

# 平成24年度「糖鎖統合データベースと研究支援ツールの開発」の成果報告

代表研究機関：産業技術総合研究所

共同研究機関：立命館大学・野口研究所・理化学研究所・創価大学

2013年1月21日



【研究開発計画】

研究支援ツールの構築

化学系実験プロトコル



- ・合成のプロトコル・反応・精製方法を取りまとめる(野口研)
- ・合成化合物(リソース)の提供(産総研)
- ・公開を目指す

有機合成:  
システム構築(産総研)・運営支援  
ワーキンググループ(野口研究所・東北大学  
岐阜大学・東海大学・産業技術総合研究所)  
によるデータ蓄積し始める



生物学系実験プロトコル



実務者のノウハウ  
毎年30のプロトコルを執筆・公開  
(産総研・立命館)



その他様々なツール開発



**統合化支援**

新規公開&既存DBの統合(国内)  
未公開データをDB化支援(国内)  
統合化技術開発  
海外DBとの連携

国際会議(2回開催)を通して情報交換・共同開発

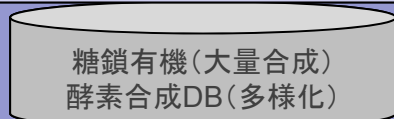
アジア・欧米との連携・統合化のための基盤整備

異分野のDB

RDF化の準備  
URIの準備

連携強化

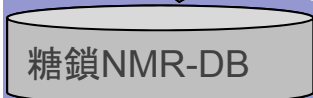
合成方法・生成物確認



糖鎖有機(大量合成)  
酵素合成DB(多様化)



NMRのデータの蓄積を継続



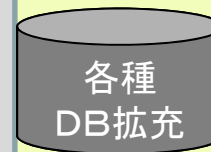
合成物の確認  
(研究支援)

JCGGDBの機能強化と拡張

DB間連携・有機的統合化を推進

- 阻害剤のデータベース
- 糖転移酵素の基質特異性のDB
- グリコシダーゼの基質特異性のDR  
Human Glyco-Disease Genes Database
- モデル生物のフェノタイプ情報

遺伝子名  
糖鎖構造  
構造名称



各種  
DB拡充

GlycoEpitope(抗糖鎖抗体DB)の  
高機能化立命館大学)

- ・糖タンパク質の実験データを充実
- ・論文で報告されている糖鎖修飾情報を充実させる(産総研)

# 実験データの公開や DBの高機能化

# 糖タンパク質データベース(マウス版)公開 GlycoProtDB

産総研



内容 IGOT法を用いて実験による糖鎖付加位置を同定したデータを公開した。

マウスの組織名と捕集分子(レクチンやHILICなどのカラム)の組み合わせで得られた糖ペプチド情報とその糖鎖付加位置を分かりやすく表示した。

マウスの情報を2012年4月に公開した。

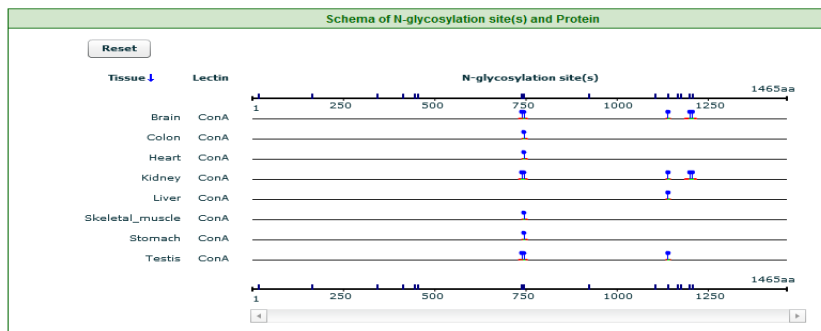
今後は、糖転移酵素のノックアウトマウスの糖タンパク質の情報を追加する予定となっている。

GeneSymbols: 1819  
Glycosites: 5657

GlycoProt ID: GP\_MMU\_00000096

Protein Name: PREDICTED: membrane protein CH1 [Mus musculus].  
Protein Accession Number: XP\_980145  
Protein GI: 94364482  
Gene Symbol: AI848100  
Gene Name: expressed sequence AI848100  
Gene ID: 226551

### Glycosylation Sites



### Sequence / Motifs

1	MTKLVLSIKVR	SESVKLE	REFYGGKKK	AGSGRRITSP	RYPCAEEARA	SESGQAIERR	KTEGYRTSCS	SASEWVGPAQ	VGRLLRWFGA	LSAEGAWPAL
101	TERSGPRAGR	GWGGSGGGR	RRRARASGQT	ERRRQSGPER	SCASLIFLPL	WASRGGHPGR	EPLRNRSEFP	TALRTLGPIL	AFILLRLHLG	LDSGCGGEDV
201	YTVVQQLLET	TKLSLPVVEA	LPTVDLHRES	SRVVGSETI	ENSSSSSTSE	RTPVSELDV	EKSGTSLIAK	PCVEQPRAD	CDAGRADDAD	APVQDAFVS
401	PPESLVQGH	ENVSSSHGKE	KVTKSEFESK	VSVEQDGGD	PKSALNISTD	LKNESSDYTK	PGSDTPTST	SPKDFEDIFT	FDWKKKVMK	VEKKSLSLTK
501	QSLHPSINGG	PHATKKVQVK	RNNYASVECK	AKILAAPEA	KSTSALIEN	MDLVMNLNCS	TKIWFVIELC	EPIQVQFDI	ANYELFSSFP	KDFPLDINTV
601	YPTNWKIKLG	TFHGRDERNV	QSPFLDQMY	ACVVMFVKY	IKVLLSHFG	SEHFCPLSLI	RVFGTSMVEE	YEEIADSVQY	SERQELFDD	YDFPLDINTV
701	EDKSKRLLIG	SATNAILMV	NIAANILGAK	TEDLTEG	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE
801	EASPTVTLG	GGSGQDRSS	SWFSETHL	CELTIGICCI	SSSEYIYEW	CVRIALYRG	ASVYFSGKD	FVFPQSLILL	PVLSVSVVFP	QPSVQVDE
901	NMERAEATVD	LDLSSVHGQ	HLTHIVDTI	ELKPSVPTI	SGQLLDDVP	ENSLKVEG	SESVKSEGGY	IPSLQATKE	SVFDDKTEK	KTESFSSAEK
1001	LSVIYETSKV	NEVMDNTVKE	DILSTEVVTK	FRETVPVPM	NTATVPEGS	VETKPSIADT	LKHTVTPMD	PSLVEKDE	QSEDELLAG	LQRTATDFYA
1101	ELQNSIDLGY	GNGNLVHGS	GKESVFMRLN	NRKALEV	ISL	ISL	ISL	ISL	ISL	ISL
1201	TQLVLS	SAT	VAEIKREVSD	RQSYLVMSLV	LCVVLGMLC	MQRCTTSQF	DGDIKSLPK	SNQVPSKRC	FSYDMMNLC	RRTSFLIRS
1301	DPNDLYIVPE	LKFSPEKKK	RCKYKTEKE	TIKPADLHP	IANGDIKRC	PFTNQDFSN	MGEVHSSYK	GPPSEGSST	SSQSESYFC	GISACTSLCN
1401	GGTKRTEK	RALKRRRSVK	QDQGLIKAL	IQTKSSGLNS	LHDIKGNKE	ITVGARVTA	VSGHI			

### N-glycosylation sites

Potential sites	Position	Sequon	Peptide fragment*1	Unique or Shared	Brain								Colon	Heart	Kidney	Liver	Skeletal_muscle	Stomach	Testis		
					ConA	ConA	ConA	ConA	ConA	ConA	ConA	ConA	ConA	ConA	ConA	ConA					
738	NKS	731-752	TEDLTEG	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	
			731-752	TEDLTEG	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE
744	NAT	740-752	SISE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	
			740-752	SISE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE	ISSE
1139	NMS	1135-1145	ALEV	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	
			1135-1145	ALEV	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL
1199	NMT	1186-1216	QTEAHLQAQLT	MTQLVLS	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	
			1186-1216	QTEAHLQAQLT	MTQLVLS	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL
			1186-1216	QTEAHLQAQLT	MTQLVLS	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL
1206	NLS	1186-1216	QTEAHLQAQLT	MTQLVLS	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	
			1186-1216	QTEAHLQAQLT	MTQLVLS	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	
			1186-1216	QTEAHLQAQLT	MTQLVLS	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	ISL	

\*1: .:Motifs potential site ■:Asn (glycosylation+180) ■:PyroGln (N terminus) ■:Met (oxidation) ■:Deamination (NtermCys)

Kaji et al. J Proteome Res. 2012 Sep 7;11(9):4553-66

©2013 成松 久(産業技術総合研究所) licensed under CC表示2.1日本

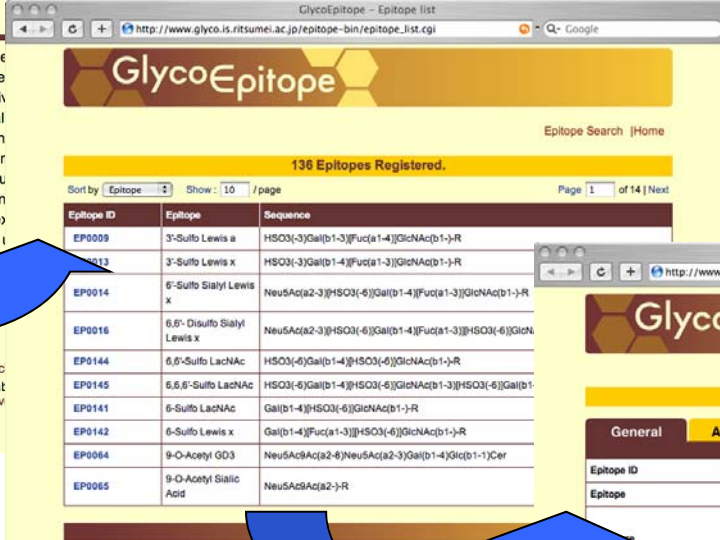




# GlycoEpiteope Database

<http://www.glyco.is.ritsumei.ac.jp/epitope/>

平成25年度  
新規抗体30件を  
収載予定



## 糖鎖抗原・抗体の 機能データベース

〈背景〉平成17年 立命館大学糖鎖工学研究センターに創設  
研究者がインターネット経由でデータベースの作成に参加

〈内容〉 171 Epitopes, 608 Antibodies (2013.01.18)

糖鎖エピソードを発現する糖タンパク質、糖脂質、プロテオグリカン、植物多糖  
糖鎖エピソードの合成や分解に関する生合成酵素および分解酵素  
糖鎖エピソードの発現する時期・場所、糖鎖エピソードの関連する疾病  
糖鎖認識抗体の入手先

引用文献: PubMedへのリンク, PDFファイルに直接リンク



# 横断検索のMicrodata埋め込み

産総研

- 「検索エンジンのための」メタデータであるmicrodataを各DBの詳細ページに埋め込むことで、一般の検索エンジンや横断検索での結果表示の際にリッチスニペット化が期待できる。
- ※ユーザは検索結果がどのようなエントリにヒットしたか理解しやすくなる。
- 「[BH12.12/ schema.org](http://BH12.12/schema.org)のライフサイエンス用語彙の作成」で暫定的に定義されたボキャブラリ(プロパティ)に準拠した。

- 12月の分子生物学会のブース展示の際に、Microdataの埋め込みに協力することになった。

- 今年度、埋めこみ作業中
  - 病原体と糖鎖(PACDB)
  - ノックアウトマウスを用いた  
機能糖鎖科学データベース
  - 糖鎖関連疾患遺伝子データベース
  - JCGGDB リポート
- HyperEstraiierのindexに取り込んだもの
  - PACDB

```

<B><FONT color="RED"><a itemprop="url" href="http://jcgddb.jp/sea
<span itemprop="keywords">Carbohydrate binding</span>
<span itemprop="keywords">Disease</span>
<TABLE border=0 cellpadding=0 cellspacing=0 width=800>
<TR><TD><LI>Respiratory Tract Disease / <FONT COLOR="RED">
<span name="Host">
<meta itemprop="description" content="Legionnaires' Disease is caus
<meta itemprop="description" content="host" />
<meta itemprop="url" content="http://www.uniprot.org/taxonomy/9606
<meta itemprop="entryID" content="9606" />
</span><TD>caused by&nbsp;&nbsp;&nbsp;&nbsp;<U><I>Legionella pneumophila
<span itemprop="taxon" itemscope itemtype="http://schema.org/Biol
<meta itemprop="name" content="Legionella pneumophila" />
<meta itemprop="description" content="pathogen" />
</span></span>

```

# 統合化の技術開発

# 収集している用語リスト

産総研



各種DB

糖鎖構造・名称・ID  
遺伝子・GeneSymbol, ID  
生物種情報・Taxonomy ID (略称)  
実験プロトコル(方法名・装置名・名称・遺伝子・糖鎖構造)



論文

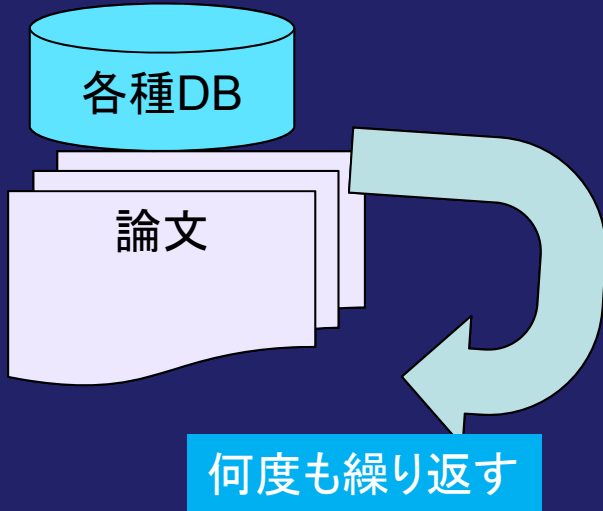
Abbreviationと展開形  
単語、フレーズ(単・複数形など)  
糖鎖構造・表記(シノニム)  
人名(著者名)  
化学物質名



# 収集過程・シノニム吸収フィルター

産総研

単語やフレーズの表記ゆれを集める

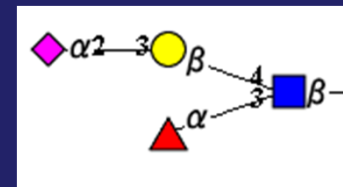


Sialyl-Lewis x  
 sLex  
 Si-Lex  
 sialyl Lewis x  
 sialyl Lewis x epitope  
 Sialex  
 sialyl Lewis X tetrasaccharide  
 NeuAc[alpha]3Gal[beta]4[Fuc[alpha]3]GlcNAc-R  
 Neu5Ac[alpha]2-3Gal[beta]1-4(Fuc[alpha]1-3)GlcNAc

同じものを意味する表記には同じ番号になるように処理する

JWG000001015

文章中のシノニムを吸収するフィルターとなる



# SweetWordFinder

産総研



URL 又は 英文

コピー & ペースト

興味のある単語をクリックするだけ

GD1b

GT1b

実行ボタン

用語集に登録されているキーワードにリンクが付く(最長文字列にヒット)

Keyword

Binding of pathogenic bacteria and bacterial toxins to host cell surfaces is an essential step in establishing infection in tissues and producing toxic effect. Glycosphingolipids on cell surfaces are receptors for binding to cells.

There are many bacterial toxins that bind to ganglioside, an acid glycosphingolipid, as the receptor on cell surface and invade host cells. The best known of these is the cholera toxin, an enterotoxin produced by *Vibrio cholerae*, and its specific cell surface receptor was identified as ganglioside GM1. Cholera toxin consists of a monomeric B subunit that binds to GM1 and an A subunit with direct toxic activity. The binding of B subunit to membrane GM1 may induce a conformational change in the toxin, resulting in the entry of the A subunit into the cell. The B subunit contains 108 amino acid residues, and Arg-55 and Asp-59 participate in binding to GM1. The mechanism of binding of the heat-labile toxin produced by enterotoxigenic *Escherichia coli*, which is structurally related to cholera toxin is similar to the mechanism of cholera toxin binding. Other ganglioside-binding bacterial toxins include tetanus toxin (GQ1b), botulinum toxins (GT1b and GT2b) and delta toxin produced by *Clostridium perfringens* (GM2). Shiga toxin produced by *Shigella dysenteriae* and Vero toxin produced by enterohaemorrhagic *E. coli* bind to neutral glycosphingolipids having an alpha-1,4 galactose moiety in

Search

Binding of pathogenic bacteria and bacterial toxins to host cell surfaces is an essential step in establishing infection in tissues and producing toxic effect. Glycosphingolipids on cell surfaces are receptors for binding to cells. There are many bacterial toxins that bind to ganglioside, an acid glycosphingolipid, as the receptor on cell surface and invade host cells. The best known of these is the cholera toxin, an enterotoxin produced by *Vibrio cholerae*, and its specific cell surface receptor was identified as ganglioside GM1. Cholera toxin consists of a monomeric B subunit that binds to GM1 and an A subunit with direct toxic activity. The binding of B subunit to membrane GM1 may induce a conformational change in the toxin, resulting in the entry of the A subunit into the cell. The B subunit contains 108 amino acid residues, and Arg-55 and Asp-59 participate in binding to GM1. The mechanism of binding of the heat-labile toxin produced by enterotoxigenic *Escherichia coli*, which is structurally related to cholera toxin is similar to the mechanism of cholera toxin binding. Other ganglioside-binding bacterial toxins include tetanus toxin (GQ1b), botulinum toxins (GT1b and GT2b) and delta toxin produced by *Clostridium perfringens* (GM2). Shiga toxin produced by *Shigella dysenteriae* and Vero toxin produced by enterohaemorrhagic *E. coli* bind to neutral glycosphingolipids having an alpha-1,4 galactose moiety in

Word ID: JWG000003582 [Merge IDs]

GPH00001659	• GD1b
GW00001088	• GD1b

JCGG-STR024670

Glycobiology, Volume 20, Issue 1, Pp.41 INT[Line:6]

Binding of LT-IIa has been well characterized and involves GD1b, GM1, GT1b, GQ1b, GM2, GD1a, and GM3 (Fukuta 1988).

Glycobiology, Volume 21, Issue 10, Pp.1373 INT[Line:7]

Ganglioside biosynthesis occurs by sequential glycosylation reactions via two major pathways, designated the "a-pathway" (GM2, GM1a and GD1a) and the "b-pathway" (GD3, GD2, GD1b, GT1b and GQ1b), with a common precursor, GM3, in which analogous steps in the two pathways are catalyzed by the same glycosyltransferases (Ledeen and Yu 1982; Hettler 2003).

Glycobiology, Volume 20, Issue 1, Pp.41 INT[Line:4]

Cholera toxin binds primarily to GM1, although it also exhibits weaker affinity binding to GM2, GD1a, GM3, GT1b, and GD1b (Bennett and Cuatrecasas 1975).

Glycobiology, Volume 20, Issue 6, Pp.668 ABS[Line:3]

All gangliosides (GM1, GM2, GM3, GD1a, GD1b, GD3 and GT1b) examined showed good neutralizing capacity against VacA.

Word ID: JWG000019991 [Merge IDs]

GPH00001831	• GT1b
GW00001175	• GT1b

JCGG-STR005131

Glycobiology, Volume 20, Issue 1, Pp.41 INT[Line:6]

Binding of LT-IIa has been well characterized and involves GD1b, GM1, GT1b, GQ1b, GM2, GD1a, and GM3 (Fukuta 1988).

Glycobiology, Volume 21, Issue 10, Pp.1373 INT[Line:7]

Ganglioside biosynthesis occurs by sequential glycosylation reactions via two major pathways, designated the "a-pathway" (GM2, GM1a and GD1a) and the "b-pathway" (GD3, GD2, GD1b, GT1b and GQ1b), with a common precursor, GM3, in which analogous steps in the two pathways are catalyzed by the same glycosyltransferases (Ledeen and Yu 1982; Hettler 2003).

Glycobiology, Volume 20, Issue 1, Pp.41 INT[Line:4]

Cholera toxin binds primarily to GM1, although it also exhibits weaker affinity binding to GM2, GD1a, GM3, GT1b, and GD1b (Bennett and Cuatrecasas 1975).

Glycobiology, Volume 20, Issue 6, Pp.668 ABS[Line:3]

All gangliosides (GM1, GM2, GM3, GD1a, GD1b, GD3 and GT1b) examined showed good neutralizing capacity against VacA.

# 統合化のインターフェース (試作)

JCGGDB Alliance DBs

遺伝子情報・基質特異性

オーソログ情報

Word ID: JWG000019315 [Merge IDs]

AB000775	• Abbreviation "GD3" stands for CD60a.
GPH00000823	• CD60a
GPH00001661	• GD3
GW00000529	• CD60a
GW00001090	• GD3
	ST0758:GD3 ST0915:GD3
JCGG-STR003473	

連携を強化

<a href="#">Glycobiology, Volume 20, Issue 1, Pp.78</a> INT[Line:18]	GD3 (NeuAc[alpha]2-8NeuAc[alpha]2-3Gal[beta]1-4Glc[beta]1-1'Cer; CD60a), a b-series disialoganglioside, is known to be highly expressed in embryonic brains, but its concentration rapidly decreases after birth (Hgamukote 2007).
<a href="#">Glycobiology, Volume 21, Issue 9, Pp.1161</a> INT[Line:3]	Further, O-acetylated variants of the ganglioside GD3, namely 7-O-GD3 and 9-O-GD3 have been described to be expressed on T and B lymphocytes in various intensities (Kniep 1993; Fox 2001).
<a href="#">Glycobiology, Volume 20, Issue 1, Pp.78</a> ABS[Line:5]	In this study, we focused on GD3, a b-series ganglioside that is enriched in the immature brain and the subventricular zone (SVZ) of the postnatal and adult brain, and evaluated the usefulness of GD3 as a cell-surface biomarker for identifying NSCs.
<a href="#">Glycobiology, Volume 21, Issue 9, Pp.1161</a> ABS[Line:1]	GD3 (CD60a) and its 9-O-acetylated variant (CD60b) are intracellular regulators of apoptosis in T lymphocytes.
<a href="#">Glycobiology, Volume 21, Issue 9, Pp.1161</a> INT[Line:19]	We have recently identified the human CASD1 (Capsule Structure 1 Domain containing 1) gene encoding the Cas1 protein with an SGNH hydrolase and a hydrophobic transmembrane domain which may catalyze 7-O-acetylation of GD3 (Arming 2011).

阻害剤・構造

ターゲットの酵素名、EC番号、Ki(阻害定数)、IC50(50%阻害濃度)、阻害の様式、構造式、CAS番号、分子量

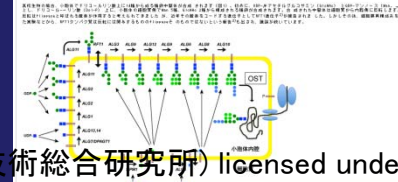
疾患・遺伝子

線虫

フェノタイプ

パスウェイ・遺伝子

GD3、CD60a、NeuAc[alpha]2-8NeuAc[alpha]2-3Gal[beta]1-4Glc[beta]1-1'Cerを同じものとして処理する。



# 糖鎖科学・実験プロトコル集

## GlycoPOD

産総研  
立命館大学

# 実験プロトコルデータベース~GlycoPOD~

立命館・産総研

**List of protocols**

- **Isolation & structural analysis of glycans**
  - Fluorescent labeling of glycans and HPLC analysis
  - Glycan Analysis by MS<sup>n</sup> Spectral Matching
  - Metabolic labeling of glycans by radioactive sugars
  - Permethylated for glycan analysis
  - Serial lectin affinity chromatography
- **N-Glycans**
  - Endo-β-galactosidase digestion
  - Endo-β-N-acetylglucosaminidase D digestion (Endo-D)
  - Endo-β-N-acetylglucosaminidase F digestion (Endo-F)
  - Endo-β-N-acetylglucosaminidase H digestion (Endo-H)
  - Endo-β-N-acetylglucosaminidase M digestion (Endo-M)
  - N-Glycanase digestion
  - Release of N-glycans by hydrazinolysis
- **O-Glycans**
  - Glycosidase digestion of O-glycoproteins and related O-glycans
  - Rapid and sensitive analysis of O-glycans using an in-line flow glycan-releasing system
- **Glycolipids and related compounds**
  - Assay of glucocerebrosidases using HPLC and fluorescent substrates



**Isolation & structural analysis of glycans**

- Fluorescent labeling of glycans and HPLC analysis

1.0 Shin-ichi Nakakita, Shunji Natsuka [Q&A 0](#) [Rating 0](#)

- Glycan Analysis by MS<sup>n</sup> Spectral Matching
- Metabolic labeling of glycans by radioactive sugars



**Fluorescent labeling of glycans and HPLC analysis**

Authors: Shin-ichi Nakakita, Shunji Natsuka (Ver 1.0) [Q&A](#) [Rating](#)

Introduction Protocol References

Category	Isolation & structural analysis of glycans
Protocol Name	Fluorescent labeling of glycans and HPLC analysis
Authors	Shin-ichi Nakakita* Department of Functional Glycomics, Life Science Research Center, Kagawa University Shunji Natsuka Department of biology, Faculty of science, Niigata University *To whom correspondence should be addressed
Keywords	Pyridamination Reducing amination

GlycoPODは、経験に依存する部分が多い糖質科学の実験において、糖鎖研究の初心者にもわかりやすい実験手法を公開している。

現在、156名の糖鎖科学研究者によって執筆された、15分野187の実験プロトコルが公開されている。平成24年度は約30プロトコルの執筆をお願いしている。そして、米国のグループも参加している。

各操作のノウハウや注意書きも追記できる。現在は、ユーザが執筆者とQ&Aについてシステム上でやり取りできるように改良中。

最終年度は、ニューテクノロジーを中心に30前後のプロトコルの追加に加えて、全てのプロトコルの見直しも行うことにしている。

# 糖鎖合成支援ツールの開発

# 糖鎖合成支援ツールの開発

野口研

産総研

有機化学による糖鎖合成の  
プロトコル・合成反応をDB化

糖質合成データベースワーキンググループ(CSDWG)によるデータ収集。野口研&産総研によるデータ入力。

東北大学・東海大学・東京工業大学・新潟薬科大学・岐阜大学・大阪大学

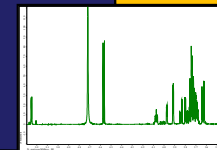
合成

Glycan Synthesis DB

- ・反応物・生成物などの構造検索
- ・合成経路検索
- ・ツールの開発

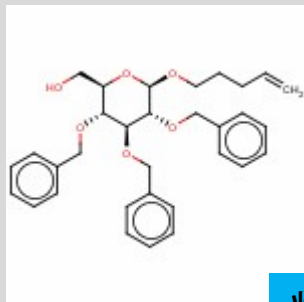
合成に必要な基質・生成物から糖鎖の骨格を抽出する機能の開発

確認



Glyco NMR-DB

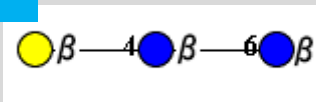
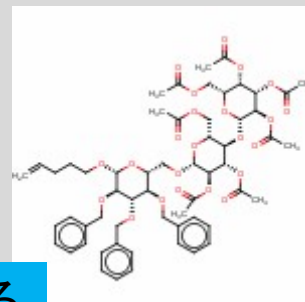
- ・構造のチェック
- ・NMRのスペクトル比較ツール
- ・実験条件の統一化されたデータ



変換ツール



糖鎖生物学者が理解できる  
CFGシンボル形式



合成できる構造が入手可能な研究リソースとなる

共同研究の促進

# GlycoNMRデータベース開発

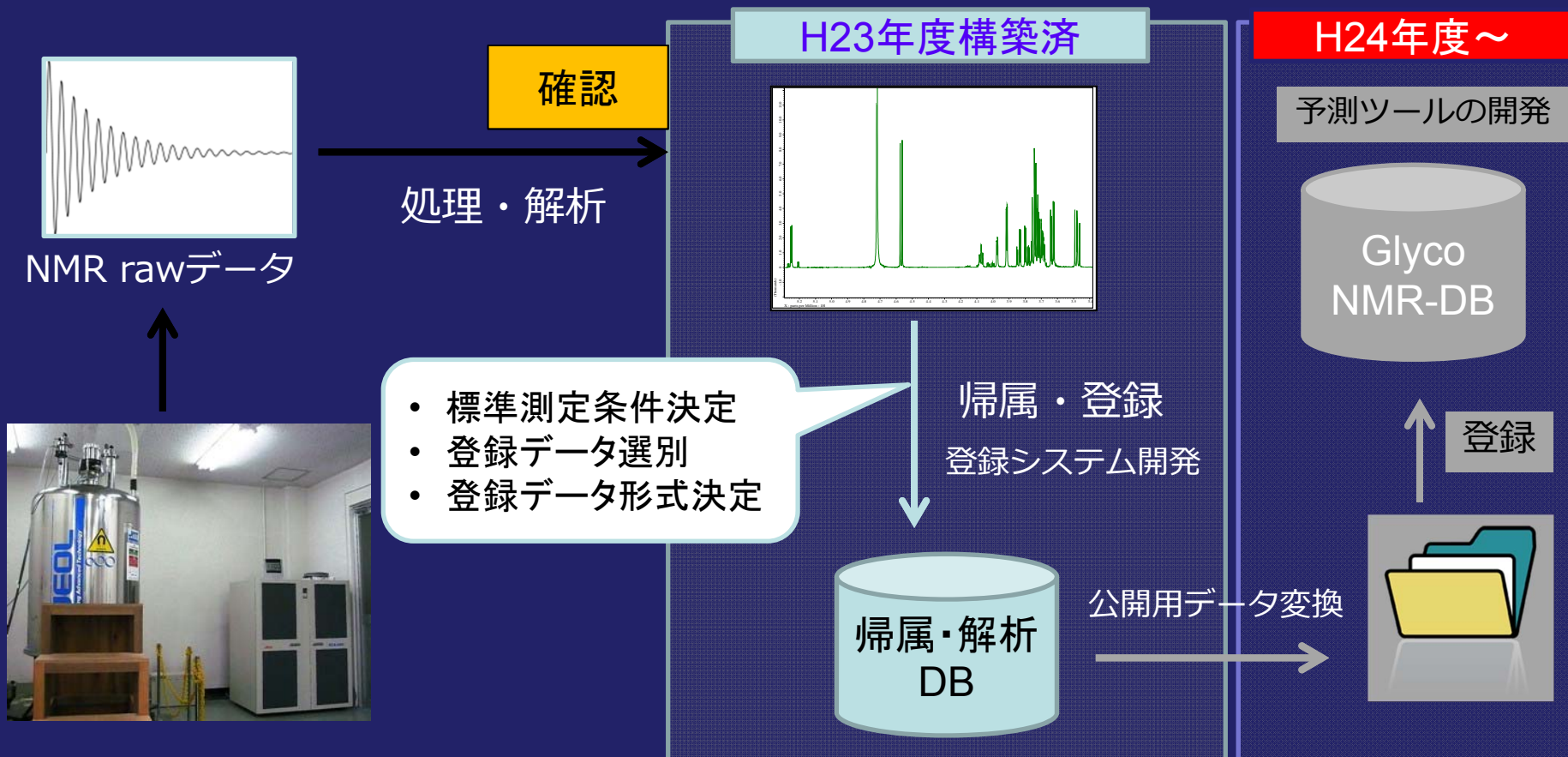
～糖鎖構造生物学・糖鎖合成化学～

## 糖鎖NMR（核磁気共鳴）データベースの特徴

- 高分解能(600MHz)
- 統一測定条件
- スペクトルの完全帰属

## 既存糖鎖NMR-DBの問題点

- 低分解能
- 測定条件が不均一
- 部分的な帰属データのみ





# GlycoNMR Database

野口研

糖鎖のNMRスペクトルは、糖鎖構造に於ける重要な情報源です。本データベースは、糖鎖研究において測定されたNMRスペクトルを解析し、分子構造、化学シフト、カップリング定数、スペクトル等をデータベース化して、閲覧・検索等ができるようにするものです。

NMRデータ: Galactose

### GlycoNMR Database

Spectra Search

検索条件

TOP Regist GlycoNMR ID Chemical Name Glycan Spectrum

GlycoNMR Spectrum ID: JCGG-SPC000001

Chemical Name: Partial match

Carbon/Hydrogen number of Monosaccharide: H1 C1

Monosaccharide Symbol: "Gal", "Man"

Glycan Type:

Linkage of Glycan:

Chemical Shift: all <sup>13</sup>C NMR <sup>1</sup>H NMR (ppm): range

Standard of chemical shift: DSS TSP TMS

Temperature (degree C): range

Solvent: [ or and ] D2O CDCl3 CD3OD DMSO-d6 DMF-d4 CD2Cl2

The maximum number you wish to display: 88 search Clear

検索結果

---

GlycoNMR Database

Search result

検索結果

entry	JCGG-SPM_ID	JCGG-SPC_ID	Compound Name	Nucleus	Temp.(degree C)	Solvent	Standard	Regist Date
1	JCGG-SPM000001	JCGG-SPC000001	D-galactopyranose	<b>1H</b>	30.0	D2O	DSS	2013-01-12 19.04.03.0
2	JCGG-SPM000002	JCGG-SPC000002	D-galactopyranose	<b>13C</b>	30.0	D2O	DSS	2013-01-12 05.14.33.0
3	JCGG-SPM000003	JCGG-SPC000003	D-galactopyranose	<b>COSY</b>	30.0	D2O	DSS	2013-01-12 05.14.33.0
4	JCGG-SPM000004	JCGG-SPC000004	D-galactopyranose	<b>HMQC</b>	30.0	D2O	DSS	2013-01-12 17.42.29.0

### GlycoNMR Database

D-galactopyranose [JCGG-SPC000001]

TOP Regist Search Download/Display

Structure ChemicalName Type Linkage Other Spectra Spectrum

ChemicalShift CouplingConstant δ & J Condition Comment All

Chemical Name

D-galactopyranose

D-galactose

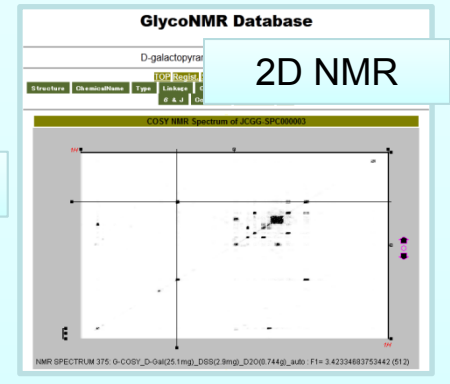
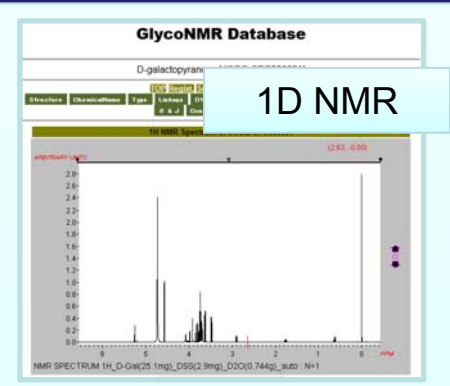
D-Gal

1 [Style][Label][With Atom Name]

2 [Style][Label][Position Label on Atom][Centered]

3 [Zoom][Zoom in]

化学構造



### GlycoNMR Database

D-galactopyranose [JCGG-SPC000001]

TOP Regist Search Download/Display

Structure ChemicalName Type Linkage Other Spectra Spectrum

ChemicalShift CouplingConstant δ & J Condition Comment All

Chemical Shift

Atom Number	Atom Name	Monosaccharide	Atom	Sugar Number	Chemical Shift (ppm)	Ambiguity Index
1	H05	D-Gal	b	H1	4.57	1
2	H27	D-Gal	b	H2	3.48	1
3	H29	D-Gal	b	H3	3.63	1
4	H31	D-Gal	b	H4	3.91	1
5	H38	D-Gal	a	H1	5.25	1
6	H39	D-Gal	a	H2	3.80	1
7	H41	D-Gal	a	H3	3.84	1
8	H43	D-Gal	a	H4	3.97	1
9	H45	D-Gal	a	H5	4.07	1

Coupling Constant

distance	pair(Atom Name, Sugar Number)	Coupling Constant (Hz)
3	H26-H27, H1-H2	7.9
3	H27-H29, H2-H3	10.0
3	H29-H31, H3-H4	3.4
3	H31-H33, H4-H5	1.0

数値データ

### GlycoNMR Database

D-galactopyranose [JCGG-SPC000001]

TOP Regist Search Download/Display

Structure ChemicalName Type Linkage Other Spectra Spectrum

ChemicalShift CouplingConstant δ & J Condition Comment All

Measurement Conditions

Parameter	Value
Machine Name	JEOL
Machine Console	JEOL
Operator	DELTA2, NMR
Measurement Time	1993
Workbook	1H
Number of Transients	1
Original F2 File Name	1H
Processing Software	SISSA
Phase Correction	single_phase.nc
Phase Adjustment	180
Integration	180
Integration Constant	180
Integration Scale	180
Integration Offset	180
Integration Gain	180
Integration Phase	180
Integration Mode	180
Integration Type	180
Integration Unit	180
Integration Value	180
Integration Weight	180
Integration X	180
Integration Y	180
Integration Z	180
Integration AA	180
Integration BB	180
Integration CC	180
Integration DD	180
Integration EE	180
Integration FF	180
Integration GG	180
Integration HH	180
Integration II	180
Integration JJ	180
Integration KK	180
Integration LL	180
Integration MM	180
Integration NN	180
Integration OO	180
Integration PP	180
Integration QQ	180
Integration RR	180
Integration SS	180
Integration TT	180
Integration UU	180
Integration VV	180
Integration WW	180
Integration XX	180
Integration YY	180
Integration ZZ	180

測定条件

# GlycoNMR Database

データ収集：野口研・理研の共同  
公開環境提供：産総研

登録管理システム：  
インターネットから測定・解析したNMRデータの登録・更新・削除などの管理を行うシステムです。

**GlycoNMR Database**

Please login

Administration  Registration

OpenID:

Please enter your OpenID.

If you have any one of the following IDs, you can login to GlycoNMR.  
Login with [Google ID](#), [Yahoo! JAPAN ID](#), [mixi ID](#), [livedoor ID](#)

[What is OpenID?](#)

For the first time login, you will be asked to register your name, nickname and E-mail with GlycoNMR.  
Only limited feature will be available until necessary information is registered.



**GlycoNMR Database**

[TOP](#) [Search](#)

Registration Menu

Please choose operation.

Menu	Link
Registration	<input type="button" value="Registration"/>
Update/Delete	<input type="button" value="Update/Delete"/>



OpenIDによるユーザ認証

登録画面

更新・削除画面

**GlycoNMR Database**

Registration Site

[TOP](#) [Registration](#) [Search](#)

GlycoNMR File

Data of Spectrum (JCAMP-DX or JPEG)

To send the result via email

XML形式スペクトルデータファイル

JCAMP-DX形式スペクトルデータ

**GlycoNMR Database**

Lists of Spectra

[TOP](#) [Registration](#) [Search](#)

JCGG-SPC\_ID

Registration Date 2018 Year | Month | Day (from) ~ 2018 Year | Month | Day (to)

Status  All data  Existing data  Deleted data

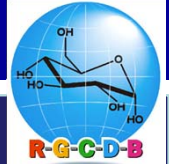
Entry	Delete	JCGG-SPC_ID	Update	Compound Name	Status	Registration Date
1	<input type="checkbox"/>	JCGG-SPC000001	<input type="button" value="UPDATE"/>	D-galactopyranose	Exist	2013-01-12
2	<input type="checkbox"/>	JCGG-SPC000002	<input type="button" value="UPDATE"/>	D-galactopyranose	Exist	2013-01-12
3	<input type="checkbox"/>	JCGG-SPC000003	<input type="button" value="UPDATE"/>	D-galactopyranose	Exist	2013-01-12
4	<input type="checkbox"/>	JCGG-SPC000004	<input type="button" value="UPDATE"/>	D-galactopyranose	Exist	2013-01-12

Delete the check data

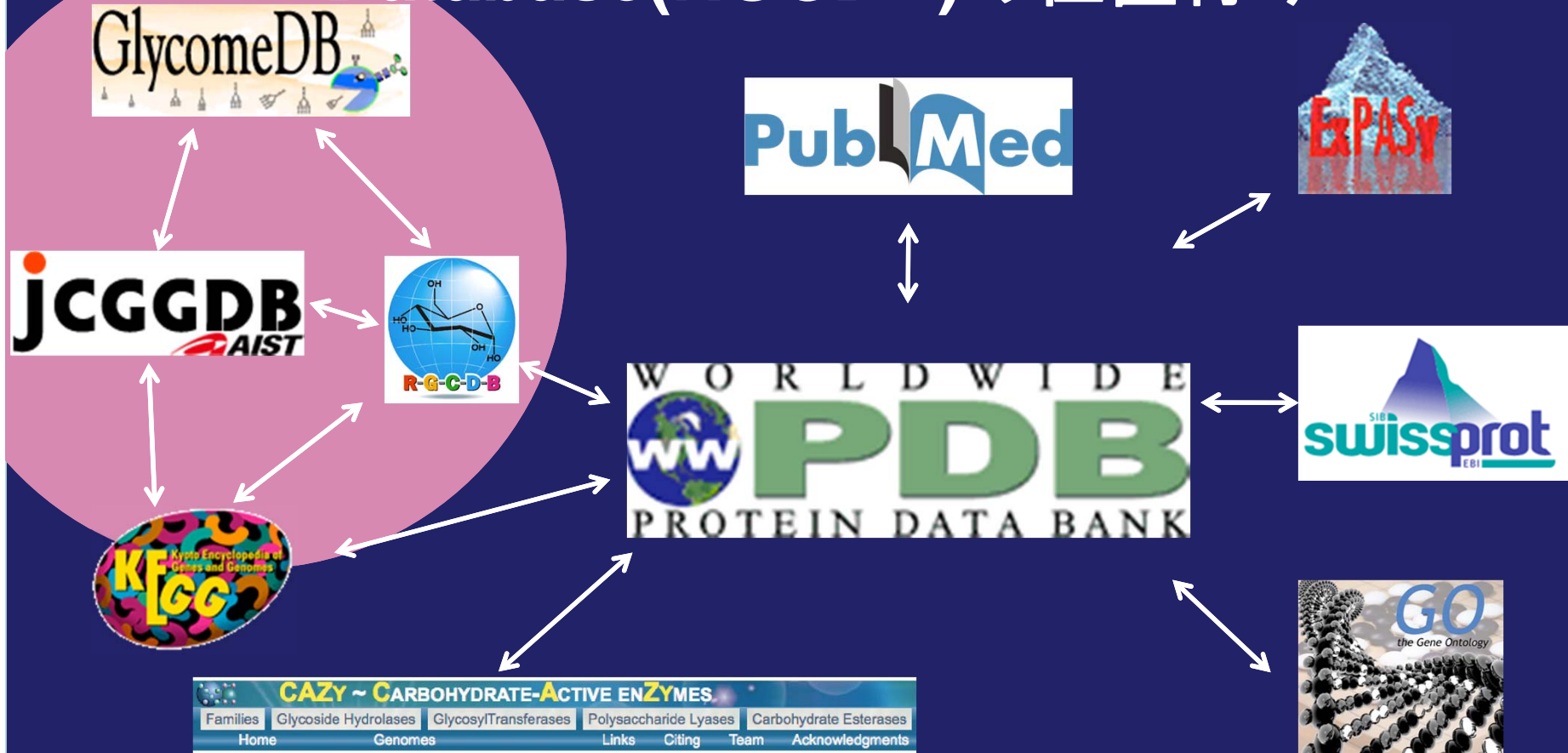
Delete the check data



# 糖鎖構造関連のPDBデータとの 連携強化



# Riken Glycan Conformation Database(RGCDB)の位置付け



PDBに登録されている糖鎖関連タンパク質についてタンパク質の情報に加えて糖鎖の情報について付加価値情報(糖鎖の構造、2面角等)を付け加えたものをWEBデータベースとして公開.

# 新規のWeb Interface

## タンパク質の検索画面

NO.	NAME	Sub Category	PDB ID
1	FC FRAGMENT OF HUMAN IgG1/UNRESOLVED PEPTIDE	nirvk	1DNZA
2	FC FRAGMENT OF HUMAN IgG1/UNRESOLVED PEPTIDE	nirvk	1DNZB
3	ISB-FC BOUND TO FCγ2R(α1/α2)	nirvk	1F6AA
4	ISB-FC BOUND TO FCγ2R(α1/α2)	nirvk	1F6AB
5	ISB-FC BOUND TO FCγ2R(α1/α2)	nirvk	1F6AD
6	FC FRAGMENT (IgG1 CLASS)	nirvk	1FC1A
7	FC FRAGMENT (IgG1 CLASS)	nirvk	1FC1B
8	IMMUNOGLOBULIN FC AND FRAGMENT B OF PROTEIN A COMPLEX	nirvk	1FC2D
9	FC RECEPTOR (INORMAL) COMPLEXED WITH FC (IgG)	nirvk	1FETA
10	FC RECEPTOR (INORMAL) COMPLEXED WITH FC (IgG)	nirvk	1FETC
11	MONOMAL FC RECEPTOR A2 BETA-2 MICROGLOBULIN/γ3 GAMMA-2A CHAIN C REGION/γ3 GAMMA-2A CHAIN C REGION	nirvk	1J1AA
12	MONOMAL FC RECEPTOR A2 BETA-2 MICROGLOBULIN/γ3 GAMMA-2A CHAIN C REGION/γ3 GAMMA-2A CHAIN C REGION	nirvk	1J1AC
13	MONOMAL FC RECEPTOR A2 BETA-2 MICROGLOBULIN/γ3 GAMMA-2A CHAIN C REGION/γ3 GAMMA-2A CHAIN C REGION	nirvk	1J1AD

インデックスページ

## 糖鎖の検索画面

## タンパク質の検索結果

## 糖鎖の検索結果

連携

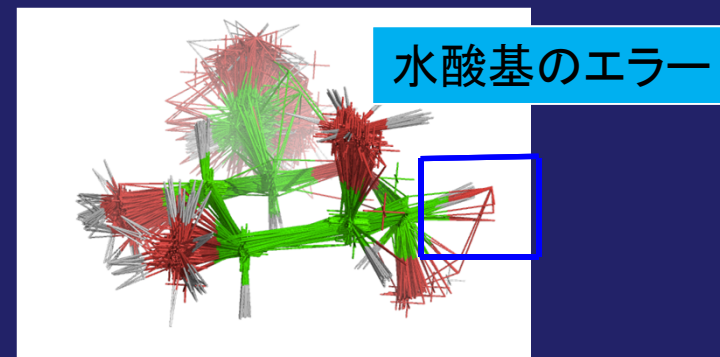
平成24年度公開予定

# PDBに置ける単糖・糖鎖の 再アノテーションプログラムの開発

## 3文字コードの曖昧さ

3 letter code	pdb name	glyco-CT	stereochemistry code	JMSDB
BGC	BETA-D-GLUCOSE	b-dglc-HEX-1:5		621212JMS0174
BMA	BETA-D-MANNOSE	b-dman-HEX-1:5		611121JMS0261
GAL	BETA-D-GALACTOSE	b-dgal-HEX-1:5		612111JMS0133
GLA	ALPHA-D-GALACTOSE	a-dgal-HEX-1:5		622111JMS0130
GLC	ALPHA-D-GLUCOSE	a-dglc-HEX-1:5		622121JMS0173
GUP	ALPHA-L-GULOPYRANOSIDE	a-lgul-HEX-1:5		622211JMS0225
GXL	ALPHA-L-GALACTOPYRANOSE	a-lgal-HEX-1:5		622111JMS0132
HSY	ALPHA-L-XYLOPYRANOSE	a-lxyl-PEN-1:5		622120JMS0352
LXC	BETA-L-XYLOPYRANOSE	b-lxyl-PEN-1:5		612120JMS0354
MAN	ALPHA-D-MANNOSE	a-dman-HEX-1:5		621121JMS0259
RIB	ALPHA-D-RIBOFURANOSE	a-drib-PEN-1:4		51112JMS0310
RIP	BETA-D-RIBOPYRANOSE	a-drib-PEN-1:5		621110JMS0314
SHD	ALPHA-D-ALTROPYRANOSE	a-dalt-HEX-1:5		612112JMS0082
XZY	BETA-D-XYLOFURANOSE	b-dxyl-PEN-1:4		52122JMS0349

## 単糖・糖鎖のエラーチェック



3文字コードで表されている単糖・糖鎖(1087個)を単糖に変換。

その後JMSDB(JCGGDB単糖データベース)のベースコードに変換⇒PDB中の単糖・糖鎖をJMSDBやJCGGDBの構造DBと連結が可能になる。

単糖のエラー ⇒ 単糖の立体構造から単糖をアノテーションするアルゴリズムの開発

グリコシド結合のエラー ⇒ glycomedbから得られた結合情報に照らし合わせてエラーの再アノテーション

グリコシド結合の欠落 ⇒ 距離・角度情報より結合情報の付加

再アノテーションプログラム(β版)をsourceforgeから公開する予定

# 予定

平成24年度

平成25年度

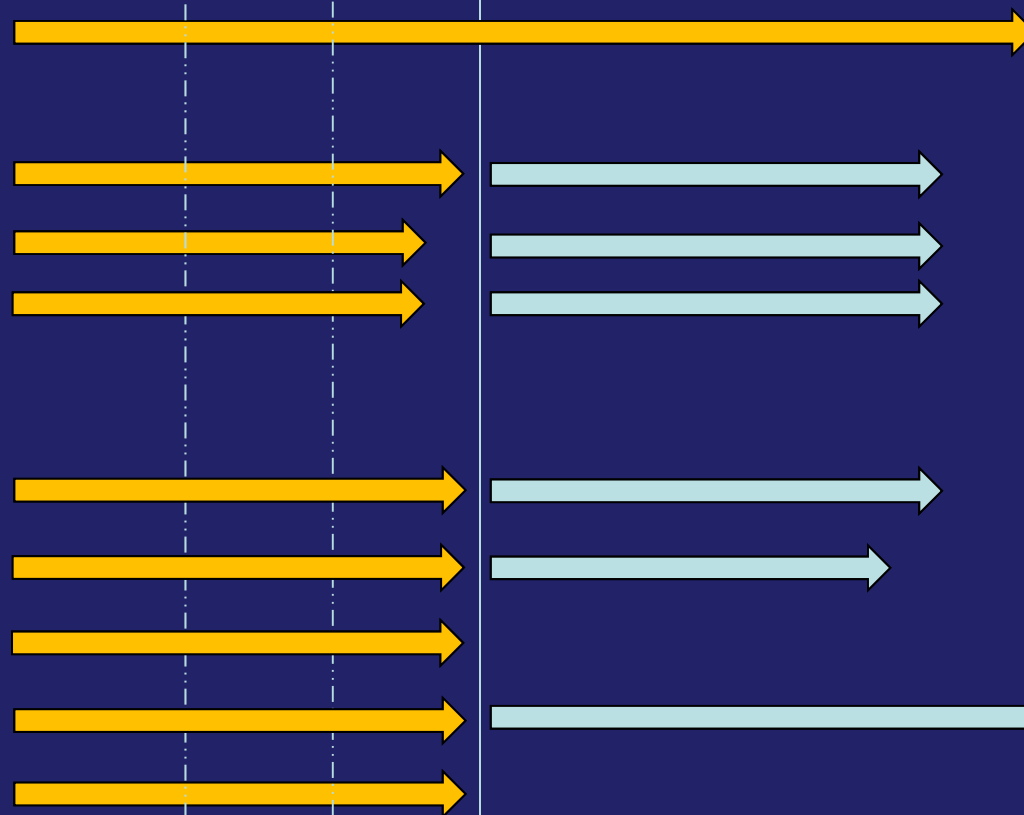
10 12

統合化インターフェース  
データ整備

配糖体  
用語・シノニム  
GlycoEpitope

共同開発

GlycoPOD  
線虫DB  
合成DB  
NMR-DB  
単糖表記の標準化



# (仮)外部連携と標準化

*National Institute of  
Advanced Industrial Science  
and Technology  
AIST*

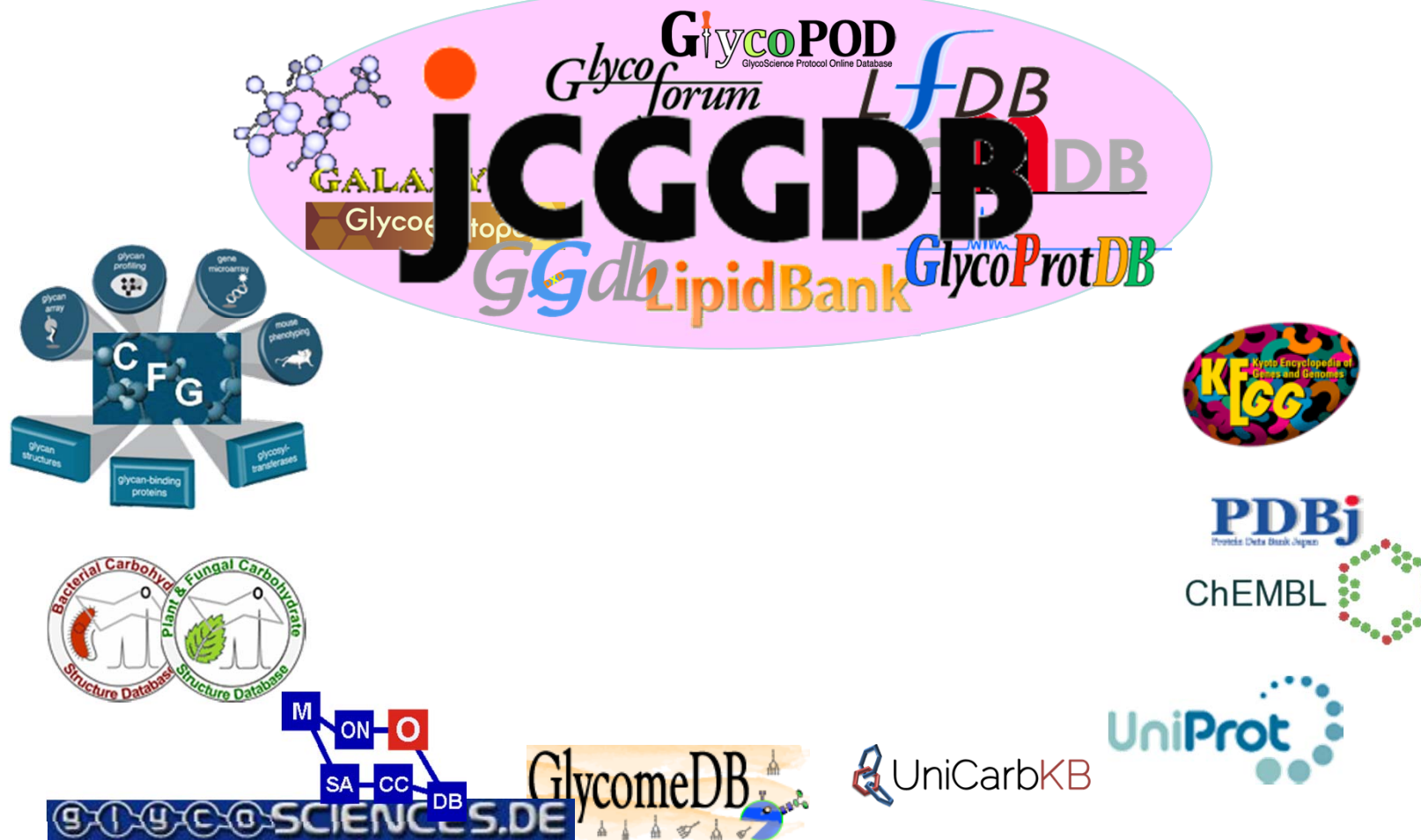


©2013 成松 久(産業技術総合研究所) licensed under CC表示2.1日本



# 糖鎖関連データベースの現状

—ハブデータベースの不在—



©2013 成松 久(産業技術総合研究所) licensed under CC表示2.1日本

# 相互接続しようとする...



# 糖鎖インフォマティクスの課題

- ハブデータベースの不在
- 標準的な表記法の不在

既存表記法の表現力不足

既存表記法では一意性を保証できない

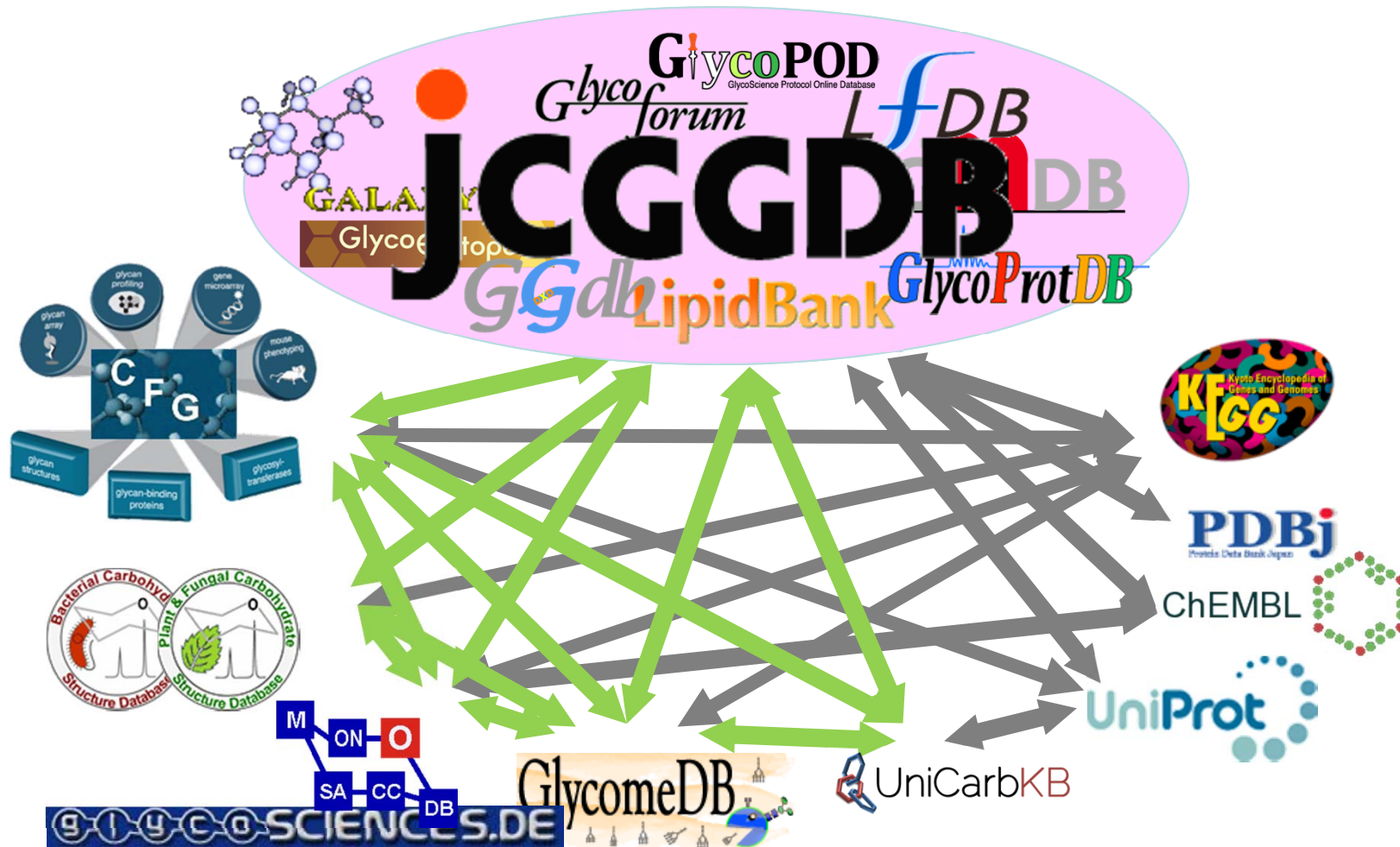
# 糖鎖インフォマティクスの課題

- ハブデータベースの不在  
→DB間の相互リンク
- 標準的な表記法の不在

既存表記法の表現力不足

既存表記法では一意性を保証できない

# 相互接続のためのRDF化と共同開発



# 糖鎖インフォマティクスの課題

- ハブデータベースの不在
- 標準的な表記法の不在

既存表記法の表現力不足

既存表記法では一意性を保証できない

糖鎖インフォマティクスの課題  
 – 糖鎖構造情報表記の不統一  
 例：単糖のさまざまな表記法

Scheme:	Name:	Separate Substituents:
MonosaccharideDB	b-dglc-HEX-1:5  (2d:1)n-acetyl	
CarbBank	b-D-GlcpNAc	
Glycosciences.de	b-D-GlcpNAc	
GlycoCT	b-dglc-HEX-1:5	•(2d-1) n-acetyl
CFG	GNb	
BCSDB	bDGlcpN	•(2-1) Ac
GLYCAM	0YB	
Protein Data Bank	NAG	
IUPAC	2-acetamido-2-deoxy-beta-D-Glucopyranose	

<http://monosaccharidedb.org>



©2013 成松 久(産業技術総合研究所) licensed under CC表示2.1日本

# 糖鎖インフォマティクスの課題

- ハブデータベースの不在
- 標準的な表記法の不在

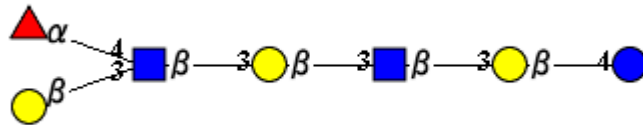
既存表記法の表現力不足

既存表記法では一意性を保証できない



# 糖鎖インフォマティクスの課題

## —不完全な糖鎖構造情報—



JCGG-STR025670

構造解析が不完全でも学術論文は受理されてしまう現状。  
曖昧な糖鎖構造を既存の表記法では表現しきれない場合がある。



©2013 成松 久(産業技術総合研究所) licensed under CC表示2.1日本

# 糖鎖インフォーマティクスの課題

- ハブデータベースの不在
- 標準的な表記法の不在

既存表記法の表現力不足

既存表記法では一意性を保証できない

*Semantic Web*に対応できる  
糖鎖表記法の開発

# 糖鎖構造情報表記の標準化

新しい糖鎖表記法の開発

WURCS:

Web3 Unique Representation of  
Carbohydrate Structures

特徴:

あいまいな部分構造を含む糖鎖構造情報も一意に表記できる。

# 糖鎖情報表記の標準化

## 糖鎖標準表記法の開発 WURCS

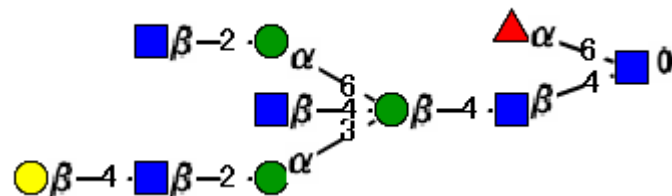
“2-acetamido-2-deoxy-β-D-Glucopyranose”などの単糖名を基にした文字列を利用して糖を表記した場合、同じ構造の糖でも表記が異なる問題がある。(α-L-Fuc vs. α-L-6-deoxy-Gal)  
 そこで、曖昧な糖鎖構造を同じ文字列で表記できる糖鎖標準表記法 (WURCS)を開発している。



Monosaccharide:  
 β-D-Galp  
 β-D-GlcpNAc  
 ?-D-GlcpNAc  
 α-L-Fucp or α-L-6-deoxy-Galp  
 α-D-Manp  
 β-D-Manp



WURCS: Monosaccharide  
 12112h-1:5  
 12122h-1:5||2:2\*NC(C)=O  
 x2122h-1:5||2:2\*NC(C)=O  
 11221m-1:5  
 21122h-1:5  
 11122h-1:5



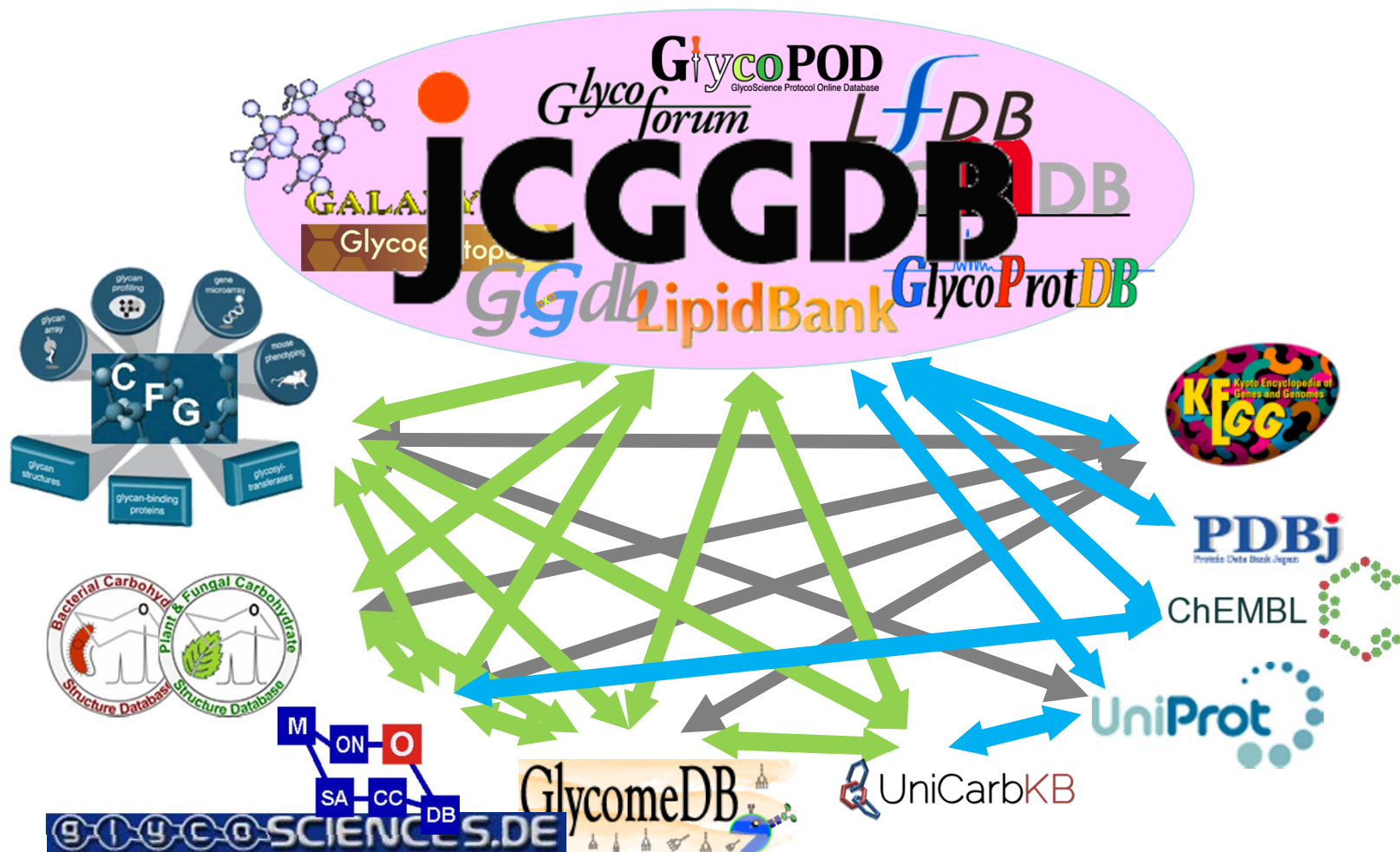
<http://jcggdb.jp/idb/jcggdb/JCGG-STR001412>

WURCS=1/[12112h-1:5]14[12122h-1:5||2:2\*NC(C)=O]12[21122h-1:5]13<[12122h-1:5||2:2\*NC(C)=O]14)([12122h-1:5||2:2\*NC(C)=O]12[21122h-1:5]16)[11122h-1:5]14[12122h-1:5||2:2\*NC(C)=O]14([11221m-1:5]16)[x2122h-1:5||2:2\*NC(C)=O]

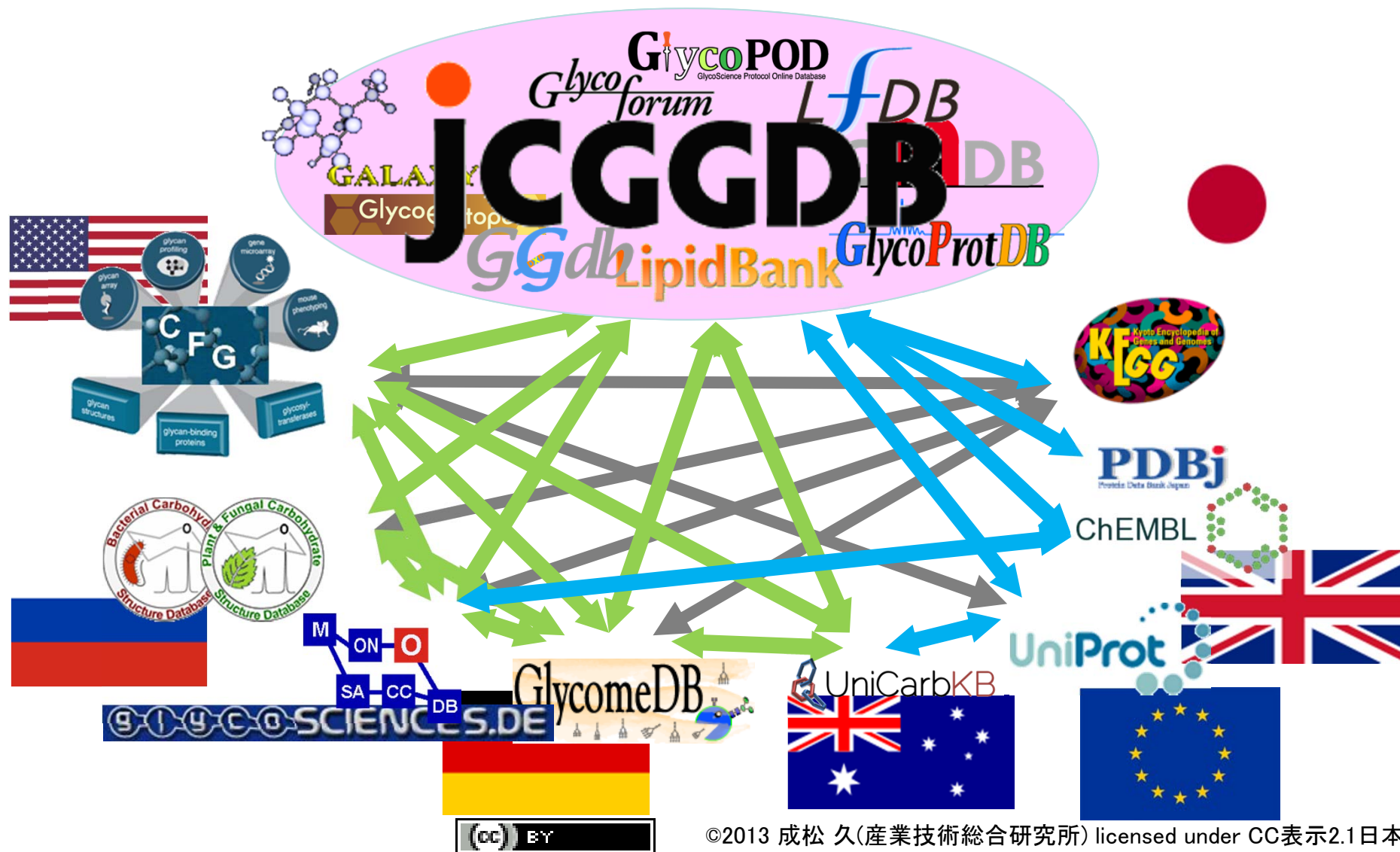
# 糖鎖構造情報表記の標準化

WURCSによる糖鎖構造表記を入出力や変換できるツールを開発して公開提供し、糖鎖表記の標準仕様とする。

# 周辺分野との連携



# RCMGから呼びかけで国際連携体制を形成



# Warren Workshop 2012

- August 8-12, 2012
- Complex Carbohydrate Research Center (CCRC)
- University of Georgia
- Athens, GA, USA



# Discussion (led by Will York)

## glyspace Registry

- Everybody can register to submit new entries (structure/Composition) and retrieve an ID for his structure
- Free and easy access with a wide variety of searches
  - By mass
  - By composition
  - By substructure
  - By Id in other resources (databases)
  - ...
- Allows structures to be entered at an appropriate level of structure fuzziness (from composition to the fully defined structure) and to explore related structures at different levels.

# 国際コミュニティ形成と共同開発

## ACGG-DB3 Okinawa Meeting



RCMGから国内外の糖鎖科学研究者に呼びかけて、糖鎖科学データの標準化へ向けたキックオフミーティングを主催した。日米欧だけでなく東アジア諸国も含めての国際合意を形成し、グローバルな連携協力体制での糖鎖科学分野のSemantic Web技術対応の基盤開発を始めることができた。本会議のレポートはGlycobiology誌からの依頼で寄稿し掲載された。

(PMID:23271684)



BioHackathon2012Toyama  
 主要な糖鎖DB (GlycomeDB, BCSDDB, MonosaccharideDB, UniCarbKB)ホルダーを招聘してRDF化を完了。

GlycoProtDBとUniProtの連携



# 連携体制と開発成果(まとめ)

- 標準化と基盤整備

  - 内外の関連データベースのRDF化

    - BioHackathon12で各DBホルダーを招待して開発

    - 糖鎖構造DB

      - GlycomeDB(米国) BCSDDB(ロシア)

      - MonosaccharideDB(ドイツ)

      - UniCarbKB(豪州)

    - 糖鎖関連DB

      - GlycoProtDB(産総研)・・・UniProtと連携

      - GlycoEpitope(立命館大学)

    - オントロジー

      - DBCLSほか・・・海外DBとも連携

    - 標準表記法

      - 野口研、創価大、産総研・・・海外DBとも連携

    - ツール開発

      - WURCS関連(野口研、創価大、産総研)

      - PDB関連(理研)

- 標準化のための活動

  - コミュニティ形成

    - ACGG-DB3会議(沖縄、4月)主催・・・キックオフミーティング

    - BH12で糖鎖グループを組織

    - ACGG-DB4(濟州島、10月)

    - 国外会議への参加

      - Warren Workshop(米国)など

# 2013年度連携関連予定

## 標準化と基盤整備

WURCS関連ツールを内外の連携先に提供  
PDB糖鎖データ再アノテーションツールの公開  
Semantic Web関連サービスの提供  
既存関連DBのRDF化継続  
糖鎖Ontology整備継続

## 標準化のための活動

糖鎖データ収集の国際標準化検討  
ACGG-DB5会議(5月中旬予定)、ACGG-DB6(下期)  
Beilstein Symposium(Potsdam、6月)  
共同ツール開発  
BioHackathon13(6月)に参加