURCS to GlycoCT.
- Discussion of tools using glyco-enzymes for the simulation of glycan biosynthesis.
- Discussion of GlycoPOST development, including sample data, metadata ontology and controlled vocabularies to provide a user-friendly interface to deposit glycomics data (detailed above).

Table I. Relevant resources described during this workshop

Resource name (URL)	Description	Resource type
CAZy (www.cazy.org) Lombard et al. (2014)	The carbohydrate-active enzymes database	Carbohydrate-active enzymes: curated family classification
CAZypedia (https://www.cazypedia.org) CAZypedia Consortium (2018)	The encyclopedia of carbohydrate-active enzymes	Curated information on the structure, specificity and mechanism of CAZy families
CSDB: Carbohydrate Structure Database (http://csdb.glycoscience.ru) Toukach and Egorova (2016)	Structure, bibliography, taxonomy and NMR data for prokaryotic, plant and fungal glycans	Glycomics, glycoconjugates and NMR
GLIC: Glycoinformatics Consortium (https://glic.glycoinfo.org/)	Database resources, software tools and repository for glycomics related data	Glycomics, glycoproteins and glycolipids
Glycomics@ExPASy (https://www.expasy.org/glycomics)	Glycomics dedicated section of ExPASy portal	Glycomics and glycoproteomics
Glycopedia (https://www.glycopedia.eu/)	Wikipedia of glycomics information	Glycomics
GlycoStore (https://www.glycostore.org) Zhao et al. (2018)	Glycan structure and experimental database	Glycomics
GlyGen (http://glygen.org/)	Data integration project for glycomics, genomics and glycoproteomics data	Glycomics and glycoproteins
GlyTouCan (https://glytoucan.org) Tiemeyer et al. (2017)	The international glycan structure repository	Glycan structures
GRITS Toolbox (http://www.grits-toolbox.org/)	Annotation program for glycomics MS data	Glycomics, mass spectrometry and data analysis
LipidBank (http://jcbl.jp/wiki/) Yasugi and Seyama (2007)	Database of lipid structures with their original publication records	Glycosphingolipids and glycolipids
MIRAGE (http://www.beilstein-institut.de/en/projects/mirage) Yasugi and Seyama (2007); York et al. (2014)	Minimum reporting initiative for glycomics	Glycomics and standards
UniCarbKB (http://www.unicarbkb.org) Campbell et al. (2014)	Glycoprotein (site-specific and global) and glycan structure annotations	Glycoproteins
VirtualGlycome (http://www.VirtualGlycome.org) Liu et al. (2017); Cheng, Zhou and Neelamegham (2017)	Open source tools for the analysis of glycosylation networks	Glycomics and glycoproteomics
ACGG-DB (https://acgg.asia/db/)	Semantic web of the integrated databases for glycogenes, lectins and glycoproteins	Glycomics and glycoproteomics

Fig. 1. Group photo of the participants of this workshop.

- Discussion of ontology, including the steps to take for updating GlycoRDF, development of a new Glycoconjugate Ontology called GlycoCoO (https://github.com/glycoinfo/GlycoCoO) and plans for publication.
- UniProt-related discussions, including the incorporation of ProtVista (Watkins et al. 2017) web components to display glycoproteins in GlyCosmos and the potential of using UniRules (https://f1000research.com/posters/6-1366) to identify common elements and annotations related to glycoproteins in UniProt.

In summary, much progress was made in stimulating international collaborations between researchers in lipidomics, proteomics, glycomics and informatics at this workshop. With several newly funded bioinformatics projects now underway, it is expected that increased collaborations and integration will be needed and implemented in the next 3–5 years. This will require the continued collaboration with those who attended, as well as more connections with those in other fields, including genomics and metabolomics (Figure 1).

Contributors:

- Issaku Yamada (The Noguchi Institute, Tokyo 173-0003, Japan)
- Sachiko Akase (Graduate School of Engineering, Soka University, Tokyo 192-8577, Japan)

- Kiyohiko Angata (Glycoscience and Glycotechnology Research Group, National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8568, Japan)
- Kazuhiro Aoki (Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, USA)
- Nobuyuki P. Aoki (SparqLite LLC., Tokyo, 192-0032, Japan)
- Masanori Arita (National Institute of Genetics, Mishima 411-8540, Japan; RIKEN Center for Sustainable Resource Science, Yokohama 230-0045, Japan)
- Matthew P. Campbell (Institute for Glycomics, Griffith University, Gold Coast, 4222, Queensland, Australia)
- Nathan Edwards (Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, Washington DC, 20007, USA)
- Akihiro Fujita (Graduate School of Engineering, Soka University, Tokyo 192-8577, Japan)
- Leyla Garcia (European Bioinformatics Institute EMBL-EBI, Hinxton, Cambridgeshire CB10 1SD, UK)
- Susumu Goto (Database Center for Life Science, Research Organization of Information and Systems, Kashiwa 277-0871, Japan)
- Bernard Henrissat (Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, Marseille, France)
- Masae Hosoda (Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan)
- Yasushi Ishihama (Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan)
- Toshisuke Kawasaki (Research Center for Glycobiotechnology, Ritsumeikan University, Shiga 525-8577, Japan)
- Thukaa Kuoka (Graduate School of Engineering, Soka University, Tokyo 192-8577, Japan)
- Genji Kurisu (Institute for Protein Research, Osaka University, Osaka 565-0871, Japan)
- Sunmyoung Lee (Graduate School of Engineering, Soka University, Tokyo 192-8577, Japan)
- Frederique Lisacek (Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland)
- Julien Mariethoz (Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, Geneva, 1211, Switzerland)
- Hisashi Narimatsu (Glycoscience and Glycotechnology Research Group, National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8568, Japan)
- Sriram Neelamegham (Chemical Engineering and Medicine, Sate University of New York, Buffalo, NY 14260, USA)
- Shoko Nishihara (Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan)
- Shujiro Okuda (Graduate School of Medical and Dental Sciences, Niigata University, Niigata, 951-8510, Japan)
- Tamiko Ono (Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan)
- Nicolle H. Packer (Department of Molecular Sciences, Macquarie University, 2109, Sydney and Institute for Glycomics, Griffith University, 4222, Queensland, Australia)
- René Ranzinger (Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, USA)
- Daisuke Shinmachi (SparqLite LLC., Tokyo, 192-0032, Japan)
- Masaaki Shiota (Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan)
- Philip Toukach (Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow 119991, Russia)
- Shinichiro Tsuchiya (Graduate School of Engineering, Soka University, Tokyo 192-8577, Japan)
- Yu Watanabe (Graduate School of Medical and Dental Sciences, Niigata University, Niigata, 951-8510, Japan)
- Akiyasu C. Yoshizawa (Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan)
- Kiyoko F. Aoki-Kinoshita (Faculty of Science and Engineering, Soka University, Tokyo 192-8577, Japan)

Funding

Acknowledgements

Conflict of interest statement

None declared.

References

Campbell MP, Peterson R, Mariethoz J, Gasteiger E, Akune Y, Aoki-Kinoshita KF, Lisacek F, Packer NH. 2014. UniCarbKB: Building a knowledge platform for glycoproteomics. *Nucleic Acids Res*. 42: D215–D221.

CAZypedia Consortium. 2018. Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology*. 28:3–8.

Cheng K, Zhou Y, Neelamegham S. 2017. DrawGlycan-SNFG: A robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology*. 27:200–205.

Kinjo AR, Bekker GJ, Wako H, Endo S, Tsuchiya Y, Sato H, Nishi H, Kinoshita K, Suzuki H, Kawabata T et al. 2018. New tools and functions in data-out activities at Protein Data Bank Japan (PDBj). *Protein Sci*. 27: 95–102.

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. *Mol Cell Proteomics*. 12:991–995.

Liu G, Cheng K, Lo CY, Li J, Qu J, Neelamegham S. 2017. A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. *Mol Cell Proteomics*. 16:2032–2047.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 42:D490–D495.

Okuda S, Watanabe Y, Moriya Y, Kawano S, Yamamoto T, Matsumoto M, Takami T, Kobayashi D, Araki N, Yoshizawa AC et al. 2017.

jPOSTrepo: An international standard data repository for proteomes. *Nucleic Acids Res*. 45:D1107–D1111.

Tiemeyer M, Aoki K, Paulson J, Cummings RD, York WS, Karlsson NG, Lisacek F, Packer NH, Campbell MP, Aoki NP et al. 2017. GlyTouCan: An accessible glycan structure repository. *Glycobiology*. 27:915–919.

Toukach PV, Egorova KS. 2016. Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. *Nucleic Acids Res*. 44: D1229–D1236.

UniProt Consortium T. 2018. UniProt: The universal protein knowledgebase. *Nucleic Acids Res*. 46:2699.

Watkins X, Garcia LJ, Pundir S, Martin MJ, the UniProt Consortium. 2017. ProtVista: Visualization of protein sequence annotations. *Bioinformatics*. 33:2040–2041.

Yasugi E, Seyama Y. 2007. [Lipid database "LipidBank" and international collaboration]. *Tanpakushitsu Kakusan Koso*. 52:1357–1362.

York WS, Agravat S, Aoki-Kinoshita KF, McBride R, Campbell MP, Costello CE, Dell A, Feizi T, Haslam SM, Karlsson N et al. 2014. MIRAGE: The minimum information required for a glycomics experiment. *Glycobiology*. 24:402–406.

Zhao S, Walsh I, Abrahams JL, Royle L, Nguyen-Khuong T, Spencer D, Fernandes DL, Packer NH, Rudd PM, Campbell MP. 2018. GlycoStore: A database of retention properties for glycan analysis. *Bioinformatics*. 10:e0144578.