

# NBDC統合化推進プログラム

「糖鎖統合データベースおよび  
国際糖鎖構造リポジトリの開発」

成松 久

平成27年8月4日 サイトビジット資料

 © 2015成松 久 (理化学研究所) licensed under CC表示2.1日本

# 「糖鎖統合データベースおよび国際糖鎖構造リポジトリの開発」 達成目標

1. **国際糖鎖構造データリポジトリシステムを開発する**
2. **全糖鎖構造データの標準化を行う**
3. **糖鎖関連データベースの標準化ならびにセマンティックウェブ対応によって、ライフサイエンスデータベース全般との統合を目指す**
  - a. **新しい実験解析技術によって生産されるデータに対応するために新規開発するデータベースも含める(糖ペプチド、レクチンマイクロアレイなど)**

# 1. 国際糖鎖構造リポジトリの開発

糖鎖構造データの登録システムの設計及び開発:

- ・GlySpaceと連携し、完了した
- ・論文をNARデータベース特集号に投稿する予定

**GlyTouCan**  
THE GLYCAN REPOSITORY

<b>40072</b> Glycans	<b>61</b> Motifs	<b>793</b> Monosaccharides
-------------------------	---------------------	-------------------------------

## 2. 全糖鎖構造データの標準化を行う

- WURCSバージョン2の開発
- WURCS-RDFの開発
- GlyTouCanへの導入

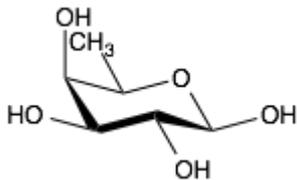
→ 全て完了

現在WURCSバージョン2の論文を執筆中

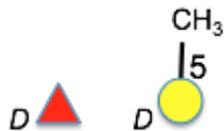
# 糖鎖構造データ標準化の必要性

## 単糖の表記の曖昧さ など

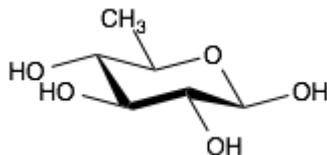
D-Fucose  
6-Deoxy-D-galactose



$\beta$ -D-Fucp



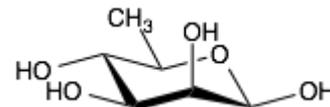
D-Quinovose  
6-Deoxy-D-glucose



$\beta$ -D-Quip



D-Rhamnose  
6-Deoxy-D-mannose



$\beta$ -D-Rhap



# WURCS:

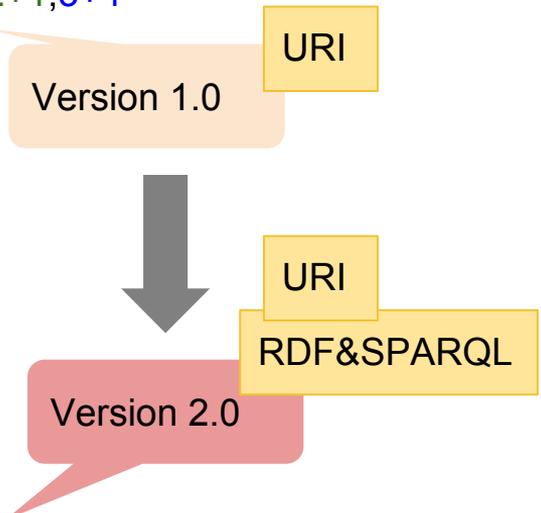
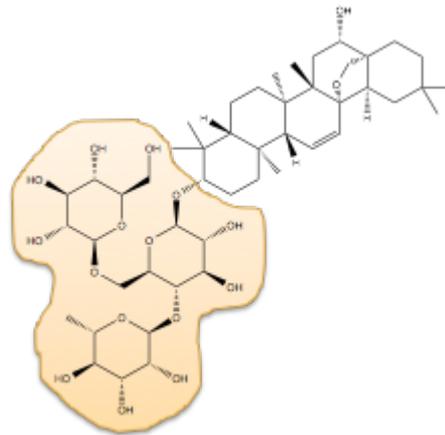
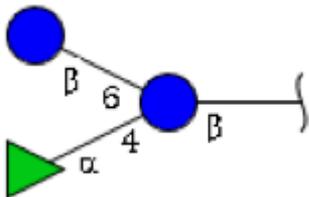
# 曖昧さを排除した表記法

## Web3 Unique Representation of Carbohydrate Structures

*J. Chem. Inf. Model.* 2014, 54, 1558–1566.

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

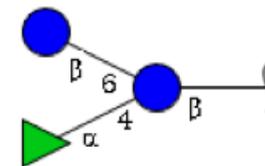
G08841ZJ



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

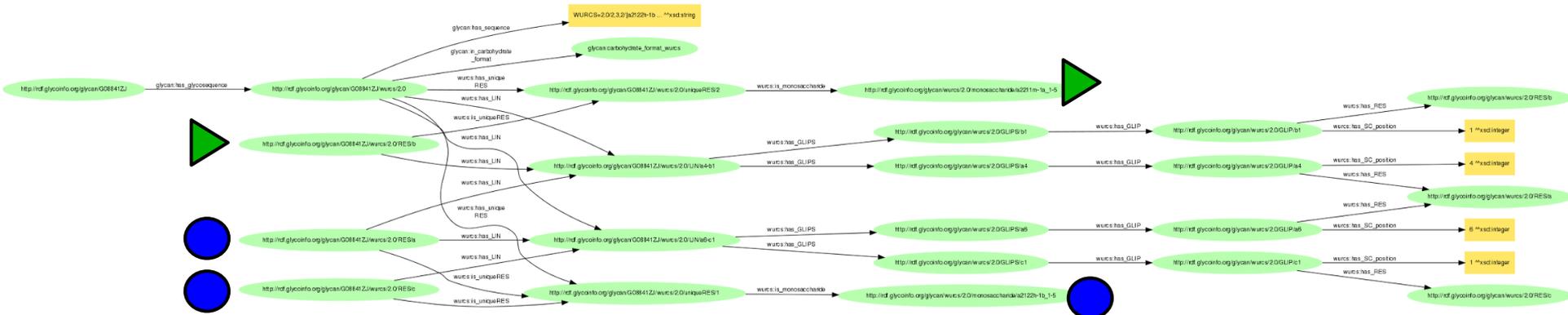
# WURCS-RDF

G08841ZJ



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

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

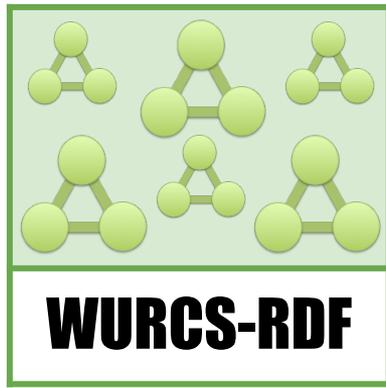


# WURCS-RDF を用いた糖鎖構造検索

**WURCS** WURCS=2.0/... 文字列

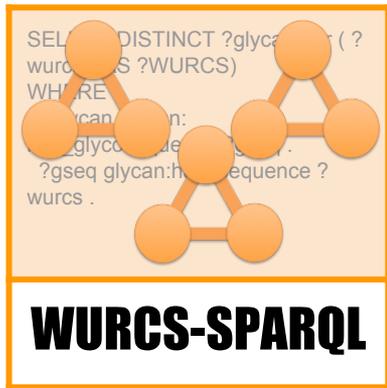
**WURCS to WURCS-RDF**

**WURCS to SPARQL Query**



**WURCS-RDF**

約260万トリプル

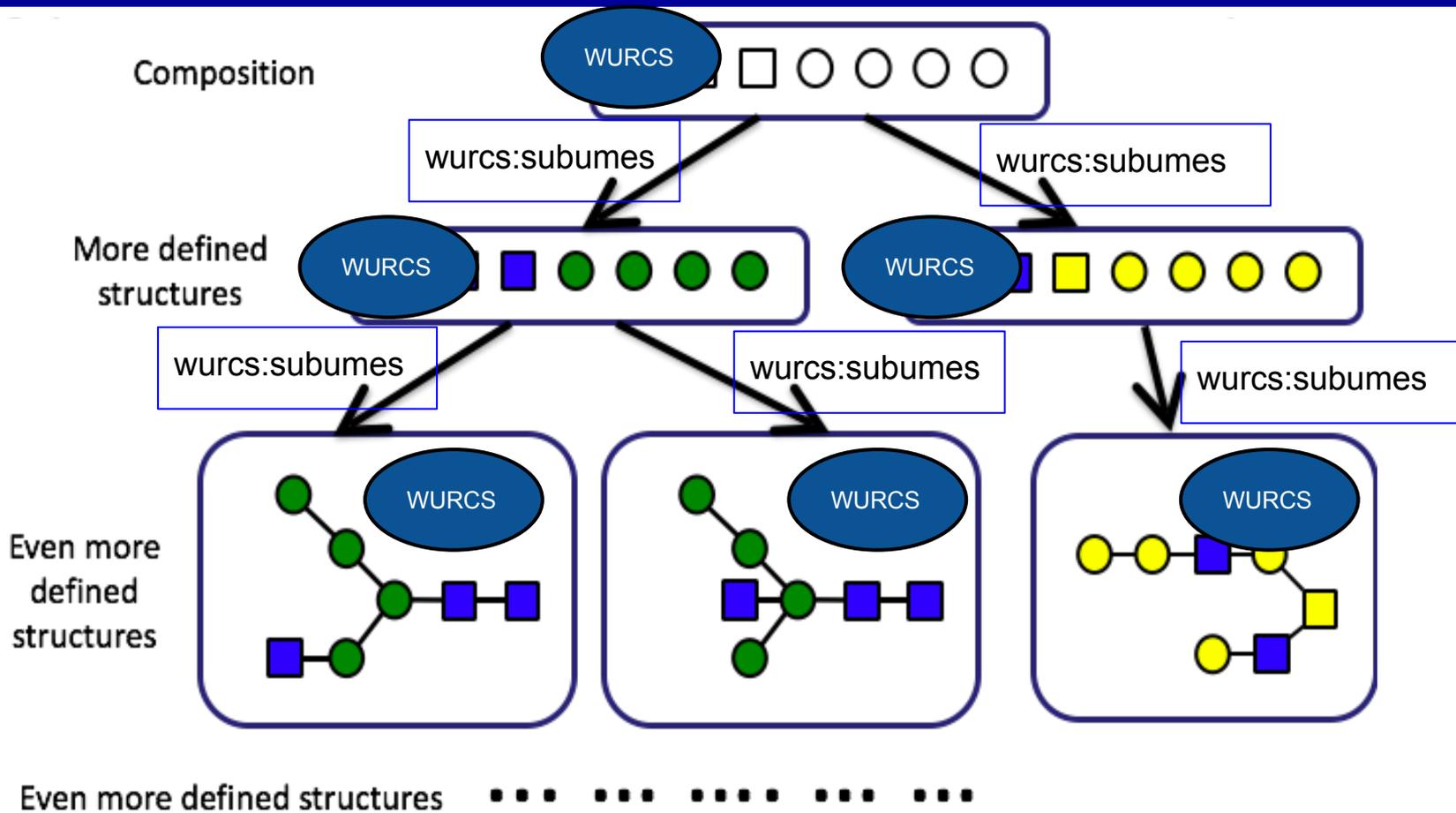


```

SELECT DISTINCT ?glycan (?
wurcs ?WURCS)
WHERE
  glycan:
glycan:
?gseq glycan:hasSequence ?
wurcs .
    
```

**WURCS-SPARQL**





### 3a. JCGGDBのセマンティックウェブ化版:ACGGDB

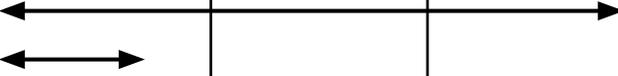
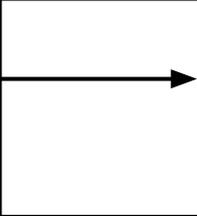
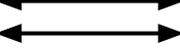
	オントロジー開発	RDF化	スタンザ(UI)開発
<b>GlycoProtDB: 糖タンパク質の糖鎖結合部位情報(更新システム)</b>	完了	完了	完了
<b>LfDB: レクチンの糖鎖結合親和性データ(更新システム)</b>	完了	完了	開発中
<b>GDGDB: Glyco- Disease Genes DB</b>	完了	完了	完了
<b>PacDB: Pathogen adherence DB</b>	完了	完了	完了
<b>GSSGDB: GlycoSite-Specific GlycoformDB: 糖タンパク質の糖鎖組成</b>	開発中	開発中	—
<b>GGDB: GlycoGene DB</b>	開発中	—	—

## 3b. GlycoEpitope、GlycoPODの開発

- GlycoEpitopeの開発
  - 糖鎖構造情報を整理し、WURCS形式で表記できるようにし、GlyToucanに登録した
  - その際に、GlycoEpitopeに手動で入力されたデータの確認が行われたため、データを改善することもできた
  - 運用システムをver. 2にアップグレードし、操作性を大幅に改善した
  - 現在、エピトープ173項目、抗体613項目を収載し公開している
- GlycoPODの開発
  - 公開中プロトコルの書式統一を行うために、再校正を進め、ほぼ完了した
  - 本年度新たに6項目の実験手法を追加収載し、合計で213項目の実験方法を公開している

※オントロジー開発を検討したところ、フリーテキストが多いため、保留中

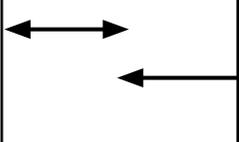
# 研究開発計画(創価大・野口研・新潟大)

研究開発項目	H26年度	H27年度	H28年度
1. 国際糖鎖構造リポジトリ開発(創価大) ・サーバー設置と運用 ・糖鎖構造データの登録システムの設計開発 ・検索閲覧システムの開発			
2. 糖鎖構造データの標準化とツール開発 ・WURCSの拡張(野口研・創価大学) ・糖鎖構造データのWURCS変換(野口研・創価大学) ・WURCS解析ツールの開発(野口研・創価大学) ・糖鎖構造相互変換ツールの開発(野口研・創価大学) ・糖鎖構造のRDFデータモデルとオントロジーの開発(野口研・創価大学)			
3. GlycoEpitopeデータベースのセマンティックウェブ化(新潟大・立命館大) ・GlycoEpitopeオントロジーの開発 ・GlycoEpitopeのRDF化 ・GlycoEpitopeのインターフェイス開発			

# 進捗状況(創価大・野口研・新潟大)

研究開発項目	H26年度	H27年度	H28年度
1. 国際糖鎖構造リポジトリ開発(創価大) <ul style="list-style-type: none"> <li>・サーバー設置と運用</li> <li>・糖鎖構造データの登録システムの設計開発</li> <li>・検索閲覧システムの開発</li> </ul>			
2. 糖鎖構造データの標準化とツール開発 <ul style="list-style-type: none"> <li>・WURCSの拡張(野口研・創価大学)</li> <li>・糖鎖構造データのWURCS変換(野口研・創価大学)</li> <li>・WURCS解析ツールの開発(野口研・創価大学)</li> <li>・糖鎖構造相互変換ツールの開発(野口研・創価大学)</li> <li>・糖鎖構造のRDFデータモデルとオントロジーの開発(野口研・創価大学)</li> </ul>	WURCS1.0 → WURCS2.0 39964 WURCS	WURCSFramework WURCSStoGlycoCT 前倒しで着手(WURCS-RDF)	
3. GlycoEpitopeデータベースのセマンティックウェブ化(新潟大・立命館大) <ul style="list-style-type: none"> <li>・GlycoEpitopeオントロジーの開発</li> <li>・GlycoEpitopeのRDF化</li> <li>・GlycoEpitopeのインターフェイス開発</li> </ul>	Ver.2へのアップグレード完了 2015.6.10完了し、GlyTouCanに連携済み	スタンザ開発	

# 研究開発計画(産総研・立命館)

研究開発項目	H26年度	H27年度	H28年度
4. 糖鎖科学統合DBのオントロジー開発とRDF化及び糖鎖発現プロファイル統合データベースの新規開発(産総研) <ul style="list-style-type: none"> <li>・GlycoProtDBのフルセマンティックウェブ化</li> <li>・糖タンパク質およびレクチンデータベースの開発</li> <li>・既存の糖鎖データベースのセマンティックウェブ化</li> </ul>			
5. GlycoPODデータベースのセマンティックウェブ化(産総研・立命館大) <ul style="list-style-type: none"> <li>・GlycoPODオントロジーの開発</li> <li>・GlycoPODのRDF化</li> <li>・GlycoPODのインターフェース開発</li> </ul>			

研究開発項目	H26年度	H27年度	H28年度
4. 糖鎖科学統合DBのオントロジー開発とRDF化及び糖鎖発現プロファイル統合データベースの新規開発(産総研) <ul style="list-style-type: none"> <li>・GlycoProtDBのフルセマンティックウェブ化(鈴木)</li> <li>・糖タンパク質およびレクチンデータベースの開発                             <ul style="list-style-type: none"> <li>- GlycoForm(鹿内)</li> <li>- LfDB(鈴木)</li> </ul> </li> <li>・既存の糖鎖データベースのセマンティックウェブ化                             <ul style="list-style-type: none"> <li>- GDGDB, PacDB(イエレナ、藤田)</li> <li>- GGDB(鈴木)</li> </ul> </li> </ul>	予定通り完了 	今年度中に完了予定 更新システム・スタンザほぼ完了  	オントロジー、RDF、スタンザ化完了 データ更新中・オントロジー開発中  
5. GlycoPODデータベースのセマンティックウェブ化(産総研・立命館大) <ul style="list-style-type: none"> <li>・GlycoPODオントロジーの開発</li> <li>・GlycoPODのRDF化</li> <li>・GlycoPODのインターフェース開発</li> </ul>	書式統一、DBの改良 	新しいプロトコルの追加 MIRAGEとの連携を検討 	

# 国際連携

- GlyTouCan
  - 米国Glyspaceプロジェクト成果活用
  - 多言語化
    - 日本語、中国語、ロシア語、フランス語、ドイツ語
  - 海外の糖鎖関連RDF化データの統合
    - GlycomeDB, UnicarbKB, MonosaccharideDB, BCSDDB
- 普及標準化
  - MIRAGE guidelinesにおいて採用
    - GlyTouCanアクセッション番号
    - WURCS
- アジアとの連携
  - ACGG-DB公開 (<http://acgg.asia/db/>)
  - ACGG国際会議主催予定 (2015.11.12-15 宮城県)

# 研究グループ会議

第1回DB会議: 2014.4.10 @つくば

～

第26回DB会議: 2015.7.16@DBCCLS 柏の葉

次回開催: 2015.8.20(予定)