



創価大学

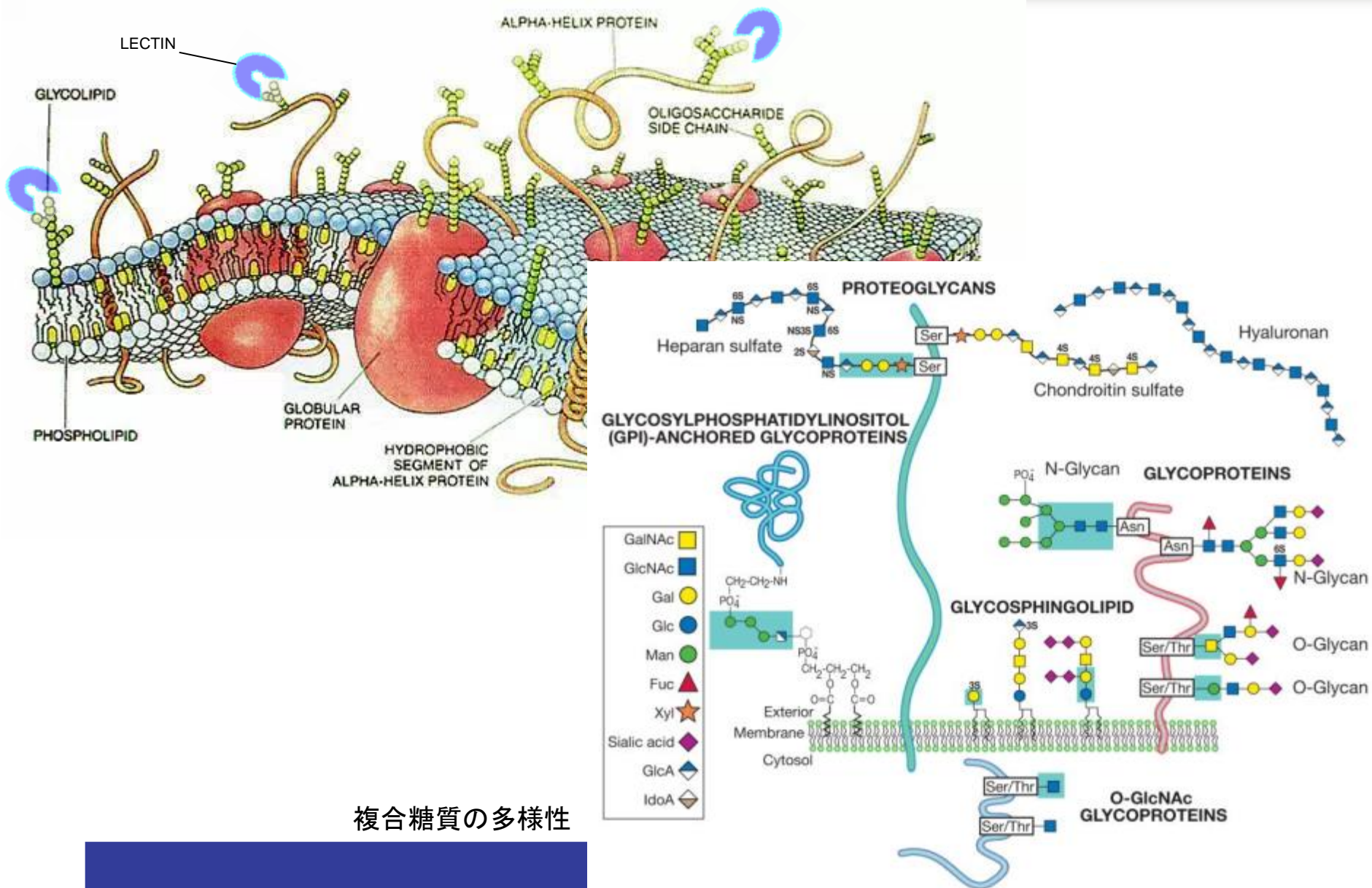
Discover your potential  
自分力の発見

平成29年度統合化推進プログラム

# 「糖鎖科学ポータル構築」

代表 創価大学 木下聖子

# 糖鎖が関わる生体内の環境



複合糖質の多様性

# 糖鎖研究の位置づけ

米国では 2012 年に米国学術研究会議(the National Research Council of the National Academies)が報告書 Transforming Glycoscience: A Roadmap for the Future (2012) を発表 :

糖鎖研究の対象は**医学 (Medicine)**にとどまらず、**エネルギー (Energy)** や**材料科学 (Materials)** にまで及ぶとしており、これまでこのような大局的な視点で糖鎖研究の将来が位置づけられたことはなかった。

米国ではNIHにおいて、糖鎖に関連する研究が強力に進められている。

一方、世界で使用されている標準的な糖鎖解析法は日本発であり、また**重要な糖鎖構造の解析および6割以上の糖鎖遺伝子のクローニングやその機能解析が我が国の研究者によってなされている**。したがって、糖鎖科学は我が国が世界をリードしてきた分野である。

しかし、これらの様々な糖鎖遺伝子などについての情報が未だに網羅的に整理されておらず、文献調査以外には糖鎖が関与するパスウェイ情報などが手に入れられない状況である。このような問題を解決するためにGlyCosmos Portalを開発し、これまで糖鎖研究をリードしてきた我が国が中核となって糖鎖科学ポータルを提供する。 3

# 第5回ACGG-DB 会議 決議事項

大連

2013年6月22日

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)

- システムのフレームワーク

プロトタイプは日本の現在のグループで開発

- 支援と普及促進

糖鎖関連の学会： IGO, ICO, Society for Glycobiology, JSCR, JCGG, HUPO 等



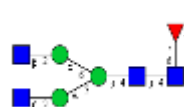
# Glycan Databases



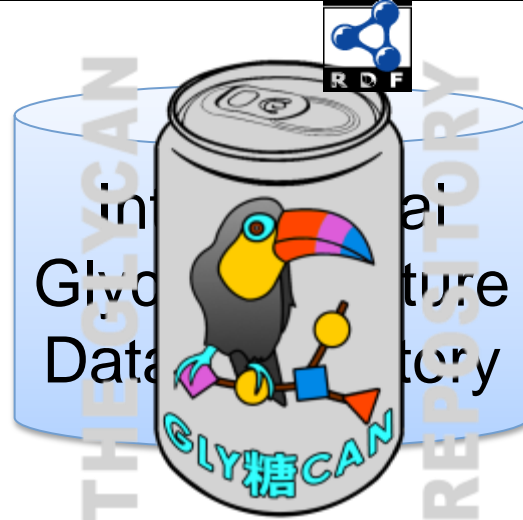
Unique Glycan ID

+

Structure Data



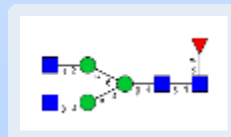
Sharing



② Obtain Unique Glycan ID

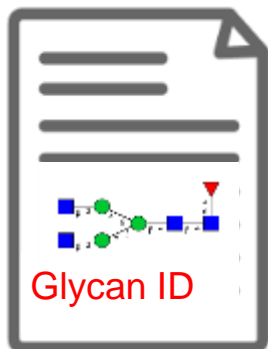
① Structure Registration

## Knowledge-Sharing of Glycan Knowledge via Internet



④ Curation

Annotation Data



Research Paper

③ Manuscript submission

MIRAGE  
WIBACE



Glycan Research

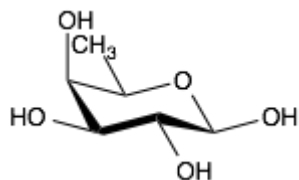
# GlyTouCan開発にあたって

- 糖鎖構造の表現
  - 曖昧性に対応する
  - 任意な文字列で表す
- リポジトリのシステム開発
  - アクセッション番号のシステム
  - ユーザー登録システム
  - 上記糖鎖構造表記システムの導入
- リポジトリの普及

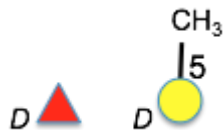
# 糖鎖情報の標準化の必要性

- 単糖表現の曖昧性

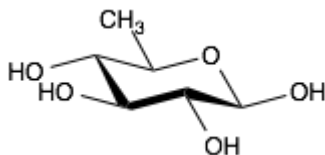
D-Fucose  
6-Deoxy-D-galactose



$\beta$ -D-Fucp



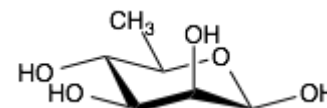
D-Quinovose  
6-Deoxy-D-glucose



$\beta$ -D-Quip



D-Rhamnose  
6-Deoxy-D-mannose



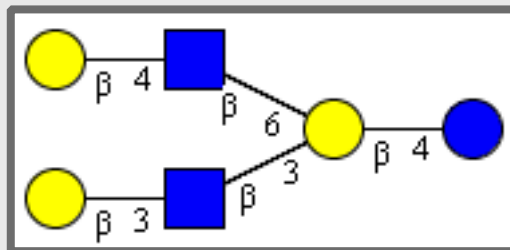
$\beta$ -D-Rhap



- 微生物や植物の糖類。 。 。

# Text formats for glycans

```
ENTRY Structure Glycan
NODE 6
  1 glcnac -16 2
  2 glcnac -16 -2
  3 gal -24 2
  4 gal -8 0
  5 gal -24 -2
  6 D-glc 0 0
EDGE 5
  1 4:b1 6:4
  2 2:b1 4:3
  3 5:b1 2:3
  4 1:b1 4:6
  5 3:b1 1:4
///
```



```
ENTRY Structure Glycan
NODE 6
  1 GlcNAc -16 -2
  2 Gal -8 0
  3 GlcNAc -16 2
  4 Gal -24 -2
  5 D-Glc 0 0
  6 Gal -24 2
EDGE 5
  1 2:b1 5:4
  2 1:b1 2:3
  3 4:b1 1:3
  4 3:b1 6:6
  5 3:b1 2:4
///
```

Uniquely represent all glycan structures as a text string

**W**eb3 **U**nique **R**epresentation of **C**arbohydrate **S**tructures\*

\*Matsubara M, Aoki-Kinoshita KF, Aoki NP, Yamada I, Narimatsu H. WURCS 2.0 update to encapsulate ambiguous carbohydrate structures. *J. Chem. Inf. Model*, 57(4):632-637, 2017.



# 糖鎖構造表記法 WURCS

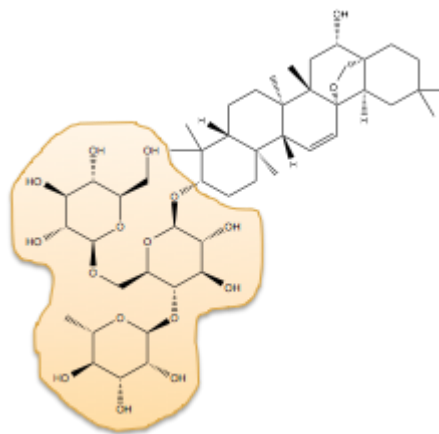
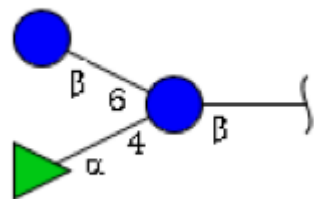
## WURCS: Web3 Unique Representation of Carbohydrate Structures

- セマンティックウェブを指向した新しい糖鎖構造表記法
- JST・NBDC統合化推進プログラムで開発
- GlyTouCanの基盤として利用
  - 糖鎖構造の同一性
  - 糖鎖構造検索
  - 分子量計算

- Version 1.0 published in 2014  
*J. Chem. Inf. Model.* 2014, 54, 1558–1566.
- New version 2.0 published in 2017  
*J. Chem. Inf. Model.* 57(4):632-637, 2017.

WURCS=1.0/3,2/[12122h|1,5][12211m|1,5][12122h|1,5]1+1,3+6|2+1,3+4

G08841ZJ



Version 1.0

URI



URI

RDF&SPARQL

Version 2.0

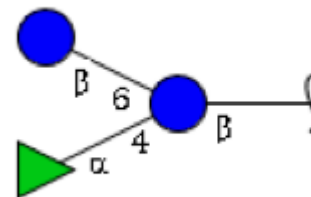
WURCS=2.0/2,3,2/[a2122h-1b\_1-5][a2211m-1a\_1-5]/1-2-1/a4-b1\_a6-c1

WURCS=<Version>/<Counts>/<UniqueRES>/<RES Seq>/<LIN>

# WURCS-RDF

- WURCSで表記した糖鎖構造から、糖鎖のRDFグラフ構造を生成することが可能
- SPARQL検索だけで部分構造検索に利用することができる

G08841ZJ

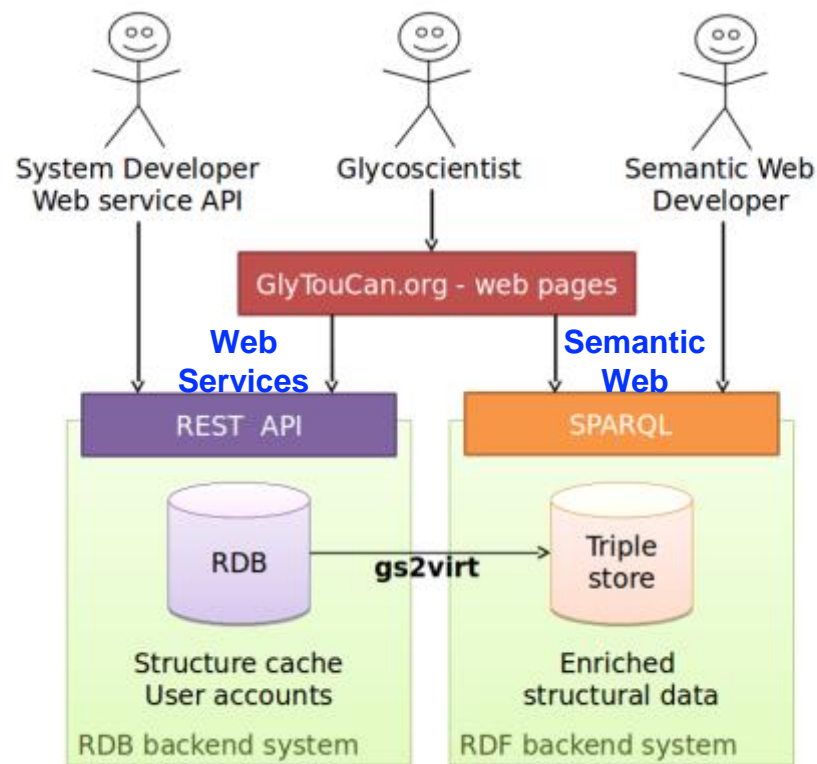


WURCS=2.0/2,3,2/[a2122h-1b\_1-5][a2211m-1a\_1-5]/1-2-1/a4-b1\_a6-c1



# GlyTouCan 1.0

<http://www.glytoucan.org>



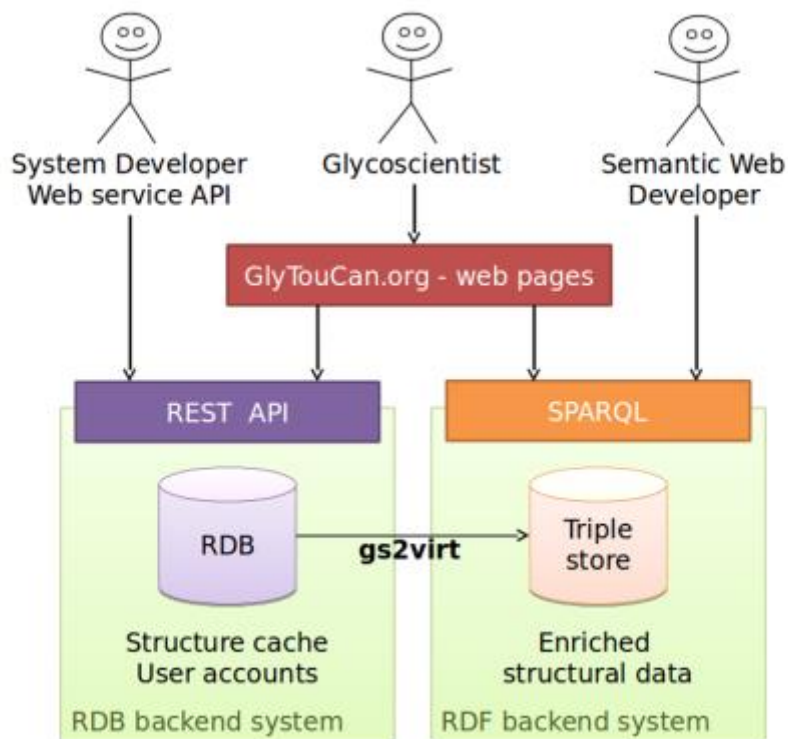
Why Toucan?

TOU = 糖 = “Sugar” in Japanese  
 GlyTouCan is a CAN of GlyCANs

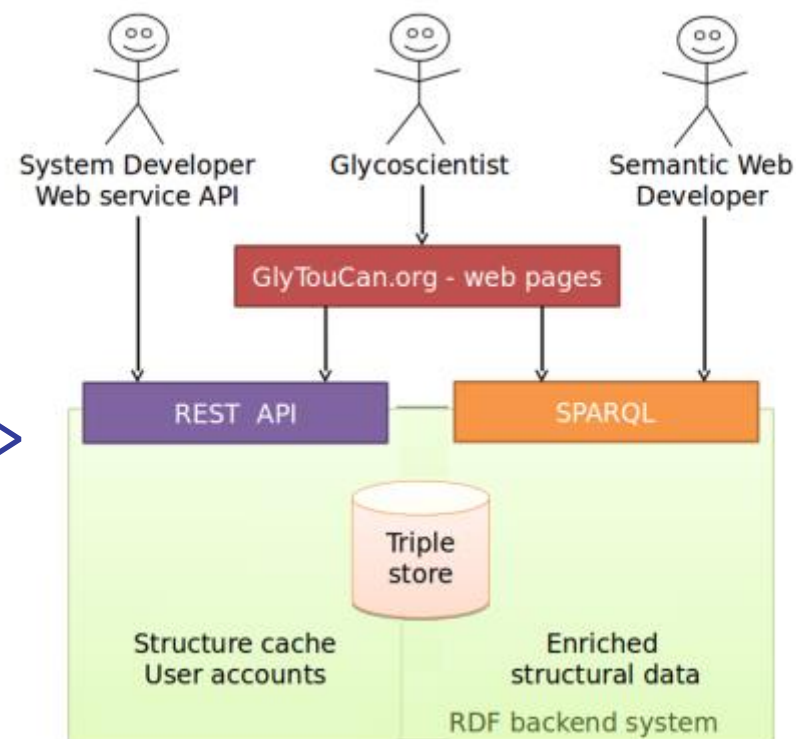
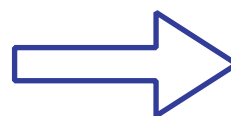
Aoki-Kinoshita, et al. *Nucleic Acids Research, Database Issue*, 2016

# GlyTouCan version 2 (Toco)

## Version 1.0

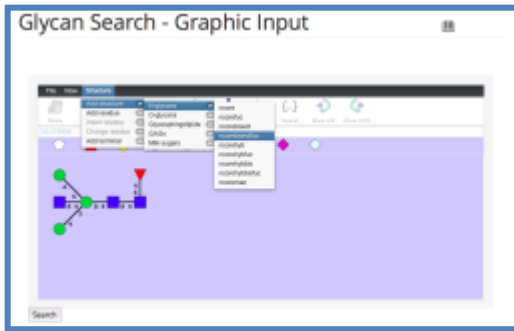


## Version 2.0

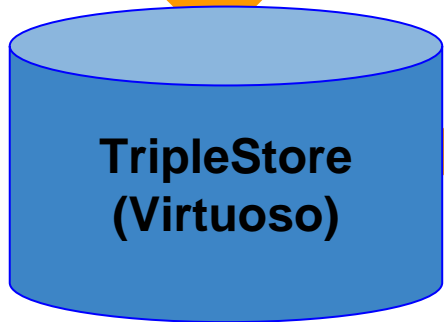


- Aoki-Kinoshita et al, *Nucleic Acids Research, Database Issue*, 2016
- Tiemeyer, Aoki-Kinoshita et al, *Glycobiology*, 2017 (white paper)

## 糖鎖構造の登録

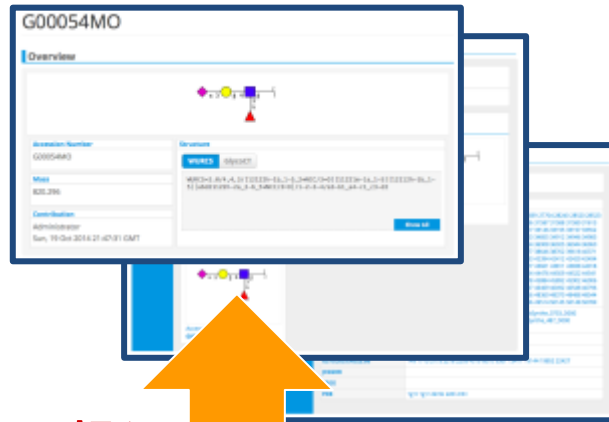


ロード



SPARQ

## 糖鎖構造のエントリー



可視化



可視化



## 糖鎖構造のリスト





# Browsing the repository (View All)

- Currently contains over 80,000 glycan entries and 800 monosaccharides

The screenshot shows the top navigation bar with links for Registration, Search, View All, and Preferences. A search bar is on the right with a 'Sign in' button and an 'Accession Number' field. The main banner features a blue background with a glycan structure diagram and the text 'GlyTouCan THE GLYCAN REPOSITORY'. Below the banner, three statistics are displayed: 82071 Glycans, 61 Motifs, and 800 Monosaccharides.

A search bar with the text 'Google Custom Search' and a 'Search' button.

The 'Glytoucan Schedule' widget includes a title bar, navigation buttons for 'Today', 'Week', 'Month', and 'Agenda', and a main content area with links for 'Showing events after 8/8. Look for earlier events' and 'Showing events until 9/15. Look for more'. It also features a 'Google Calendar' integration button and a note about the time zone: 'Events shown in time zone: Tokyo'.



## What is GlyTouCan?

GlyTouCan is the international glycan structure repository. This repository is a freely available, uncurated registry for glycan structures that assigns globally unique accession numbers to any glycan independent of the level of information provided by the experimental method used to identify the structure(s). Any glycan structure, ranging in resolution from monosaccharide composition to fully defined structures can be registered as long as there are no inconsistencies in the structure.

## Tweets by @glytoucan

A tweet from @glytoucan stating: 'GlyTouCan has been updated with all links from GlycomeDB!' dated Aug 4, 2017.





# Graphical Search for Glycans



Search

View All

Preferences

Sign In

Sign Up


Accession Number





## Glycan Search - Graphic Input





File View Structure


  
Delete

  
Copy


  
Paste


  
Orientation

  
Select all

  
Select none









  
Bracket

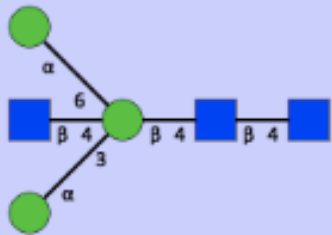
  
Repeat

  
Move CW

  
Move CCW

1st Linkage  2nd Linkage

Search

17

# Graphical Search Results

Search View All Preferences Sign In Sign Up

## Substructure search results

### Input Query

Accession Number	Image
G35619UY	

### 328 Number of Glycans

1 2 ... 17 >

Accession Number ▲ ▼	Image
G99954ZJ	


[Registration](#)
[Search](#)
[View All](#)
[Preferences](#)
[Profile](#)



## Glycan Registration - Text Input



Input your glycan structure(s) below.

Sequence

GlycoCT and WURCS  
are supported

Conversion utilities from  
others, such as IUPAC,  
are currently being  
developed

Add Another

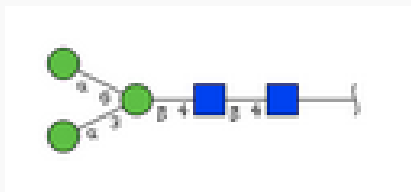
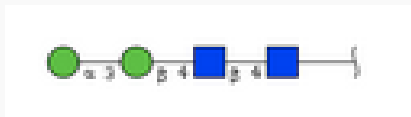
Example of GlycoCT condensed format: The glycan containing repeating units in GlycoCT format.

```

RES
1r:r1
REP
REP1:8o(4+1)2d=-1--1
RES
2b:b-dgal-HEX-1:5
3s:n-acetyl
4b:b-dglc-HEX-1:5 | 6:a
5b:b-dgal-HEX-1:5
6s:n-acetyl
7b:a-dgal-HEX-1:5
8b:b-dglc-HEX-1:5
LIN
1:2d(2+1)3n
2:2o(3+1)4d
3:4o(4+1)5d
4:5d(2+1)6n
5:5o(4+1)7d
6:7o(3+1)8d
    
```

## Registration Confirmation

The following structure(s) have already been registered.

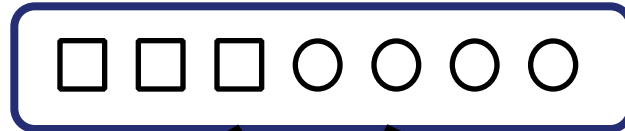
	original Structure	Structure	Image	ID
No.1	ENTRY test1 Glycan NODE 5 8 1 1 GlcNAc 0 1 2 GlcNAc -9 1 3 Man	RES 1b:x-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:a-dman-HEX-1:5 LIN 1:1d(2+1)2n		G20624LQ
No.2	ENTRY test2 Glycan NODE 4 8 1 1 GlcNAc 0 1 2 GlcNAc -9 1 3 Man 4 Man	RES 1b:x-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 LIN 1:1d(2+1)2n 2:1alpha(4+1)3d		G37216UM

Download this data.

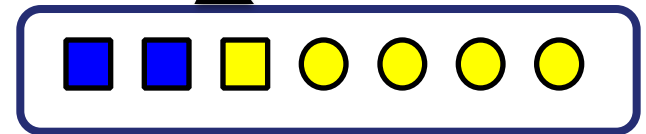
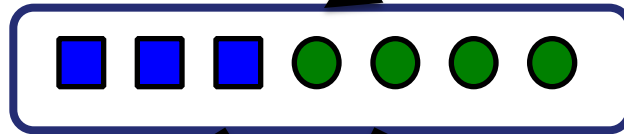
[Download](#)

# Version 2.0: Subsumption

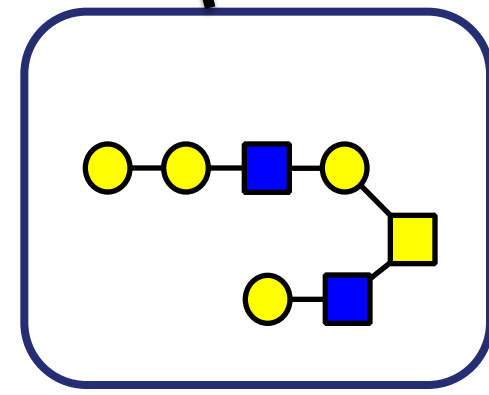
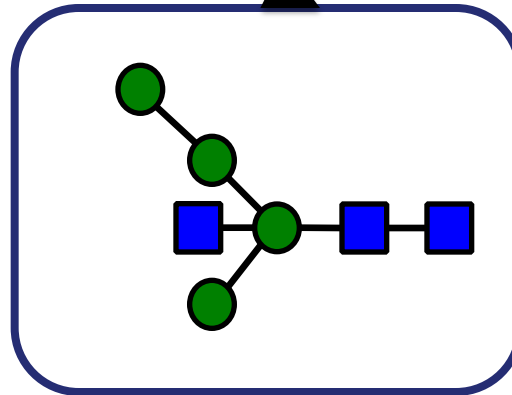
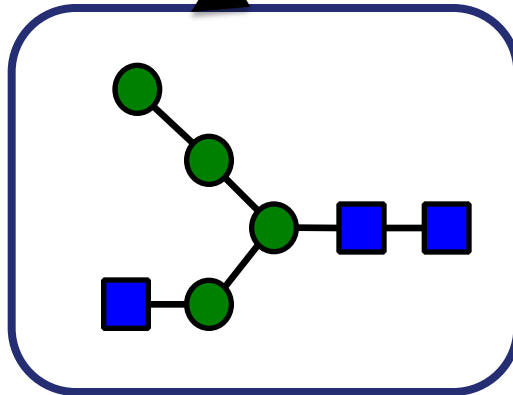
Base Composition



Monosaccharide composition



Glycosidic Topology



Even more defined structures    • • •    • • •    • • • •    • • •    • • •

# GRAB tool

- Glycan RDF association browser

**GRAB Graph**  
G68883LA

GRAB Tree  
Return

GRAB allows users to easily understand the relationships of glycan structures in an informational graphic.  
**GRAB Tree** displays multiple relationships in tree format. Relationships such as motifs, superstructures, substructures, and topology will be displayed as they become available.  
**GRAB Graph** displays the relationships in four directions: superstructures (Above), substructures (Below), subsumes (Left), and subsumed by (Right) with the selected glycan at the center.

**Mouse over function:** User can easily check the selected glycan's structure and accession number.  
**Right-click function:** A menu will display where Users can select to display the glycan's entry page, to show a new GRAB page with the selected glycan, copy the glycan accession number, etc.

The interface shows a central glycan structure with four directional arrows:
 

- Superstructure** (Up): A green box containing a larger glycan structure.
- Substructure** (Down): A purple box containing a smaller glycan structure.
- Subsumes** (Left): A red box containing a glycan structure.
- Subsumed by** (Right): A blue box containing a grid of glycan structures, with one structure highlighted and a tooltip menu open.

Tooltip menu options:

- Copy this accession number
- Center on this glycan
- Open this glycan's entry page
- Show this glycan image

Page navigation: Registration, Search, View All, Preferences, Sign in, Accession Number search.

Page footer: Share (Facebook, Twitter, Google+), Terms and Conditions, Site Policy, FAQ, User Guide, Contact Us, Copyright © 2016 Glytoucan.org v1.2.1

# 糖鎖関連データベースの国際的な動き

- NIH Common Fund
- MIRAGE: minimum information required for reporting a glycomics experiment
- GLIC: glycoinformatics consortium

## RFA-RM-16-021はGlyTouCanが連携する計画である

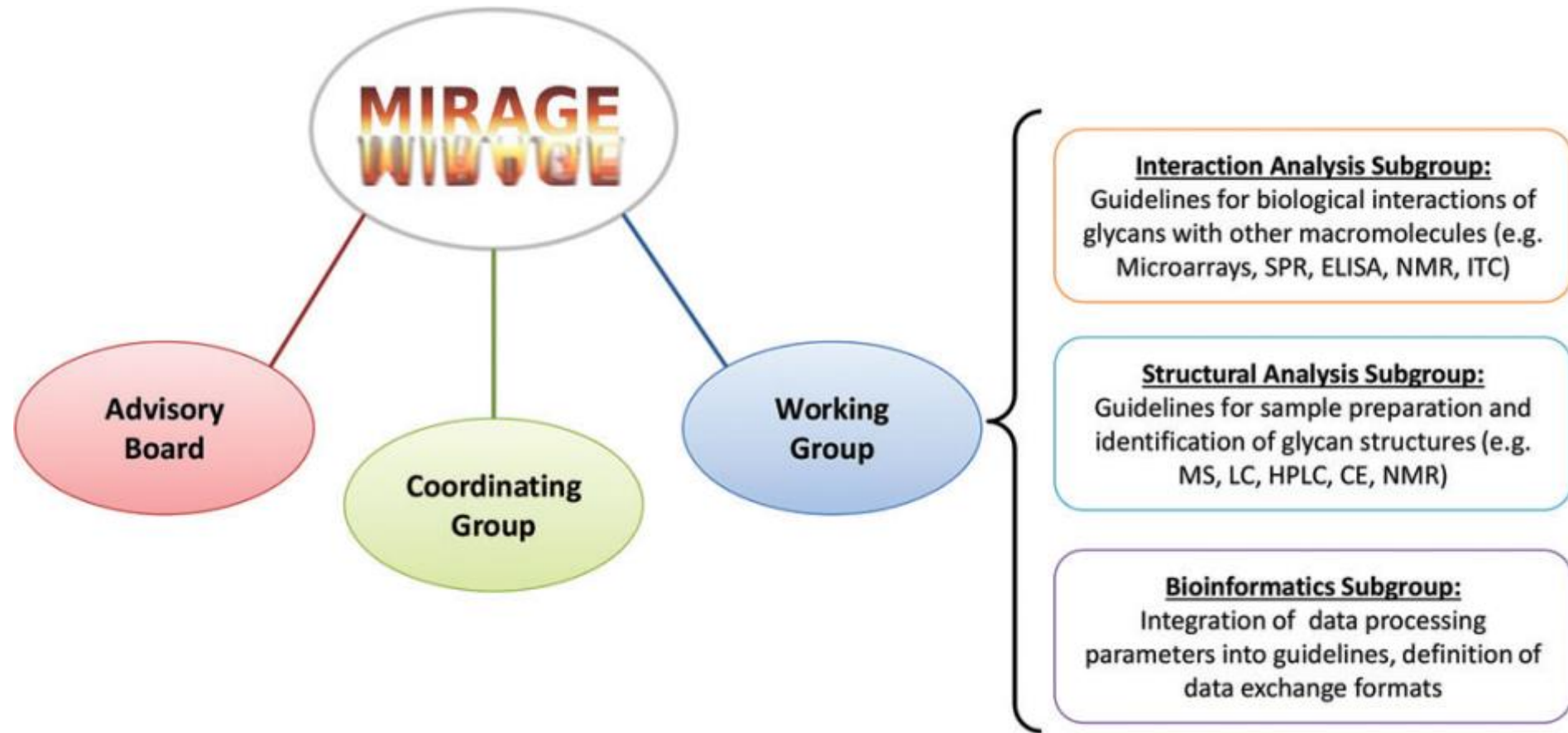
- [Facile Methods and Technologies for Synthesis of Biomedically Relevant Carbohydrates \(U01\)](#)  
[RFA-RM-16-020](#)
- [Data Integration and Analysis Tools: Accessible Resources for Integration and Analysis of Carbohydrate and Glycoconjugate Data in the Context of Comparable Gene, Protein, and Lipid Data \(U01\)](#)  
[RFA-RM-16-021](#)
- [Novel and Innovative Tools to Facilitate Identification, Tracking, Manipulation, and Analysis of Glycans and their Functions \(U01\)](#)  
[RFA-RN-16-021](#)
- [Innovative Adaptations to Simplify Existing Technologies for Manipulation and Analysis of Glycans \(U01\)](#)  
[RFA-RM-16-023](#)
- [Facile Methods and Technologies for Synthesis of Biomedically Relevant Carbohydrates \(U01\)](#)  
[RFA-RM-15-007](#)
- [Novel and Innovative Tools to Facilitate Identification, Tracking, Manipulation, and Analysis of Glycans and their Functions \(R21\)](#)  
[RFA-RM-15-008](#)
- [Novel and Innovative Tools to Facilitate Identification, Tracking, Manipulation, and Analysis of Glycans and their Functions \(U01\)](#)  
[RFA-RM-15-009](#)
- [Data Integration and Analysis Tools: Accessible Resources for Integration and Analysis of Carbohydrate and Glycoconjugate Structural, Analytical, and Interaction Data in the Context of Comparable Gene, Protein, and Lipid Data \(R34\)](#)  
[RFA-RM-14-012](#)
- [Novel and Innovative Tools to Facilitate Identification, Tracking, Manipulation, and Analysis of Glycans and their Functions \(U01\)](#)  
[RFA-RM-14-013](#)
- [Novel and Innovative Tools to Facilitate Identification, Tracking, Manipulation, and Analysis of Glycans and their Functions \(R21\)](#)  
[RFA-RM-14-014](#)
- [Facile Methods and Technologies for Synthesis of Biomedically Relevant Carbohydrates \(U01\)](#)  
[RFA-RM-14-015](#)



## Minimum Information Required for A Glycomics

### Experiment

- <http://www.mirage-beilstein.org>
- Definition of reporting guidelines for Glycomics experiments
- Guidelines **do not** define how to perform experiments
- Make results better **understandable** and **reproducible** by reporting required information
- Guidelines can be used by **journals** and **databases**



- **Kiyoko Aoki-Kinoshita**, Soka University, Tokyo, Japan
- **Catherine Costello**, Boston University, Boston, MA, USA
- **Anne Dell**, Imperial College London, UK
- **Niclas Karlsson**, Gothenburg University, Gothenburg, Sweden
- **Milos Novotny**, Indiana University, Bloomington, IN, USA
- **Nicolle Packer**, Macquarie University, Sydney, Australia
- **James Paulson**, The Scripps Institute, La Jolla, CA, USA
- **Pauline Rudd**, NIBRT, Dublin, Ireland
- **Michael Tiemeyer**, University of Georgia, Athens, GA, USA
- **Lance Wells**, University of Georgia, Athens, GA, USA
- **David Smith**, Emory University School of Medicine, Atlanta, GA, USA
- **Ten Feizi**, Imperial College London, London, UK

# Glycoinformatics Consortium (GLIC)について

野口研究所の支援によって結成した、世界中の糖鎖インフォマティクス研究者で構成されたコンソーシアム: <http://glic.glycoinfo.org>

これまで様々な国で糖鎖関連研究開発プロジェクトが実施され、その中の糖鎖インフォマティクス分野において各種データベース等が開発されてきた。世界の糖鎖インフォマティクス研究者が協調・連携しデータベースや各種ツールなどを共同で開発する仕組みを検討し、Glycoinformatics Consortiumを設立するに至った。

## GLICの目的

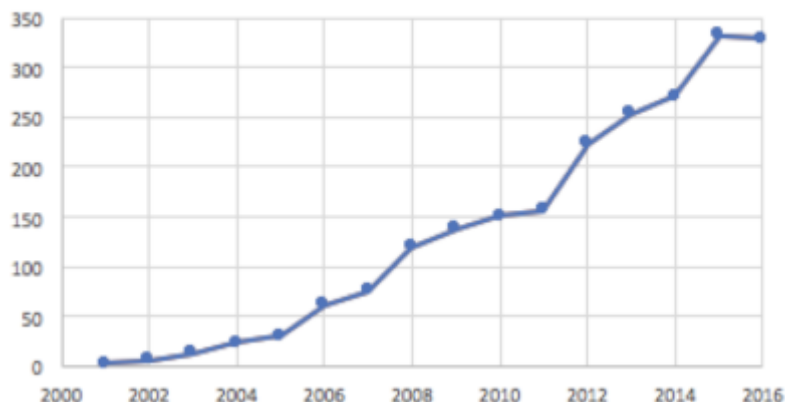
- 糖質情報学のための情報を一元化したソフトウェアリポジトリの提供および管理
- 糖質情報学・糖質化学・糖質生物学のための情報交換の場を設け、糖質科学の発展に貢献



# 糖鎖に関する文献の傾向

- 糖鎖の機能は複合糖質として考慮して初めて糖鎖の機能解明につながるということが明らかになってきた。
  - PubMedにおいて「glycomics」は2015年にピークを達し、「glycoproteomics」はこの2年間で倍増している。
  - 従って、糖鎖科学研究は糖鎖から糖タンパク質を含んだ複合糖質の研究にシフトしており、今後は複合糖質のデータがより重要かつ大量に生産されることが予測される。
- 問題点：
  - 糖鎖研究が未熟な学生や研究者は糖鎖について調べる場合、最新の情報を取得する方法は文献調査以外にない。
  - 現状のゲノム、プロテオーム、メタボローム、リピドーム、インタラクトームなどのオミックスデータベースは複合糖質をほとんど考慮しておらず、糖鎖科学に対する教育にも十分になされているとは言い難い。
- 従って、国内外の糖鎖研究を網羅的に整理し、既知な情報、未知な情報の実際の状況を把握するために、文献や関連データベースを統合し、ポータルサイトとして情報を提供することは急務である。

glycomics in PubMed

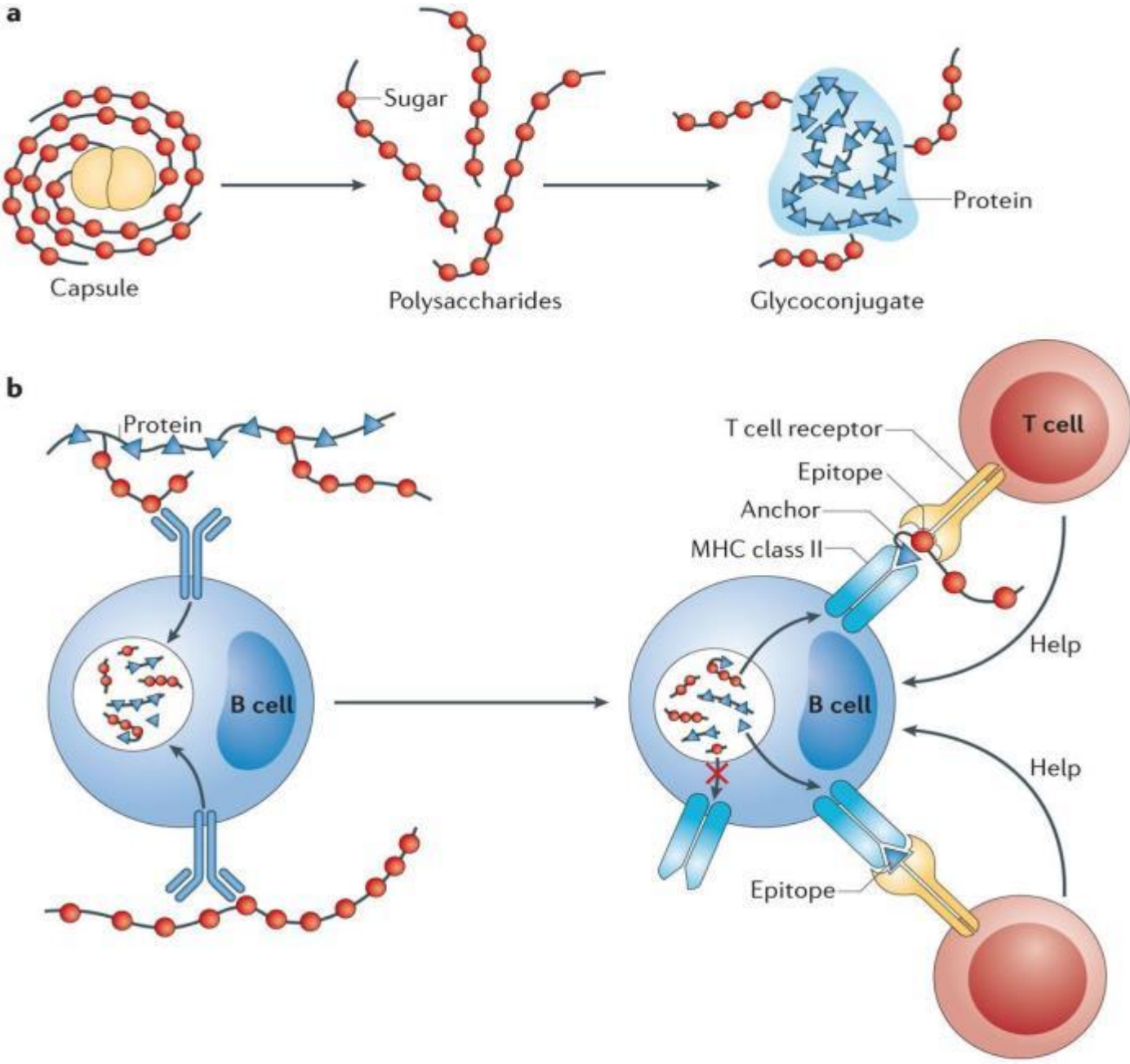


glycoproteomics in PubMed: ↗

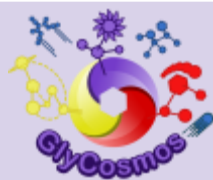


複合糖質のワクチン開発

In the 1970s, a group of scientists led by John Robbins at the US NIH revisited a paper that was published by Oswald Avery and Walther Goebel in 1929, which showed that **when the pneumococcal polysaccharide was conjugated to a protein, it was able to induce antibodies** (Figure). Robbins and colleagues used the same technology to make a covalent conjugate of diphtheria toxoid linked to the capsular polysaccharide of *H. influenzae* type b, and in 1980 they described the first conjugate vaccine. The conjugation technology heralded a major revolution in vaccinology. Within two decades, conjugate vaccines were licensed against *H. influenzae* type b, meningococcus serogroup C and seven serotypes of pneumococcus. In all cases, the vaccines induced high-affinity antibodies in infants, and they eliminated both the disease and the carriage of bacteria from the immunized population.



Rappuoli, R. & De Gregorio, E. A sweet T cell response. *Nature Med.* 17, 1551–1552 (2011).



# GlyCosmos Portal

GlycoCT  
IUPAC  
KCF  
LINUCS  
GLYDE-II  
CabosML

## Standards

Glycan Notation  
**WURCS & Tools**



全期間: 野口研・創価大

Ontology  
**GlycoRDF**

GO  
FALDO  
PIERO  
SIO  
CHEMINF

## Repositories

H29~H31: 創価大

GlyTouCan



**GlyComb  
Repository**



**ACGG-DB**

H29~: GL-i

**PDBj**  
Protein Data Bank Japan

H29~: 野口研  
2次元・3次元構造

**GlycoNAVI**  
Support System for Carbohydrate Research

*Molecular structure*

Glycoprotein

**Structural  
Glycobiology**  
H31~: 野口研

**Glycobiology  
Curation System**

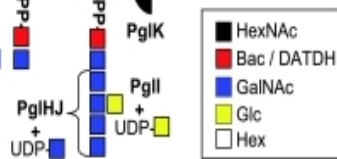
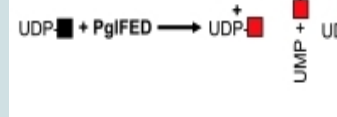
H31~H33: 創価大・新潟大

*C. jejuni*

Periplasm

Glycolipid

Cytoplasm



**Systems  
Glycobiology**  
H31~H33: 創価大



Japan Proteome Standard  
Repository Database

H29~: 新潟大



H29~: 創価大  
+ 新潟大



糖鎖が関わるパスウェイ  
(糖関連遺伝子を含む)

*Biological function*



**GlyCosmos Database**





# 国内の糖鎖科学研究コミュニティの規模

国内において7つの学会（2016年現在の会員数）

学会名	会員数
日本糖質学会	1,028
日本応用糖質学会	884
日本農芸化学会	10,530
日本脂質生化学会	521
日本質量分析学会	981
日本分子生物学会	13,305
日本生化学会	8,113

海外においても、Society for Glycobiology (SFG), EuroCarb, International Carbohydrate Organization (ICO), International Glycoconjugate Organization (IGO) などの国際会議にそれぞれ数百名の糖鎖科学研究者が参加している。

# 糖鎖研究コミュニティの要望

1. 糖鎖が特定の糖タンパク質やそのglycoformの立体構造と機能にどのように影響するかについての情報を取得したい
2. 酵素合成において、酵素の活性やその種類、入手先の情報を取得したい
3. ある糖鎖構造を持つと報告されたタンパク質・脂質を取得し、組織や細胞の種類別に整理した情報を取得したい
4. 特定の代謝やシグナル伝達経路に関連する糖鎖のリストを取得したい
5. 疾患で変化することが知られているグリコシル化経路中のタンパク質のリストを取得したい
6. 他生物種における糖鎖遺伝子のヒト・オルソログ（相同分子種）を検索して表現型付きのSNP情報を取得したい

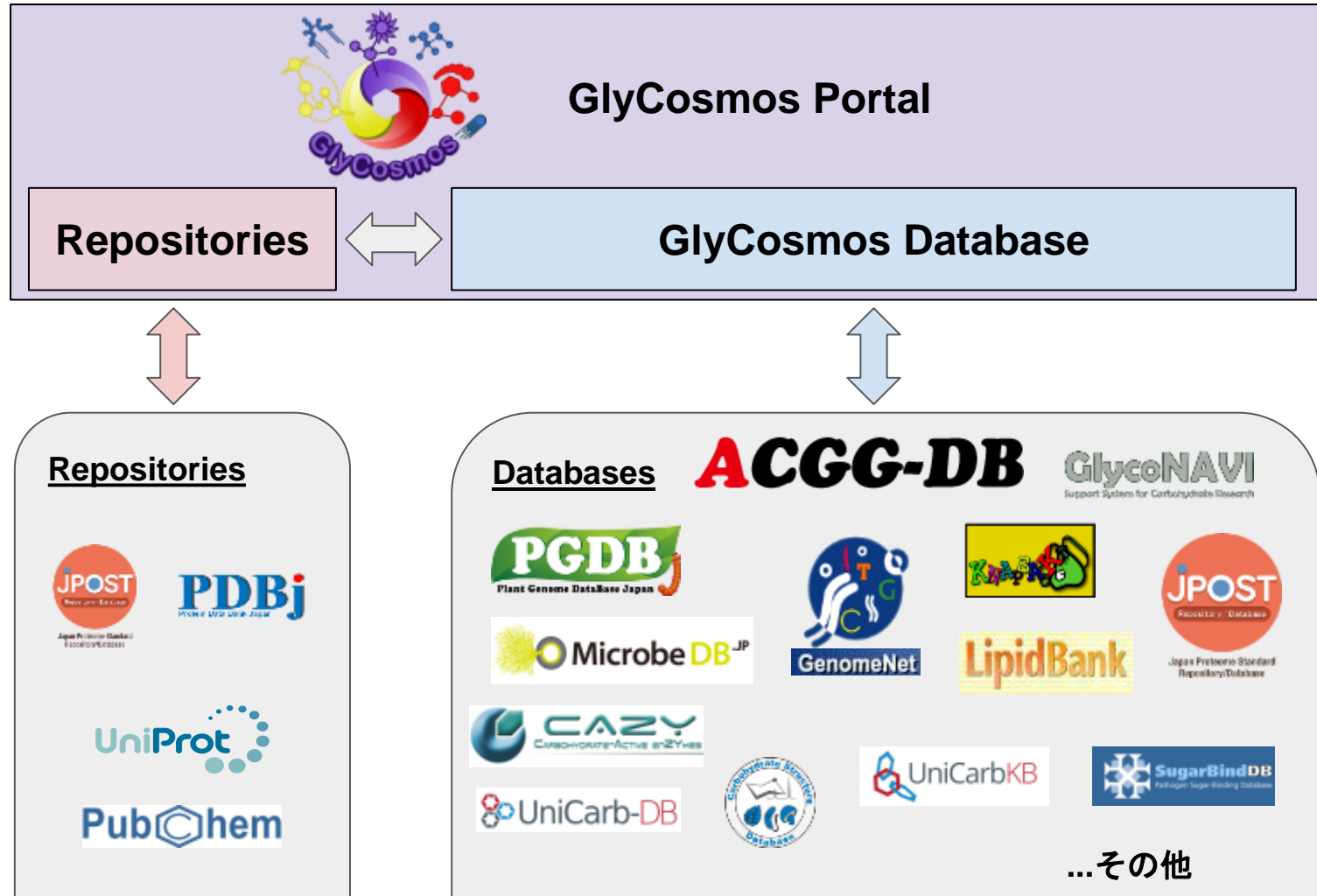
Glycoinformatics Consortium (GLIC)による調査結果, 2015年

# ねらい

- 日本の糖鎖研究コミュニティの強みである糖鎖遺伝子と糖鎖構造研究について、国内の糖鎖関係の学会と連携し、多くの国内外における関連知識を集約するデータベースとしてGlyCosmosを開発する。
- さらに、複合糖質データベース(GlyCosmosDB)を開発することで、プロテオーム、メタボローム、リピドームなどのオミックスの研究コミュニティと協働して研究開発を推進することで、より広範囲の連携を実現することが可能となる。
- 最終的には糖鎖研究に興味を持った研究者が本ポータルに訪れ、糖鎖に関する知見を容易に理解できるようになる。
  - 糖鎖遺伝子→関わっているパスウェイ、疾患、糖鎖構造、立体構造情報、複合糖質→糖鎖機能
  - 実験で確認された情報と、計算で予測された情報が格納されるため、仮説構築や事実確認に利用可能

# 本提案における統合の概要

## Integration of life science information



# 研究開発計画: 体制

共同研究グループ (3)  
GL-i 成松久

JCGGDB/ACGG-DBの維持管理およびGlyCosmosとの連携

連携研究者  
産業技術研究所  
梶裕之、久野敦、  
梶谷内晶、  
佐藤隆、岡谷千晶

研究代表者グループ  
創価大学 木下聖子

総括、糖鎖パスウェイ  
データのデータベース  
化、キュレーションシ  
ステム開発

連携研究者  
DBCLS  
金進東

JCGGDB  
ACGG-DB

GlyCosmos

アドバイザー

立命館大学  
日本糖質学会名誉会員  
JCGG監事

川崎敏祐

創価大学  
日本糖質学会理事  
JCGG副会長

西原祥子

Macquarie Univ.  
UniCarb-KB設立者  
MIRAGE Advisory Board member  
(質量分析の専門家)

Nicki Packer

Beilstein Inst.  
MIRAGEの設立者

Carsten Kettner

Univ. of Georgia  
CarbBank開発関係者  
(植物関連糖鎖の専門家)

William S. York

Imperial Coll. London  
EuroCarbDB開発関係者  
(質量分析の専門家)

Stuart Haslam

共同研究グループ (1)  
野口研究所 山田一作

糖鎖表記法WURCSの開発および複  
合糖質構造の整理・データベース化

共同研究グループ (2)  
新潟大学 奥田修二郎

糖鎖情報の自動キュレーションシ  
ステムの開発および糖鎖遺伝子デ  
ータの収集・整理

支援団体

日本糖質学会,  
JCGG, FCCA,  
MIRAGE, GLIC

# ACGG-DBの機能拡張とGlyCosmos portalとの連携 およびアジア地域との連携

## 研究概要

- 糖鎖研究における日本発のオリジナルデータの収集
- Semantic Webに対応したデータやOntologyを日本及びアジアの糖質研究者に向けて、有用な情報の提供 (<http://acgg.asia/db/>にて公開)
- マウスやヒトの糖鎖データのGlyCosmos との連携による他生物糖鎖DBとの統合

糖タンパク質



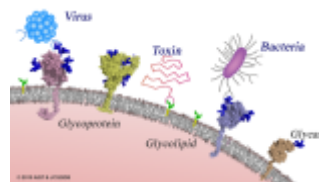
糖鎖遺伝子



レクチン



糖鎖疾患関連



## 研究計画・進捗

2017

ACGG及びACGG-DBのPortal siteを公開



オリジナルデータの追加やSemantic Webに対応した糖鎖DB間の検索機能を開発

糖鎖の構造と機能を直感的にリンクさせるViewerによる表示など、ユーザーフレンドリーなインターフェースの改良

GlycoProtDBとUniProtやGlyCosmos等の他DBとの連携や、ACGG参加国内での糖鎖DBとのクロスリンクを進める

2021

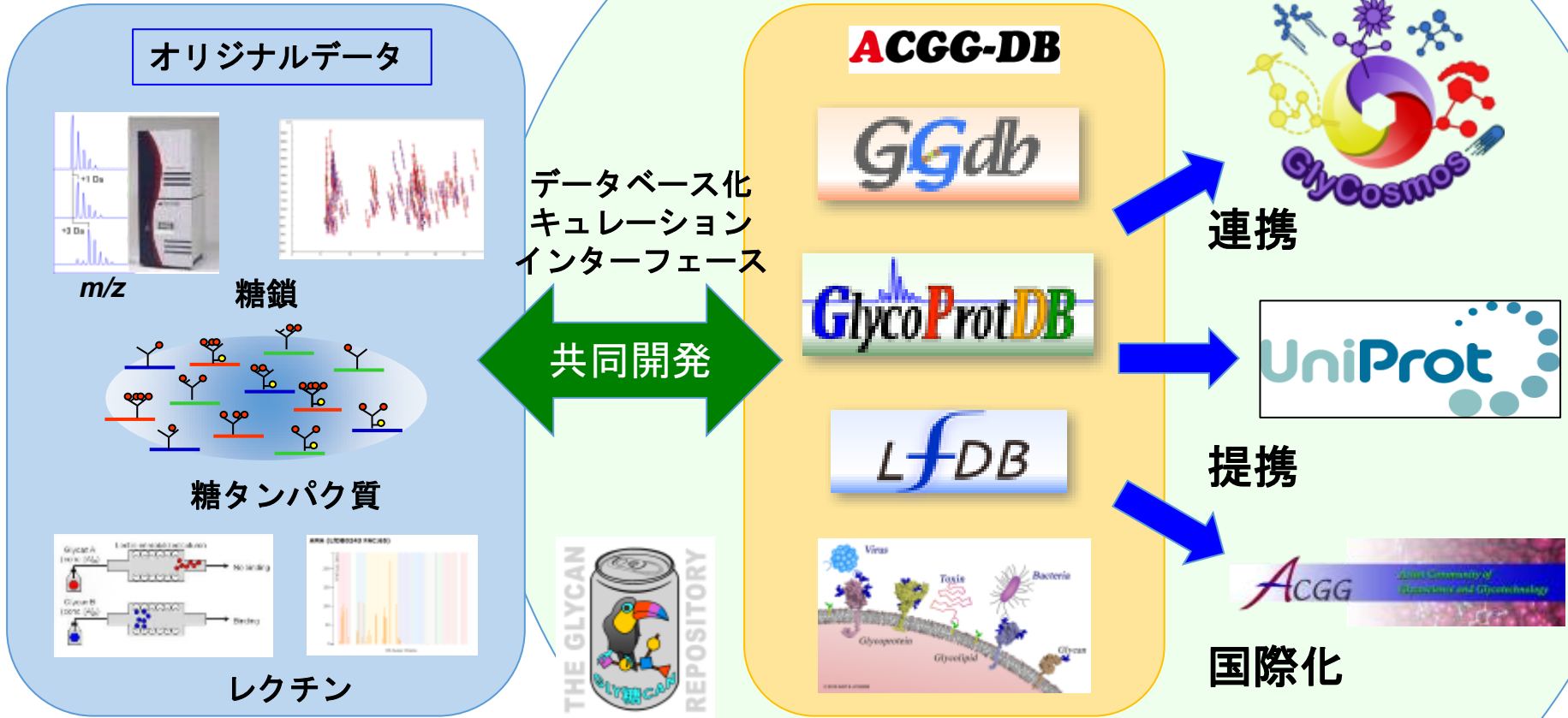
**ACGG-DB**



# ACGG-DBの機能拡張とGlyCosmos portalとの連携 およびアジア地域との連携

産総研：データ提供

GL-i：DB開発



# 主なWURCSの開発

バクテリアなど複雑な単糖を含む場合、単糖の置換基の曖昧さ表現が必要となる

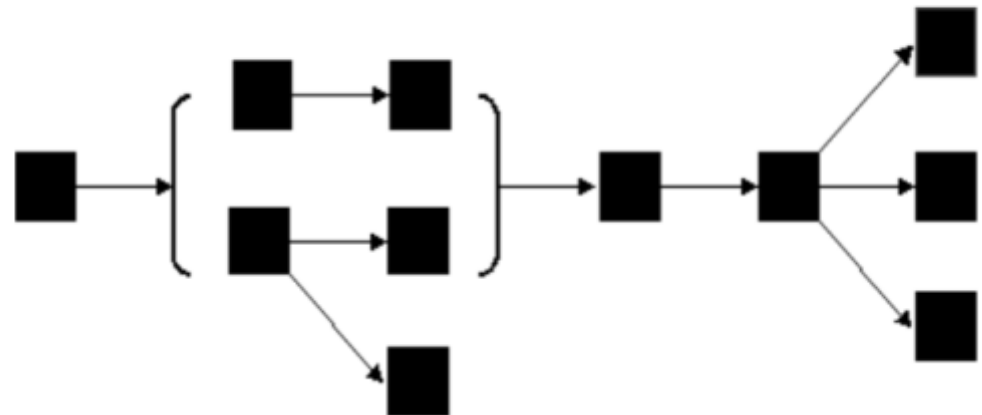
– 曖昧な置換基表現への対応

必要な曖昧表現

- OR表現
- 結合位置の候補が複数
- 位置不定の二重結合
- 繰り返し

## 2.1.4 Subgraphs: Alternative Units (a)

Alternative units are marked in the residue list with the type identifier 'a' and a canonical number. They are further broken down in the alternative unit section (ALT).





## GlyTouCanの改良：

- GlycomeDBのメタデータ導入（8月4日完了）
- コメント機能の追加
- ユーザーの個人ページの開発
- モチーフ情報の整理

## GlyCombの設計：

- アクセッション番号のシステム  
（4月28日アドバイザー会議にて案決定）
- UniProtとの打ち合わせ

KEGG GLYCANの糖鎖情報を調べると、糖鎖合成経路だけではなく、シグナリング経路、細胞間コミュニケーション、免疫系のパスウェイにも関わっていることが掲載されている。

## Pathway Maps and Structure Maps

**KEGG GLYCAN** originally represented the glycan structure database mentioned above, but now it is used in a broader sense to represent the KEGG resource integrating glycan structures, genes, pathways and diseases [2]. In the **KEGG PATHWAY** database the section on:

- **Glycan biosynthesis and metabolism**

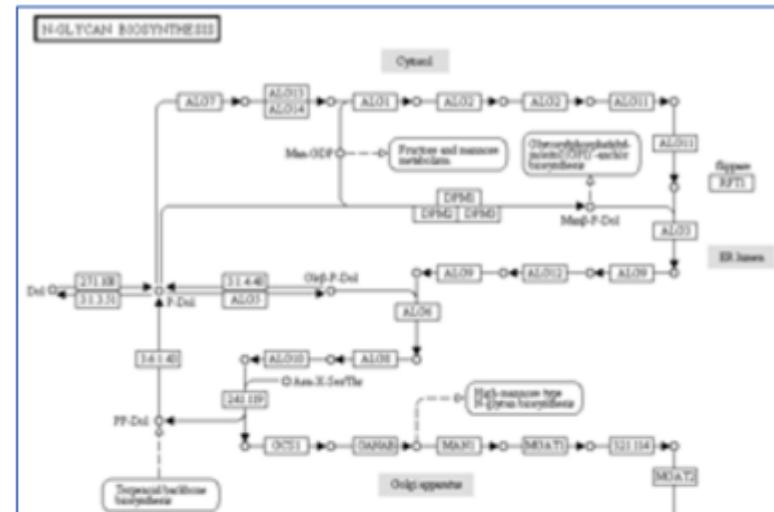
contains metabolic pathway maps for glycans. Some of them contain an additional representation of glycan biosynthesis or degradation, called the glycan structure map. For example, N-glycan biosynthesis is shown both as a pathway map and a structure map in [map00510](#).

The roles of glycans in various cellular processes may be examined with other pathway maps in the KEGG PATHWAY database, such as:

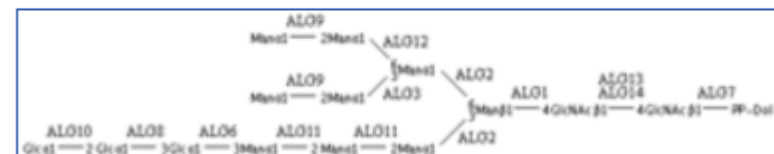
- Signaling molecules and interaction
- Cell communication
- Immune system

In contrast to the manually drawn KEGG structure maps, the Composite Structure Map (CSM) is computationally generated to represent all possible structures in the KEGG GLYCAN structure database [3]. CSM can be used to examine the structural repertoire inferred from the genomic or transcriptomic repertoire.

- **KEGG GLYCAN composite structure map**



N-glycan biosynthesis (pathway map)



N-glycan biosynthesis (structure map)

# 糖鎖関連Pathway数

- UniProtの約3万の糖タンパク質をReactomeのAnalysis Toolsに入力したところ、1660のヒトの糖タンパク質がヒットした。そしてこれらはすべてのパスウェイに関連していた。
- 糖鎖合成経路だけではなく、シグナリング経路、免疫系、細胞外マトリックス、発生など、糖タンパク質はありとあらゆるパスウェイ6773個に関わっていることがわかる。

REACTOME 3.2 59 Pathways for: Homo sapiens

Event Hierarchy:

- Cell Cycle (65/1,534) FDR: 1E0
- Cell-Cell communication (99/422) FDR: 1E0
- Cellular responses to stress (87/1,008) FDR: 1E0
- Chromatin organization (33/507) FDR: 1E0
- Circadian Clock (18/358) FDR: 1E0
- Developmental Biology (453/2,467) FDR: 1E0
- Disease (488/2,703) FDR: 1E0
- DNA Repair (48/816) FDR: 1E0
- DNA Replication (6/150) FDR: 1E0
- Extracellular matrix organization (300/604) FDR: 1E0
- Gene Expression (209/4,014) FDR: 1E0
- Hemostasis (396/1,754) FDR: 1E0
- Immune System (1,012/4,024) FDR: 1E0
- Mitophagy (7/81) FDR: 1E0
- Metabolism (607/2,934) FDR: 1E0
- Metabolism of proteins (578/3,151) FDR: 1E0
- Muscle contraction (84/304) FDR: 1E0
- Neuronal System (232/952) FDR: 1E0
- Organelle biogenesis and maintenance (30/647) FDR: 1E0
- Programmed Cell Death (68/663) FDR: 1E0
- Reproduction (21/30) FDR: 1E0
- Signal Transduction (1,392/5,352) FDR: 1E0
- Transmembrane transport of small molecules (383/850) FDR: 1E0
- Vesicle-mediated transport (157/1,259) FDR: 1E0

# 具体的な実施計画：GlyCosmosの設計

## パスウェイ情報の収集・整理・統合

- WikiPathwaysの導入（10月～）
- Reactomeの糖タンパク質が関連するパスウェイを導入する
- 糖タンパク質情報とのリンクを付加し、糖鎖の関与をメタデータとして記載する

## 複合糖質のデータ整理・統合

- タンパク質および脂質DBより、複合糖質の構造を解析・整理し、化学構造を用いた統合を実施

## 糖鎖遺伝子の環境分布情報の収集・整理

- 環境メタゲノムデータからの糖鎖関連遺伝子探索
- CSDBと連携し、環境と糖鎖構造とのリンク付けを実施

# 達成目標

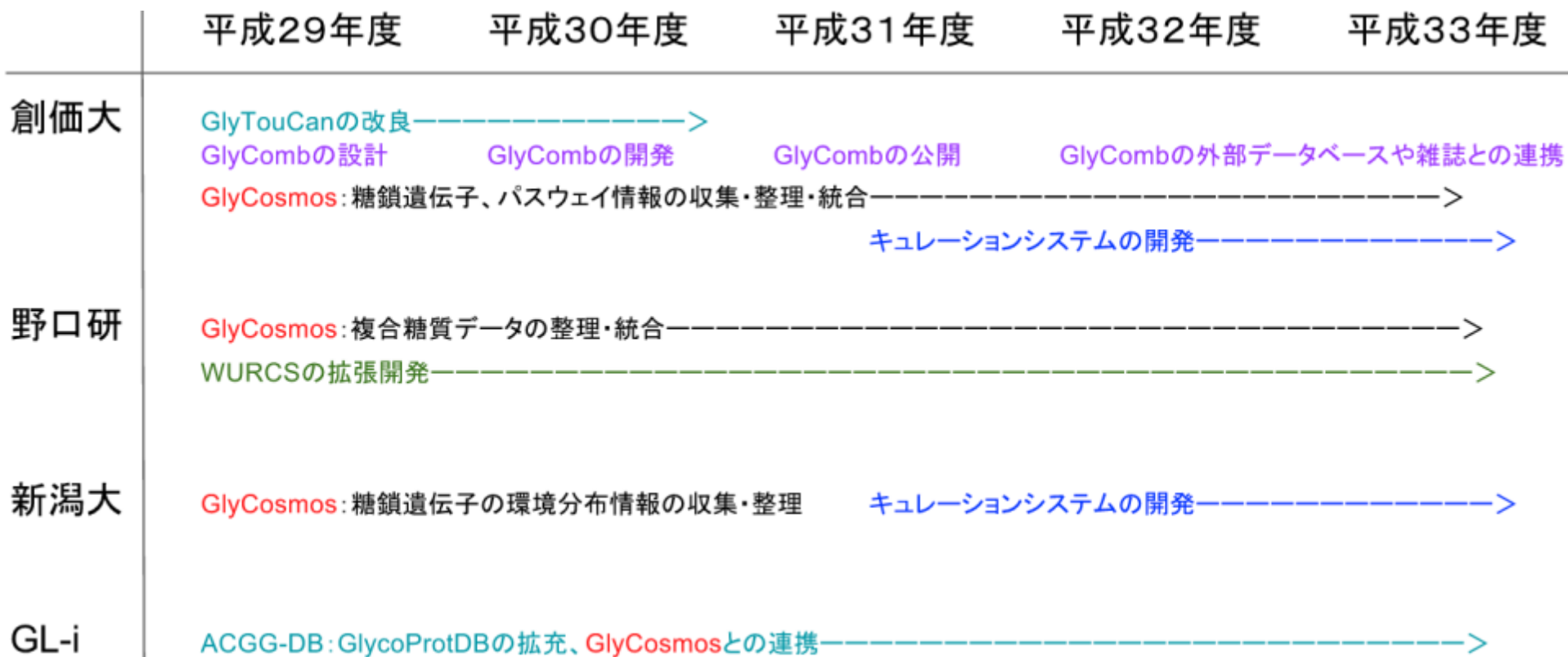
## ～第3年次末

- **GlyCosmos Portal :**
  - GlyCombが稼働、糖タンパク質を登録し、アクセッション番号を取得可能
- **GlyCosmos Database :**
  - 微生物のゲノム情報より糖鎖関連遺伝子を抽出し、糖転移酵素を糖鎖構造とリンクし、GlyTouCanやパスウェイ情報と統合したユーザーインターフェースが開発済
  - BCSDDBに格納されている半分以上の糖鎖構造に対する合成しうる遺伝子をGlyCosmosから公開
  - GlycoProtDBおよび全PDBデータ中の糖鎖構造とのリンクとStanzaによる可視化が完成

## ～第5年次末

- **GlyCosmos Portal :**
  - GlyTouCanおよびGlyCombリポジトリの利用が普及し、各国からユーザーが利用
  - 複合糖質の情報から糖鎖の合成パスウェイおよび糖転移酵素とリンクし、ゲノムとの関係がわかるユーザーインターフェースが完成
  - GlyCosmosは日本糖質学会の公式データベース
- **GlyCosmos Database :**
  - 植物のゲノム情報より糖鎖関連遺伝子を抽出し、糖転移酵素を糖鎖構造とリンクし、糖鎖に関連するパスウェイを検索しやすいユーザーインターフェースが開発済
- **キュレーションシステム:**
  - N型糖鎖とO型糖鎖のパスウェイが網羅していて、一般ユーザーがコメントを投稿でき、キュレーターが編集を記録できるシステムが完成

# 研究開発計画：実施計画・スケジュール



# データ提供者、データ利用者との連携・協業の計画

## データ提供者

- ウェブ上のデータ提供者
  - GL-iおよび産総研の連携研究者
  - 日本糖質学会会員
- データベースのデータ提供者
  - PDBj, LipidBank, MicrobeDB.jp, jPOST, PGDB, KEGG, PubAnnotation
  - GlyTouCan Partners (ACGG-DB(JCGGDB), UniCarbKB, UniCarb-DB, CSDB, SugarBind, ....)

## データ利用者

本提案においては、以下の会議でハンズオンセミナーやチュートリアルを開催し連携や協業について協議する予定である：

- 日本糖質学会・バイオインフォマティクス・ワークショップ
- 分子生物学会年会ブース出展
- ACGG-DB会議・ハンズオンセミナー
- ワーレンワークショップ国際会議・Beilstein Glyco-Bioinformaticsシンポジウムのデモセミナー
- Society for Glycobiology (SFG)ハンズオンセミナー
- Glycoinformatics Consortium (GLIC)ハッカソン

なお、GlyCosmos PortalにWeb formも用意し、フィードバックを受け入れられるように開発を進める。

# 搭載するデータの量・範囲の妥当性について

初期のGlyCosmos Portalが搭載するデータ数：

## Repository

- GlyTouCan >81,000 entries
- GlyComb >57,000 entries (LipidBank + UniProt)

## Database

- >57,000 glycoconjugate entries (UniProt + PDB + LipidBank)
- >600 glycogenesに加え、レクチンやパスウェイ関係の遺伝子など多数

根拠：

## Glycoconjugates:

PDBに糖が含まれるエントリー数：8,366

UniProtの糖鎖修飾されたタンパク質の数：29,994 (GlycoProtDB：12,894)

LipidMAPS(LipidBank)の糖脂質の数：27,509

## Genes

GGDBのエントリー数（ヒト）：221

マウスにおける糖関連遺伝子の内部調査の結果：388

CAZy Genomes: Bacteria 7853、Viruses 331、Archaea 283、Eukaryota 212

## Pathways

Reactome のパスウェイ：6773

KEGGとの連携も検討中

(追加実施) 総合グライコミクスの複合糖質の数：約200種類＋組織、細胞種別、パスウェイ情報



# GlyComb、GlyCosmosポータル、GlyCosmos Databaseの アクセス数見込み

以下の左表はGlyTouCanの平成27年度のアクセス数であり、右表は平成23年におけるJCGGDBのデータベース別のアクセス数であり、JCGGDBは月間約2万のアクセス数がある。

GlyTouCan	平成27年度のアクセス数
訪問者数	4,610
訪問数	10,844
ページ数	49,122

平成27年6月公開開始

データベース名	月間平均アクセス数
GGDB	14,381
LfDB	1,967
GlycoProtDB	3,061

Synthesiology, Vol.5, No.3, pp190-203 (2012)

- GlyCombはGlyTouCanのアクセス数よりは多くなる見込みで年間アクセス数1万件を見込む。
- GlyCosmos DBはパスウェイや他のゲノム情報を含むため、JCGGDB（年間アクセス数約24万件）より多くのアクセス数の予想である
- GlyCosmosポータルは上記のアクセス数に加え、糖鎖関係者以外のアクセスが予想されるため、年間で約30万件のアクセス数を見込む

# GlyCosmosで期待する効果

- glycoproteomics研究が進み、その情報基盤を作成する必要がある
- また、糖鎖研究の初心者でも糖鎖について容易に調べられるポータルを作成し、糖鎖研究者以外にも幅広い分野の研究者が利用できるポータルを目指す
- 糖鎖情報に興味を持ち、糖鎖の重要性がわかってきた時に、必要な情報を容易に得られるようにする