

糖鎖科学における 研究データ管理

山田一作

やまだ いっさく



公益財団法人野口研究所

第3回 SPARC Japan セミナー2019
2020年2月7日(金) 国立情報学研究所

自己紹介 + α

もともとは、有機化学が専門

野口研究所で糖鎖科学と出会う

→ 糖鎖構造は面白く難しい

糖鎖構造は面白く難しい

化学には元素記号がある

炭素 $\rightarrow$ **C** $\rightarrow$ ^{12}C

糖鎖科学では

グルコース $\rightarrow$ **Glc** $\leftarrow$ これでいい？

Glc の化学構造式は？

Glc の化学構造は？

Glc は、グルコース？

EMBL-EBI ChEBI

glucose
Examples: iron*, InChI=1S/CH4O/c1-2/h2H,1H3, caffeine
Search Advanced

Home Advanced Search Browse Documentation Download Tools About ChEBI Contact us Submit

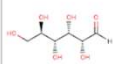
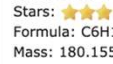
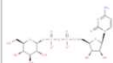
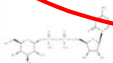
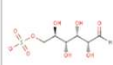
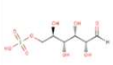
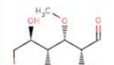
ChEBI > Search Results

Search Results for All in ChEBI
AND glucose in All
Edit Search

Show more data from EMBL-EBI

Download your results

842 entries found, displaying 1 to 15.

<u>aldehydo-D-glucose</u>  CHEBI:42758 Stars: ★★★★★ Formula: C6H12O6 Mass: 180.15588 Charge: 0	<u>glucose</u>  CHEBI:17234 Stars: ★★★★★ Formula: C6H12O6 Mass: 180.15588 Charge: 0 Definition: An aldohexose used as a source of energy and metabolic intermediate.	<u>2-deoxy-D-glucose</u>  CHEBI:15866 Stars: ★★★★★ Formula: C6H12O5 Mass: 164.15650 Charge: 0 Definition: A deoxyglucose that is D-glucose in which the hydroxy group at position 2 has been replaced by a hydrogen. It is an anti...
<u>CDP-α-D-glucose(2-)</u>  CHEBI:137927 Stars: ★★★★★ Formula: C15H23N3O16P2 Mass: 563.302 Charge: -2	<u>CDP-α-D-glucose</u>  CHEBI:28942 Stars: ★★★★★ Formula: C15H25N3O16P2 Mass: 565.318 Charge: 0	<u>D-glucose</u>  CHEBI:17634 Stars: ★★★★★ Formula: C6H12O6 Mass: 180.15588 Charge: 0 Definition: A glucose with D-configuration.
<u>D-glucose 6-sulfate(1-)</u>  CHEBI:57904 Stars: ★★★★★ Formula: C6H11O9S Mass: 259.21100 Charge: -1	<u>D-glucose 6-sulfate</u>  CHEBI:16809 Stars: ★★★★★ Formula: C6H12O9S Mass: 260.21900 Charge: 0	<u>3-O-methyl-D-glucose</u>  CHEBI:73918 Stars: ★★★★★ Formula: C7H14O6 Mass: 194.18250 Charge: 0

“Glc”とは何？

内容

- 糖鎖の機能
- 糖鎖科学における研究データ
- 糖鎖の構造
- 糖鎖科学における標準化
- 糖鎖科学のリポジトリ

糖鎖の機能

- 蛋白質や脂質の機能向上
 - 溶解性、細胞局在、蛋白質品質管理
- 細胞間情報伝達
 - 細胞の認識や接着、病原体の認識、がん増殖・転移などに関与
- 個の識別
 - ABO型の血液型

糖鎖科学における研究データ

データ

糖鎖科学研究の手法の例

- 合成
 - 化学合成及び酵素合成
- 分析
 - 一次構造解析
 - 相互作用
 - 機能
 - イメージング

■

■

■

鍵 → 糖鎖構造

糖鎖の構造

構造

糖鎖科学研究の鍵

- 合成
- 分析



糖鎖構造データ
糖鎖構造分析データ

糖鎖構造のデータ

構造

文字列表記

名称

Blood Group B Type 2

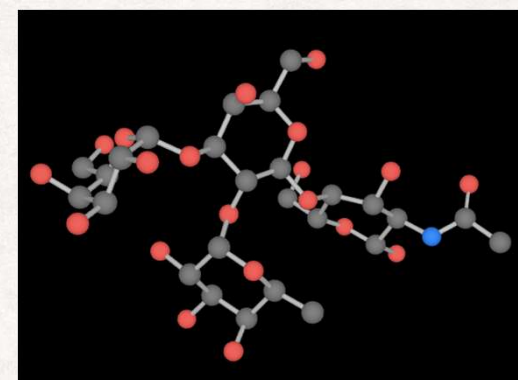
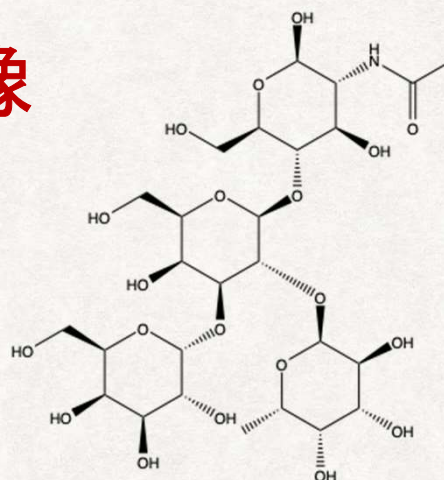
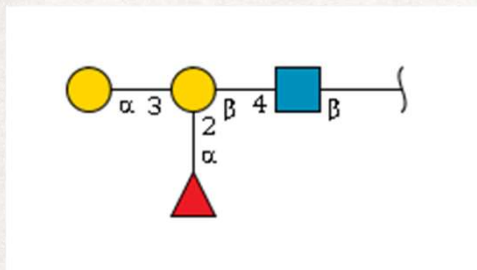
IUPAC

Fuc(a1-2)[Gal(a1-3)]Gal(b1-4)b-GlcNAc

GlycoCT

LINUCS

画像



一次構造

三次元構造

糖鎖科学における標準化

JIS Z 8002（標準化及び関連活動—一般的な用語）における 標準化の定義

「実在の問題又は起こる可能性がある問題に関して、与えられた状況において最適な秩序を得ることを目的として、共通に、かつ、繰り返して使用するための記述事項を確立する活動。」と定義

- 相互理解
- 互換性確保
- 品質の確保
- 正確な情報の伝達

参考：福田泰和, 吉田均, 工学教育(J.of JSEE), 56-2(2008)

糖鎖構造の明確化

糖鎖科学における標準化の取り組み

- 糖鎖構造の記号
- 糖鎖構造の文字列表記
- ガイドライン
- 糖鎖構造へのID付与

糖鎖構造の記号

Symbol Nomenclature for Glycans (SNFG)

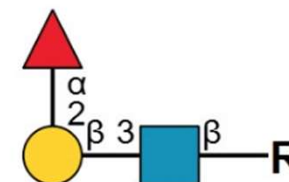
Table 1. Monosaccharide symbol nomenclature

SHAPE	White (Generic)	Blue	Green	Yellow
Filled Circle	 Hexose	 Glc	 Man	 Gal
Filled Square	 HexNAc	 GlcNAc	 ManNAc	 GalNAc
Crossed Square	 Hexosamine	 GlcN	 ManN	 GalN

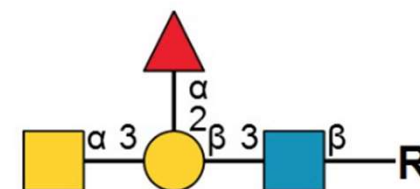
Glycobiology 25:1323-1324, 2015.
Glycobiology 29:620-624, 2019.

Blood group antigens:
H antigen on Type-1 lactosamine chain

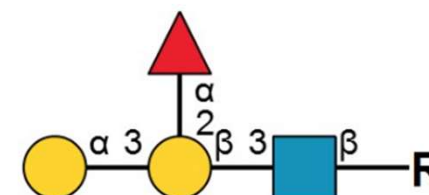
R=glycan backbone



A antigen



B antigen



<https://www.ncbi.nlm.nih.gov/glycans/snfg.html>

糖鎖構造の文字列表記



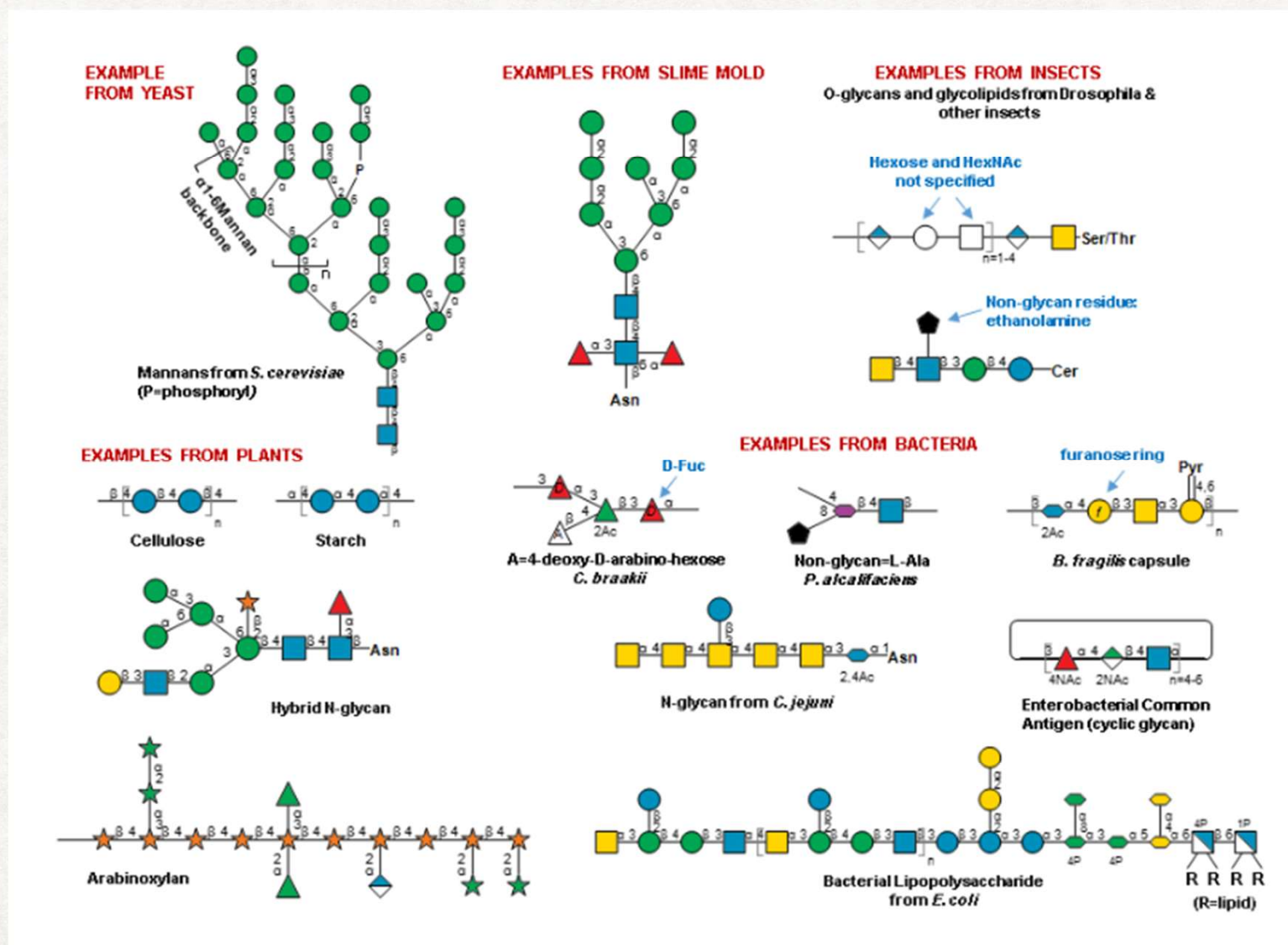
Web3 Unique Representation of Carbohydrate Structures

国際糖鎖構造リポジトリのために開発した表記法

- セマンティックウェブを指向した糖鎖構造表記法
- 糖鎖構造をユニークな文字列へ
- 糖鎖構造の同一性



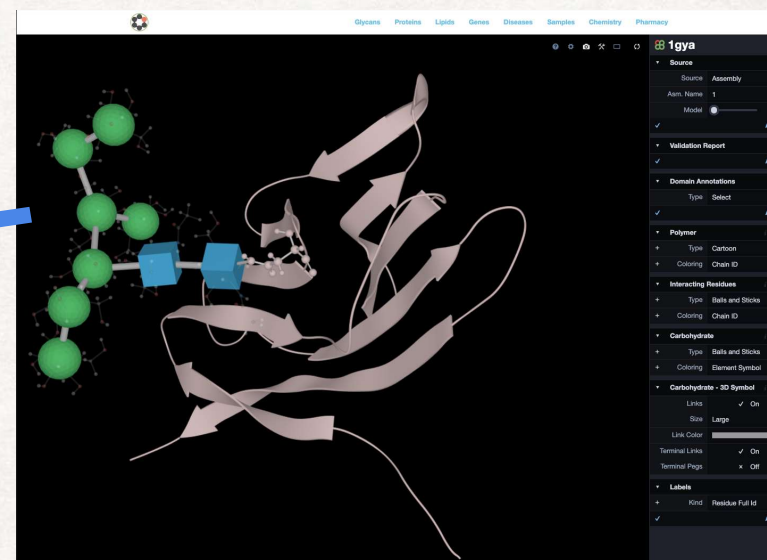
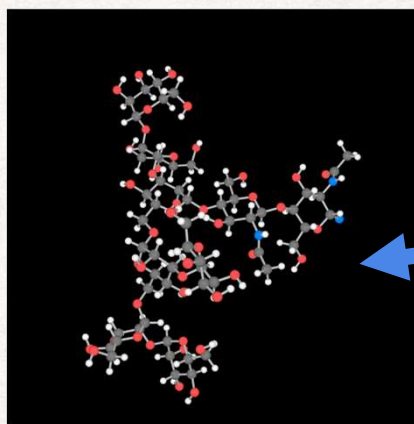
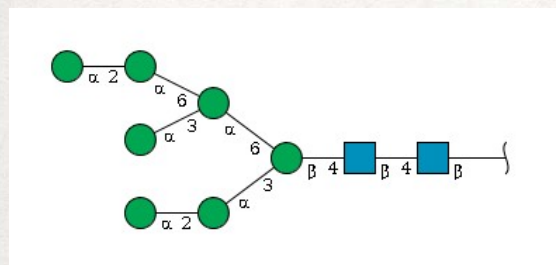
糖鎖構造の種類



WURCS

Web3 Unique Representation of Carbohydrate Structures

PDB: 1GYA
Science, 269:1273-1278, 1995



WURCS=2.0/3,9,8/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5]/1-1-2-3-3-3-3-3-3/a4-b1_b4-c1_c3-d1_c6-f1_d2-e1_f3-g1_f6-h1_h2-i1

ガイドライン

MIRAGEプロジェクト

Minimum Information Required for A Glycomics Experiment

プロジェクトの目的

科学文献のグライコムクスデータの品質を向上

グライコムクス実験を報告する際に報告すべき項目のガイドラインを提唱

- ✓ 実験方法を定義するのではない
- ✓ 実験結果を解釈しやすくし、再現できるように
- ✓ ジャーナルやデータベースもこれらのガイドラインを利用可能

MIRAGEプロジェクト

標準化

PROJECTS > MIRAGE > GUIDELINES

GUIDELINES

OVERVIEW

Guideline	Version	Published
Sample Preparation	1.0; 18 Feb. 2016	<i>Glycobiology</i> , 2016, 26(9):907-910
Mass Spectrometric Analysis	1.0; 24 Apr. 2013	<i>Mol. Cell. Proteomics</i> , 2013, 12:991-995
Glycan Microarray Analysis	1.0; 22 June 2016	<i>Glycobiology</i> , 2017, 27(4):280-284
Liquid Chromatography Analysis	1.0; 1 Mar. 2018	<i>Glycobiology</i> , 2019, online

糖鎖構造へのID付与

第5回ACGG-DB 会議（2013年6月22日 中国・大連）

参加国：日本、米国、豪州、ドイツ、ロシア、中国、韓国、台湾



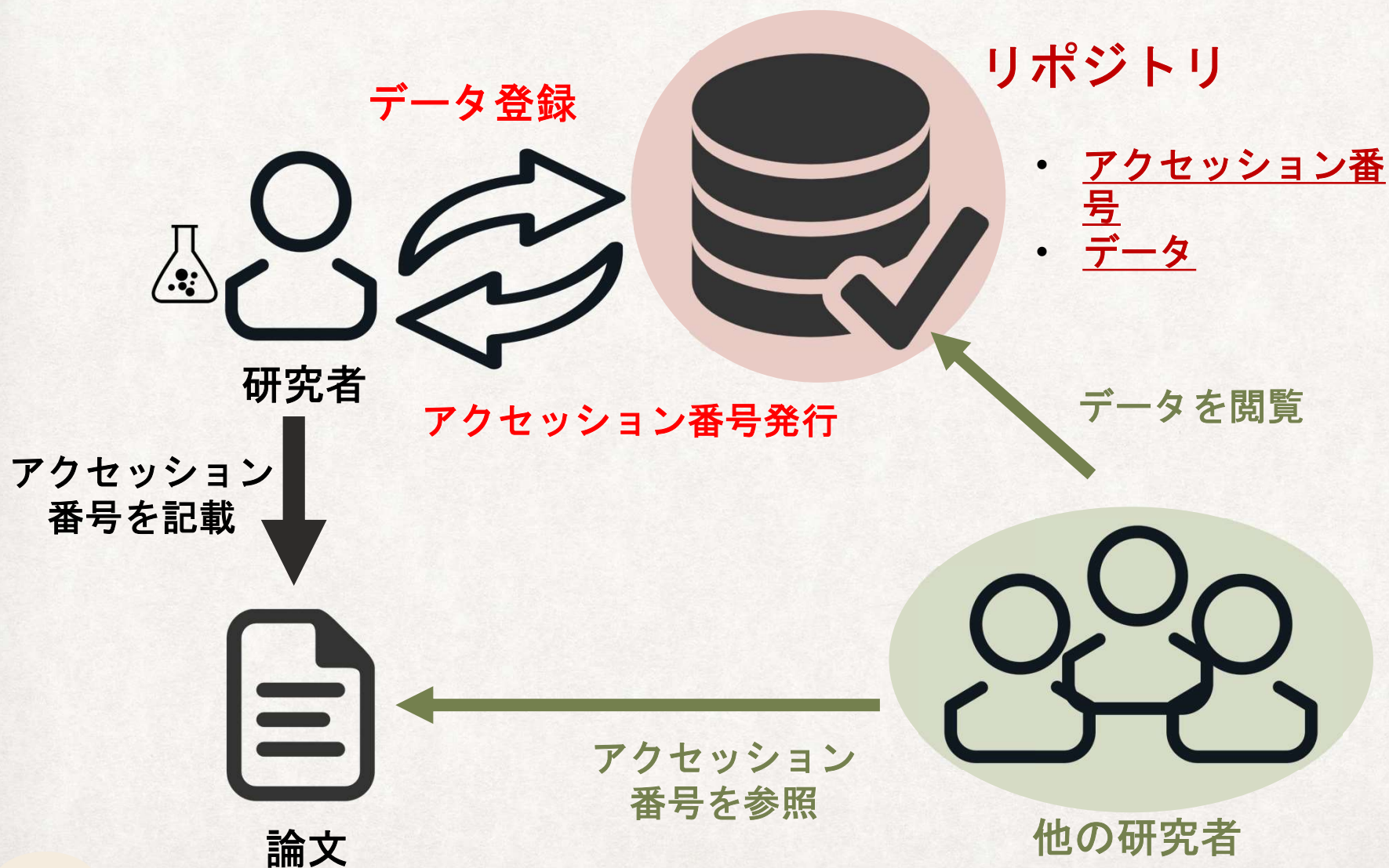
- 国際糖鎖構造リポジトリの基本合意
 - 論文投稿前のアクセッション番号の発行システム
- データの範囲
 - 糖鎖構造情報のみ
 - （メタデータ：登録者と登録日）

リポジトリ



糖鎖科学のリポジトリ

Repo



糖鎖科学のリポジトリ

Repo

標準化

SNFG



MIRAGE
WIBACE

リポジトリ



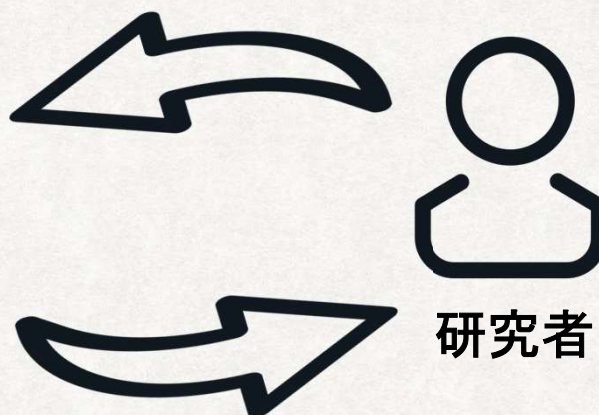
Glyco
POST

GlyTouCan

Repo

国際糖鎖構造リポジトリ

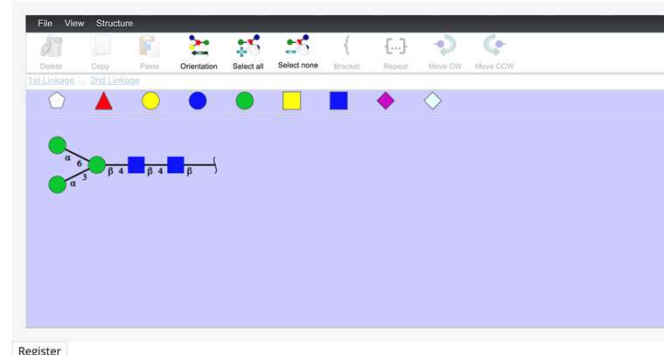
糖鎖構造データ登録



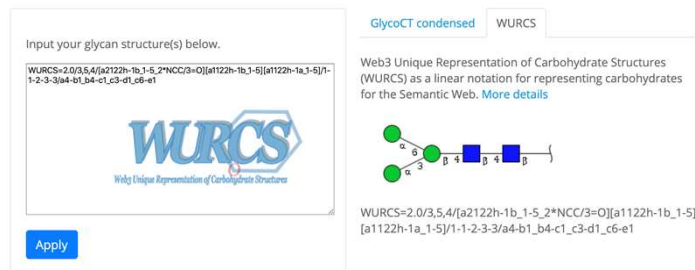
G22768V0

アクセッション番号

Glycan Registration - Graphic Input

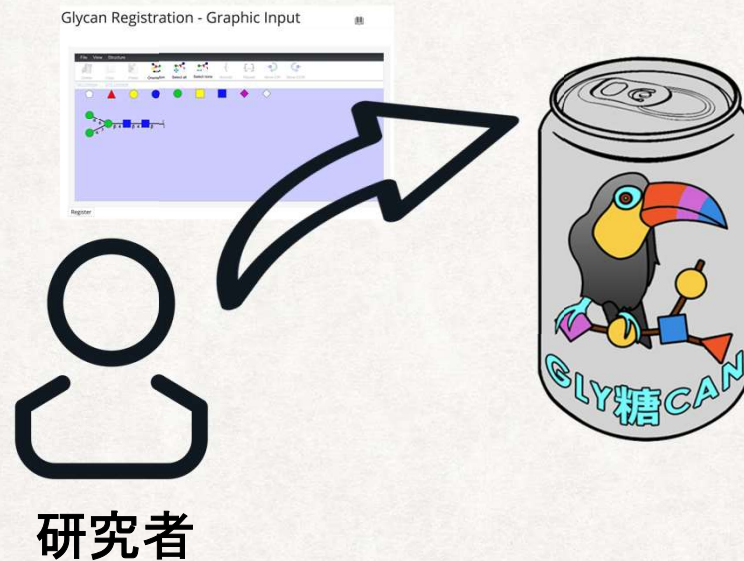


Glycan Registration - Text Input

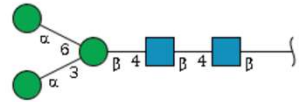


GlyTouCanの中

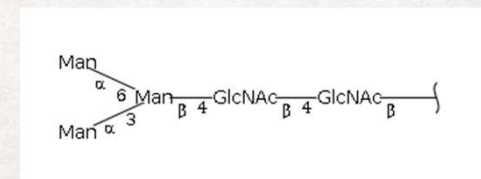
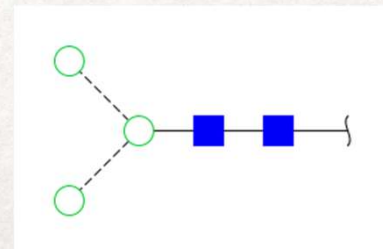
Repo



アクセッション番号

RES 1b:b-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	WURCS=2.0/3,5,4/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5] [a1122h-1a_1-5]/1-1-2-3-3/a4-b1_b4-c1_c3-d1_c6-e1 		G22768VO
SNFG			

登録される糖鎖構造の各種データを生成



UniCarb-DR

Repo

アノテーションした質量分析データのリポジトリ

The screenshot shows the UniCarb-DR website. At the top is a navigation bar with links to UniCarb-DR, MIRAGE, and About, along with Login and Register buttons. The main content area features the UniCarb-DR logo, which consists of three interlocking gears in red, blue, and grey, followed by the text 'UniCarb-DR' and 'The Glycomic MS Database and Repository'. Below this is a red 'Search UniCarb-DR' button. A paragraph of text describes the database's mission and history, mentioning its initiation in 2009 and its growth into one of the largest experimental glycomic MS databases. It also mentions that glycan nomenclature is available in 'Essentials of Glycobiology'. Below the text is a blue bar with the heading 'Search categories by following a link below' and five categories: Taxonomies (16), Tissues (26), References (24), Structures (1118), and Spectra (1588). Each category is represented by a small icon and a number.

UniCarb-DR | MIRAGE | About

Login Register

 **UniCarb-DR**
The Glycomic MS Database and Repository

[Search UniCarb-DR](#)

Carbohydrates are the most common but less known biological material on earth. This deficit can partly be explained by the lack of glycodatabases and glycobioinformatics. UniCarb-DR is an example of what can be achieved with a lot of dedication, cross-disciplinary collaborations and lack of monetary resources. In 2009 the UniCarb-DB was initiated in order to meet in-house need to store structural and MS-glycomic data, and has now grown to be one of the largest experimental glycomic MS databases, generously supported by several national and European funding for continued growth and development.

Glycans nomenclature is available in [Essentials of Glycobiology](#).

Let's explore the glycans!

Search categories by following a link below

Category	Count
Taxonomies	16
Tissues	26
References	24
Structures	1118
Spectra	1588

Excel file generation.

The forms below mirror the structure of the excel file. The number of tabs is dependent on the options selected in the previous step, and all the data captured in the forms will be transcribed into the excel file. Please also click on the "Generate" button without filling in the forms. This action will generate an empty excel file. However, it is recommended to use the interfaces since some of the fields have a defined language, which prevents validation errors later during the submission process.

The fields marked with an * are required. However, if these data is not relevant for your experiment, you can use the value NA (Not applicable).

Generate

Sample Preparation

Sample Preparation

LC settings

MS part 1

LC-MS Settings

For uploading of data to UniCarb-DB

General features — (a) Global descriptors

Date stamp

Responsible person: Provide name

Affiliation

Stable contact information

HP: Brand

Injector

Injector settings

Instrument manufacturer

Instrument model

Customizations

Ion mode

Sample Preparation

LC settings

MS part 1

Sample Preparation

LC settings

MS part 1

MS part 2

MIRAGE MS Guidelines

Guidelines for reporting mass spectrometric analysis data

General features — (a) Global descriptors

Date stamp

Responsible person: Provide name

Affiliation

Stable contact information

Instrument manufacturer

Instrument model

Customizations

Ion mode

Positive

4. Spectrum and peak list generation and annotation

For this section, if software other than that listed in 1b (Control and analysis software) is used to perform a task, it must be recorded in the case of a task.

Location of source ("raw") files

Software Name Format Link to the target area URL

Peak list generation and annotation — (b) Peak list generation

Since several different applications may be used for the data acquisition, data post processing and spectrum annotation each used software should be recorded separately together with the information what modification has been done to the data.

Software Name Version Customizations Software settings

MIRAGEガイドラインを利用してサンプル情報、LC設定、MS設定、スペクトルデータなどアノテーション済みのデータを登録

GlycoPOST

グライコプロテオミクス関連実験から得られた 質量分析データのためのリポジトリ


GlycoPOST Beta

 Data list
  Submit
  Mypage
  Help

[Login](#)
[Sign up](#)

About GlycoPOST

GlycoPOST is a mass spectrometry data repository for glycomics and glycoproteomics. It consists of a high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to our repository server is very fast and uses only web browsers – it does not require installing any additional software.

Submission conditions are in accordance with the [Minimum Information Required for a Glycomics Experiment \(MIRAGE\)](#) guidelines.



Statistics

35 projects are registered. **8** s are opened.

1144 files amount to **310.9 GB**.

Data list

ID	Project title	Description	Publication	Principal investigator	Announcement date	Detail
GPST000026	SEC-IS-MS of low molecular weight heparins	Low molecular weight heparins (LMWH) prepared by partial depolymerization of unfractionated heparin are used globally to treat coagulation disorders on an outpatient basis. Patent protection for sever...	27709905	Joseph Zaia	2020/01/25	Detail page
GPST000038	UniCarb-DR	Data from UniCarb-DR	31332			
GPST000027	GnT3 KO mouse brain	Bisecting GlcNAc is a general suppressor of terminal modification of N-glycan	31375533	Miyako Nakano	2019/09/21	Detail page




GycoPOST

Repo

実験から得られた**生データ**
(ファイル) と、実験条件
などの**メタデータ**

MIRAGE
WIBACE

GlycoPOST Beta

[Data list](#) [Submit](#) [My page](#) [Help](#) [Login](#) [Sign up](#)

About GlycoPOST

GlycoPOST is a mass spectrometry data repository for glycomics and glycobiology. It consists of a high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to our repository server is very fast and uses only web browsers - it does not require installing any additional software.

Submission conditions are in accordance with the Minimum Information Required for a Glycomics Experiment (MIRAGE) guidelines.

Statistics

33 projects are registered. 7 are opened.
1143 files amount to 310.9 GB.

**MIRAGE
WIBACE**

Data list

ID	Project title	Description	Publication	Principal investigator	Announcement date	Detail
GPST000038	UniCarb-DR	Data from UniCarb-DR	31332201	Nicolas G. Karlsson	2019/10/02	Detail page
GPST000027	GnT3 KO mouse brain	Bisecting GlcNAc is a general suppressor of terminal modification of N-glycan	31375533	Myako Nakano	2019/09/21	Detail page
GPST000031	GnT3 KO mouse kidney	Bisecting GlcNAc is a general suppressor of terminal modification of N-glycan	31375533	Myako Nakano	2019/09/21	Detail page
GPST000030	Reference glycan structure libraries of primary human cardiomyocytes and pluripotent stem cell-derived cardiomyocytes reveal cell-type and culture stage-specific glycan phenotypes	To generate glycan structure libraries, the current study used porous graphitized carbon (PGC) LC interfaced with MS (PGC-LC-MS), a technique that resolves glycan structures and enables characterization...		Rebekah L. Gundry	2019/08/31	Detail page
GPST000029	Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping	Porous graphitized carbon (PGC) based chromatography achieves high-resolution separation of glycan structures released from glycoproteins. However, the implementation of PGC-based separations in glyco...	31065629	Nicole H. Packer	2019/06/28	Detail page
GPST000024	Discrimination of isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS	We used porous graphitized carbon-LC-ESI-MS/MS to separate and detect released N- and O-glycan isomers from mammalian model glycoproteins using negative mode resonance activation CID-MS/MS. By intern...	29603058	Nicole H. Packer	2019/06/28	Detail page
GPST000009	Atlantic Salmon mucus from skin, pyloric caeca and distal intestine	Diseases cause ethical concerns and economic losses in the Salmonid industry. The mucus layer comprised of highly O-glycosylated mucins is the first contact between pathogens and fish. Mucin glycans g...	30923042	Sara Linden	2019/04/10	Detail page

1 - 7 / 7

[About](#) [Contact](#) [Privacy policy](#) [Terms of use](#)

© 2018 GlycoPOST project

GPST000026	
Title	SEC-MS of low molecular weight heparins
Keywords	heparin, low molecular weight heparin, glycosaminoglycan, LC-MS
Description	Low molecular weight heparins (LMWH) prepared by partial depolymerization of unfractionated heparin are used globally to treat coagulation disorders on an outpatient basis. Patent protection for several LMWH has expired and abbreviated new drug applications have been approved by the Food and Drug Administration. As a result, reverse engineering of LMWH for biosimilar LMWH has become an active global endeavor. Traditionally, the molecular weight distributions of LMWH preparations have been determined using size exclusion chromatography (SEC) with optical detection. Recent advances in liquid chromatography-mass spectrometry methods have enabled exact mass measurements of heparin saccharides roughly up to degree-of-polymerization 20, leaving the high molecular weight half of the LMWH preparation unassigned. We demonstrate a new LC-MS system capable of determining the exact masses of complete LMWH preparations, up to dp30. This system employed an ion suppressor cell to desalt the chromatographic effluent on-line prior to the electrospray mass spectrometry source. We expect this new capability will impact the ability to define LMWH mixtures favorably.
PubMed ID	27709905
Principal investigator	Joseph Zaia
Submitter	Joseph Zaia
Revision	0
Note	J. Zaia, K. Khatri, J.A. Klein, C. Shao, Y. Sheng, and R. Viner, Complete Molecular Weight Profiling of Low Molecular Weight Heparins Using Size Exclusion Chromatography-Ion Suppressor-High Resolution Mass Spectrometry; Anal Chem; 88; (2016) 10654-10660.
Created date	2019/05/24
Announcement date	2020/01/25

File list

File name	Experimental preset	File size	Detail
dalte_30nmoles_1_HCDOffset_-2.RAW		898.5 MB	Detail
enoxa_30nmoles_2_HCDoffsets_eL_2.RAW		904.6 MB	Detail
enoxa_30nmoles_1_HCDoffsets_eL_2.RAW		900.4 MB	Detail
dalte_30nmoles_2_HCDOffset_-2.RAW		837.9 MB	Detail

リポジトリの連携

ガイドラインに基づいて作成されているため、データの参照が容易

MIRAGE
WIBACE

MSガイドライン

メタデータのインポート、
エクスポートが可能

メタデータの互換性



実験の生データ



アノテーション済みのデータ
(糖鎖構造の情報)

リポジトリの連携



糖鎖構造のアクセス番号



アノテーション済みのデータ
(糖鎖構造の情報)



UniCarb-DR から UniCarb-DB に登録する際に
GlyTouCanのアクセス番号を追加

リポジトリの利用

Repo

GlySpace Alliance



GlyGen



GlyTouCanに登録された糖鎖構造情報は、国内外のデータベースで利用

ACGG-DB

PDBj
Protein Data Bank Japan

GlycoEpitope



GlycoNAVI
Support System for Carbohydrate Research



MatrixDB



Carbohydrate Structure Database



UniCarbKB



UniCarb-DB
The Glycomic MS Database and Repository

PubChem

まとめ

糖鎖科学における研究データの管理のために、

◆ 標準化

- ◆ 糖鎖構造の明確化のためSNFG記号、WURCS（文字列表記）、アクセス番号を標準
- ◆ データの品質向上のためにMIRAGEガイドラインを提唱

◆ リポジトリ

- ◆ 標準を活用してGlyTouCan、UniCarb-DR、GlycoPOSTの各リポジトリを運用、連携

糖鎖科学研究者は、これらのリポジトリと国内外データベースを利用することで、正確な研究データを利用可能