

糖鎖統合データベースと研究支援ツールの開発

(代表)産業技術総合研究所
立命館大学
野口研究所
理化学研究所

2012年2月24日

糖鎖統合グループの計画

統合化のための情報基盤整備

23

DB集約・データ収集・API

24

RDF化のための準備
URIの準備

アジア連携・統合化

25

欧米との連携・統合化のための準備・話し合い

ポイント

- 国内の統合
- データベース拡充
- ウェットの研究者のためのデータベース構築
- 統合化のための準備
- 国際連携(アジア・欧州・米国)

概要

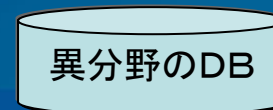


JCGGDBの機能強化と拡張

連携強化



1~3年目



異分野のDB

DBCLS
RDF化
センマンティックの利用



1~3年目

合成物の確認
(研究支援)

今までの文科省・統合DBプロジェクトの

継続

- DBとサーバの管理
- JCGGDBのデータ更新
- 新規DBの構築

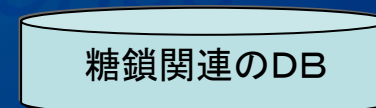
拡張

- DB間連携・有機的統合
- 研究者主導型の蓄積型DBにするための改良
- webやAPIの高機能化
- DBの知名度向上

研究支援ツールの構築

統合化支援

更なる新規&既存DBの統合(国内)
未公開データをDB化支援(国内)
海外DBとの連携



糖鎖関連のDB

1~3年目

(一例として)

糖鎖NMRDB

- 統一した条件下で糖鎖のNMRシグナル



NMRシグナルの構造への帰属
ツールの開発など

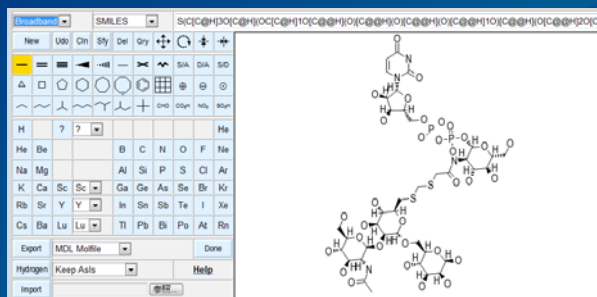
2~3年目

データ拡充



Chemistry Informatics (ChemInfo) の分野で利用されている立体構造を表記するデータ形式 (SDF, SMILES, InChI, PDB など) を利用して糖鎖構造を検索可能

ステレオケミストリを
区別して検索できる



多彩な機能

- 1) 一致検索
- 2) 部分構造検索
- 3) Super Structure
- 4) 類似検索



配糖体DB
7万構造
機能の情報

新規構築

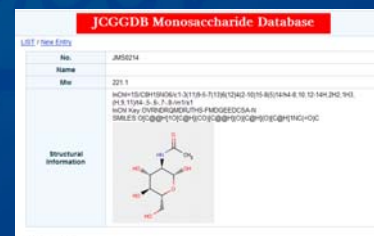
糖鎖合成支援システム

合成化合物DB: 3975構造
反応経路DB: 反応数: 887件

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

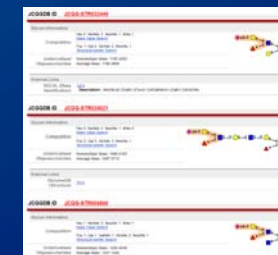
データ拡充

単糖のDB
890構造



データ拡充

糖鎖構造DB
6248構造

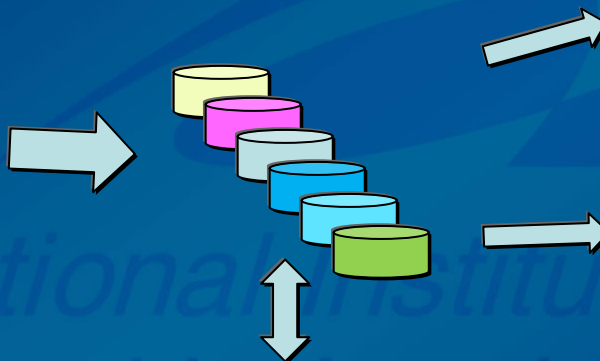
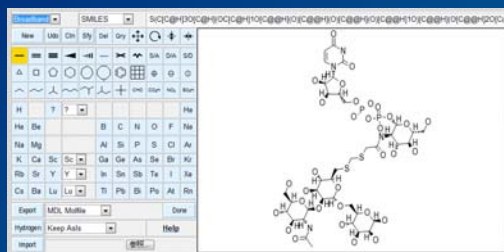


平成24年度は
URIの開発を行う

阻害剤DB 286構造
産総研のGGDB
グリコシダーゼDBと連携

化学構造データ・管理用コマンドライブラリの開発

構造検索するためのコマンドラインライブラリを開発
 ・登録・編集・削除



コマンドでメンテナンス



阻害剤DB
286構造



単糖のDB
890構造



糖鎖構造DB
6248構造

配糖体
7万件

糖鎖合成支援システム
 合成化合物DB:3975構造
 反応経路DB:反応数:887件

データ拡充

レクチンデータベース

レクチンと糖鎖の相互作用

LFDDB Lectin frontier DataBase

Advanced Search

Keyword:

Family Classification:

Select Item

- Lectin Family
- Monosaccharide Specificity
 - GalNAc(2)
 - GlcNAc(2)
 - Mann(4)
 - Gal(2)
 - MannP(2)
 - Mann(3)
 - Other(2)
- IC-50

AAL (LFF-00003f)

Lectin ID	LFF-00003f
Lectin Name	AAL
Lectin Family	Fucose-binding Lectin
Monosaccharide Specificity	Fuc
Organism (Scientific Name)	orange peel mushroom (<i>Neuria auranti</i>)
Japanese Name (Family)	トウモロコシ
Source	Fungi
Source Organ	
Specific Source Organ	Fuc
Number of CRD	
3D-fold	
Accession (External Link)	GenBank: D85776, D09573, E02022 PDB: 1UJL, 1UJC, 1OZF, 1Plm
Sequence ID	
Sequence	1 apse1pstaK iaklmeasg gqevyfaa lqgklnaap gqgqgqga agqgqkaal 41 lqglavnaak aqgqgqgq gmaaklme lqvqgqgq gqgqgqgq gmaaklme 121 vqgqgqgq lqvqgqgq gqgqgqgq agqgqkaap gqgqgqgq ataklqgqgq 241 kvqgqgqg lqvqgqgq gqgqgqgq agqgqkaap gqgqgqgq gqgqgqgq 301 lqvqgqgq apv

References
J Biol Chem. 278(29) 2003 Jul 18; 27959-27967 / Biochemistry. v64(2, No 18, 2003 Sep 30) 11993-11999

糖タンパク質のデータベース

マウスの各組織由来の糖タンパク質を
数種類のレクチンで捕集し解析したデータ

Collectrin (GLP-00000d)

General Information

Glycoprotein ID	Protein Name	Organism	Length(aa)	MW(kDa)
GLP-00000d	Collectrin	Mus musculus	222	29070

Sequence / motifs

```

1  HQALPLPLVY TQAAQVPLVY AEAALPPLVLA TQAAQVPLVY TQAAQVPLVY
41  TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY
141  TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY
181  TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY TQAAQVPLVY
    
```

Glycosylation Site

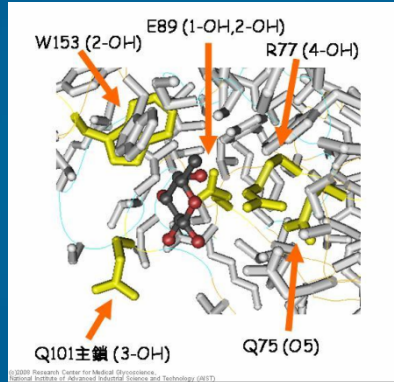
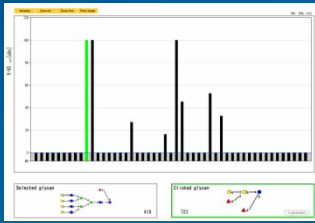
Organism	Lectin	Glycosylation Site
Organism	Mus musculus	ALL

External Link

GenBank	RefSeq	MD	Protein Link	Nucleotide Link	Swiss prot	PIR-PSD	IT
U0101162	NP_205051	M01162234	A455006.1 A458912.1 B462022.1	A455006.1 B462022.1	Q8E204		PIR011819.1099

Related Protein (homolog)

Organism (Tax_ID)	Protein Name
Homo sapiens(9606)	transmembrane protein 270 (TP_065716.1)
Canis familiaris(9615)	similar to transmembrane protein 20 (TP_054205.1)
Mus musculus(10090)	transmembrane protein 270 (TP_065651.1)
Rattus norvegicus(10116)	kidney-specific membrane protein (MP_065125.1)
Galus gallus(9031)	similar to transmembrane protein 27; collectrin; kidney-specific membrane protein; 041000L037R6 (UP_415821.1)
Homo sapiens(9606)	transmembrane protein 270 (TP_065716.1)
Canis familiaris(9615)	similar to transmembrane protein 20 (TP_054205.1)
Mus musculus(10090)	transmembrane protein 270 (TP_065651.1)
Rattus norvegicus(10116)	kidney-specific membrane protein (MP_065125.1)
Galus gallus(9031)	similar to transmembrane protein 27; collectrin; kidney-specific membrane protein; 041000L037R6 (UP_415821.1)



実験で証明された
タンパク質の糖鎖
修飾位置のDB

データ更新完了
論文投稿後に公開

更新作業中

(CC) BY

糖タンパク質統合DB



General Information

GlycoProtID	Gene Name	Accession	Taxonomy	Length	Molecular Weight
1318	RIKEN cDNA 0610007C21 gene	NP_082131	Mus musculus	223	23857.2

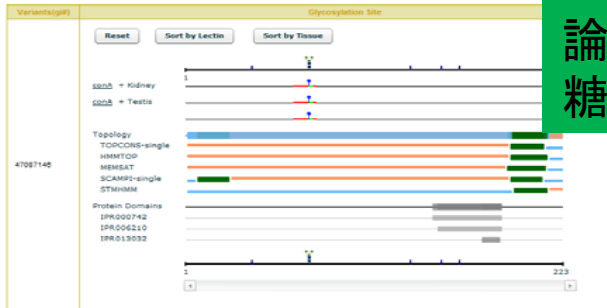
Gene Information

Gene ID	Symbol	Alias	Chromosome	Location	Description	External Links
381629	0610007C21Rik	0610007C21Rik, AI316792, Apr3, HSFC013, p1	5	5 B1J5	RIKEN cDNA 0610007C21 gene	MGI:1918918, Ensembl:ENSMUSG00000013622

Gene Ontology

ID	Class	Description	Evidence Code	PMID
GO:0005576	Component	cellular_component	ND	
GO:0016021	Component	integral to membrane	IEA	
GO:0016020	Component	membrane	IEA	
GO:0005886	Component	plasma membrane	IEA	
GO:0003674	Function	molecular_function	ND	
GO:0009156	Process	biological_process	ND	

Glycosylation Site



論文で報告された糖タンパク質のデータを集約

Sequence / Motifs

Sequence	N-glycosylation site																																																																																				
1 MARRRQQR AAALLLVLYV EALALAPED TLQDQRRG ₁₀ EYVAIVCEET EELGARDGL 41 RQRTIIGLH GQ ₁₁ LRPQC DVFQGFATL IIGQANLR DELANTFRF TQQTLLIPG 121 IYRFRQKHA WQVETFRF GYDQVQDQ EELVFRFRF RQDQANQDQ ELQVYVLDQ 181 IYRFRQKQD STELAFYFI LQETLIALI ILRGTQRKA KAP	<table border="1"> <thead> <tr> <th>Position</th> <th>Site</th> </tr> </thead> <tbody> <tr><td>10</td><td>Y</td></tr> <tr><td>11</td><td>Q</td></tr> <tr><td>12</td><td>R</td></tr> <tr><td>13</td><td>Q</td></tr> <tr><td>14</td><td>D</td></tr> <tr><td>15</td><td>Q</td></tr> <tr><td>16</td><td>R</td></tr> <tr><td>17</td><td>G</td></tr> <tr><td>18</td><td>R</td></tr> <tr><td>19</td><td>R</td></tr> <tr><td>20</td><td>Q</td></tr> <tr><td>21</td><td>D</td></tr> <tr><td>22</td><td>Q</td></tr> <tr><td>23</td><td>R</td></tr> <tr><td>24</td><td>G</td></tr> <tr><td>25</td><td>L</td></tr> <tr><td>26</td><td>V</td></tr> <tr><td>27</td><td>Y</td></tr> <tr><td>28</td><td>L</td></tr> <tr><td>29</td><td>V</td></tr> <tr><td>30</td><td>L</td></tr> <tr><td>31</td><td>V</td></tr> <tr><td>32</td><td>L</td></tr> <tr><td>33</td><td>L</td></tr> <tr><td>34</td><td>V</td></tr> <tr><td>35</td><td>Y</td></tr> <tr><td>36</td><td>L</td></tr> <tr><td>37</td><td>V</td></tr> <tr><td>38</td><td>L</td></tr> <tr><td>39</td><td>L</td></tr> <tr><td>40</td><td>V</td></tr> <tr><td>41</td><td>R</td></tr> <tr><td>42</td><td>Q</td></tr> <tr><td>43</td><td>D</td></tr> <tr><td>44</td><td>V</td></tr> <tr><td>45</td><td>F</td></tr> <tr><td>46</td><td>Q</td></tr> <tr><td>47</td><td>D</td></tr> <tr><td>48</td><td>V</td></tr> <tr><td>49</td><td>F</td></tr> <tr><td>50</td><td>R</td></tr> </tbody> </table>	Position	Site	10	Y	11	Q	12	R	13	Q	14	D	15	Q	16	R	17	G	18	R	19	R	20	Q	21	D	22	Q	23	R	24	G	25	L	26	V	27	Y	28	L	29	V	30	L	31	V	32	L	33	L	34	V	35	Y	36	L	37	V	38	L	39	L	40	V	41	R	42	Q	43	D	44	V	45	F	46	Q	47	D	48	V	49	F	50	R
Position	Site																																																																																				
10	Y																																																																																				
11	Q																																																																																				
12	R																																																																																				
13	Q																																																																																				
14	D																																																																																				
15	Q																																																																																				
16	R																																																																																				
17	G																																																																																				
18	R																																																																																				
19	R																																																																																				
20	Q																																																																																				
21	D																																																																																				
22	Q																																																																																				
23	R																																																																																				
24	G																																																																																				
25	L																																																																																				
26	V																																																																																				
27	Y																																																																																				
28	L																																																																																				
29	V																																																																																				
30	L																																																																																				
31	V																																																																																				
32	L																																																																																				
33	L																																																																																				
34	V																																																																																				
35	Y																																																																																				
36	L																																																																																				
37	V																																																																																				
38	L																																																																																				
39	L																																																																																				
40	V																																																																																				
41	R																																																																																				
42	Q																																																																																				
43	D																																																																																				
44	V																																																																																				
45	F																																																																																				
46	Q																																																																																				
47	D																																																																																				
48	V																																																																																				
49	F																																																																																				
50	R																																																																																				

Topology Prediction

Method	Prediction	Regions
TOPCONS-single	Outside	1-191
	Transmembrane	192-212
	Inside	213-223
HMMTOP	Outside	1-191
	Transmembrane	192-214
	Inside	215-223
MEMSAT	Outside	1-189
	Transmembrane	190-213
	Inside	214-223
SCAMPI-single	Outside	1-6, 212-223
	Transmembrane	7-26, 192-211
	Inside	27-191
STMHMM	Outside	1-193
	Transmembrane	194-214
	Inside	215-223

Domain Prediction

Interpro	Domain Name	Accession	Regions
IPR000742	EGF_3	PS50026	146-187
	GF-like	SM00181	149-187
	GF-like_reg_CS	PS00022	175-186
		PS01186	175-186

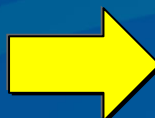
ENSEMBL	IPI	JGPI	NCBI_GI	REFSEQ	SWISSPROT	TREMBL	VEGA
ENSMUSP00000013766	IPI00471273.3	1318	47087146	NP_082131	Q8PGD0-1	A8C1S6	OTTMUSP00000025430

Related Protein

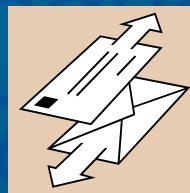
Taxonomy(ID)	Symbol(GENE_ID)	Accession(GI)
Homo sapiens (9606)	C2orf28 (51374)	NP_542159.3 (282847378)
Pan troglodytes (9598)	LOC740920 (740920)	XP_001154610.1 (114576571)
Canis lupus familiaris (9615)	LOC475699 (475699)	XP_532906.2 (73980654)
Mus musculus (10090)	0610007C21Rik (381629)	NP_082131.2 (47087146)
Rattus norvegicus (10116)	RGD1311605 (298841)	XP_216650.3 (109477992)
Danio rerio (7955)	si:ch211-101118.3 (553468)	NP_001038238.1 (229576935)

新規ツール: GlycoFASTA

入力: Multi-fasta



暗号化した URL を送付



糖タンパク質統合DB

公開用アミノ酸配列

GlycoFASTA server
 This Web service provides a system to detect currently-known glycosylation sites. It compiles the information on Glycopeptide reported in articles and runs FASTA36 for GlycopeptideDB(Subject) passing a user entry of amino acid sequences as the queries.

Please upload Multi-FASTA file 選択されていません

E-mail:

Species :

- Homo sapiens (To be released soon)
- Mus musculus
- Drosophila melanogaster
- Caenorhabditis elegans

AIST Research Center for Medical Glycoscience / CC BY-NC-ND

結果にアクセス
(一週間後に結果を削除)

GlycoFASTA server
 This Web service provides a system to detect currently-known glycosylation sites. It compiles the information on Glycopeptide reported in articles and runs FASTA36 for GlycopeptideDB(Subject) passing a user entry of amino acid sequences as the queries.

ENTRY
 1
 2
 3
 4

Title:gi|82460376|ref|NP_001014422.1| ABI gene family, member 3 (NESH) binding protein isoform 4 [Mus - 1099 aa
 Query Length:1099 a.a.

PeptideID	Region	Glycosylation sites
JPEP1540	811-938	816
JPEP1537	84-46	87

Viewer

JPEP1540
 JPEP1537

JPEP1540
 init: 177 init1: 177 opt: 177 Z-score: 337.0 bits: 68.9 E(3736): 2.2e-13
 Smith-Waterman score: 177; 100.0% identity (100.0% similar) in 29 aa overlap (811-938:1-28)

```

890 900 920 930 940
gi|824 1LDWEKPLNDTYEYEVISRENGSFGSKNKS1QITNQTFFSTVENLKPDSYEFQWPKNP
JPEP15  :                               SIQITNQTFFSTVENLKPDSYEFQWPKK
                               10 20
gi|824  LSEGPAQNTVAFSTESADPRVSEPIASGRDAIWRPFPPNSDSYSECKGKQYVWRVYKVF
    
```

JPEP1537
 init: 76 init1: 76 opt: 76 Z-score: 149.7 bits: 33.1 E(3736): 0.0059
 Smith-Waterman score: 76; 100.0% identity (100.0% similar) in 19 aa overlap (84-46:1-13)

```

10 20 30 40 50 60
gi|824  SLGQLLQSSIALALGNAGKLPKGGKPSLKHVHINTSDSILLKFLRPNANVKLEGFLLGY
JPEP15  :                               VHIITTSDSLILK
                               10
gi|824  GSNVSPNQYFPLPTGKFTFAVDAEPKYLIVRFPAPPSPGKSKSGSRPRKPLQLVVG
    
```


GlycoFASTA結果画面



GLYCOFASTA SERVER

GlycoFASTA server [Home](#)

This Web service provides a system to detect currently-known glycosylation sites. It compiles the information on Glycopeptide reported in articles and runs FASTA36 for GlycopeptideDB(Subject) passing a user entry of amino acid sequences as the queries.

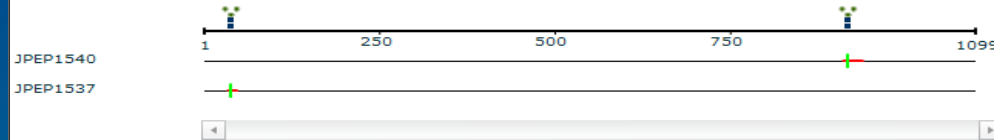
ENTRY

- 1
- 2
- 3
- 4

Title:gi|82480378|ref|NP_001014422.1| ABI gene family, member 3 (NESH) binding protein isoform 4 [Mus - 1099 aa
Query Length:1099 a.a.

PeptideID	Region	Glycosylation sites
JPEP1540	811-938	816
JPEP1537	84-46	87

Viewer



JPEP1540
initn: 177 init1: 177 opt: 177 Z-score: 337.0 bits: 68.8 E(3736): 2.2e-13
Smith-Waterman score: 177; 100.0% identity (100.0% similar) in 28 aa overlap (911-938:1-28)

```

      890      900      910      920      930      940
gi|824  ILDWEKPLNDIVTEYEVISRENGSFSGKNKSIQITNQTFSTVENLKPDTSYEFQVKPKNP
      ::::::::::::::::::::::::::::::::::::::::::::
JPEP15          SIQITNQTFSTVENLKPDTSYEFQVKPK
                    10          20
  
```

```

      950      960      970      980      990      1000
gi|824  LGEGPASNTVAFSTESADPRVSEPIISAGRDAIWTERPFNSDSYSECKGQYVVKRTWYKKF
  
```

JPEP1537
initn: 76 init1: 76 opt: 76 Z-score: 149.7 bits: 33.1 E(3736): 0.0059
Smith-Waterman score: 76; 100.0% identity (100.0% similar) in 13 aa overlap (34-46:1-13)

```

      10      20      30      40      50      60
gi|824  SLGCLLLCGSIALALGNAQKLPKGGKPSLKVHINTTSDSILLKFLRPANANVKLEGFLLGY
      ::::::::::::::::::::
JPEP15          VHINTTSDSILLK
                    10
  
```

```

      70      80      90      100      110      120
gi|824  GSNVSPNGYFPLPTEGKFTAVVDAEPKYLIWVRPAPPPSQKKSOSGKSRPRKPLQLVVG
  
```

統合化インターフェース構築の ためのAPIを設置

*National Institute of
Advanced Industrial Science
and Technology*

AIST

糖鎖関連遺伝子に関連するデータベースの統合化のための情報基盤整備

Human Glyco-Disease Genes Database

糖転移酵素の基質特異性のDB

77の病名を収録
原因遺伝子(糖鎖関連遺伝子)
専門家が書いた病態情報を収録

API構築中

API完成

阻害剤のデータベース

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

グリコシダーゼの基質特異性のDB

API完成

モデル生物のフェノタイプ情報

KOマウスDB
(名古屋大学・古川グループ
・産総研構築支援)

API完成

WormGlycoDB(仮称)
(九州大学・野村グループ
・産総研構築支援)

API完成

平成24年度からは
統合化インターフェースの開発

国際連携

*National Institute of
Advanced Industrial Science
and Technology*

AIST

アジアでの強力な連携体制づくり



2011年6月2~3日



2011年10月26日

予定: 2012年4月23-24日 アジア(中国・台湾・韓国)と欧州、米国のジョイントミーティング

- 第1回目の目的は、4カ国の糖鎖研究の現状について参加者の間で共有すること。国際協力の枠組みにおける糖鎖研究と糖鎖科学DBの可能性も議論した。

- 第2回目の目的は、新しいメンバーに向けてACGGDBの取り組みや進捗などを説明。どのような国際協力が可能なのかを探った。
- 最初の取り組みとして、糖タンパク質統合データベースの構築を提案した。

共同研究グループ

*National Institute of
Advanced Industrial Science
and Technology*
立命館大学
AIST

実験プロトコルのデータベース GlycoPOD

- 現在147プロトコルを公開
- 今年度は30プロトコルを公開した。
- 毎年30のプロトコルを執筆・公開していく予定

Protocols Authors Titles About Database

Authors

- ▶ Akihiko Kameyama
- ▶ Akiko Kinoshita-Toyoda
- ▶ Bruce Yong Ma
- ▶ Hatsumi M. Goda
- ▶ Hidenao Toyoda
- ▶ Hideto Watanabe
- ▶ Hiroaki Korekane
- ▶ Hiroshi Many
- ▶ Jun Hirabayashi
- ▶ Kazuaki Kakehi
- ▶ Keiko Furukawa
- ▶ Koichi Furukawa
- ▶ Koji Kimata
- ▶ Kyoko Kojima-Aikawa
- ▶ Makoto Ito
- ▶ Masao Yamada
- ▶ Mitsuhiro Kinoshita
- ▶ Motohiro Nonaka
- ▶ Motohiro Tani
- ▶ Naoyuki Taniguchi
- ▶ Nobuko Kawasaki
- ▶ Nongluk Sriwilaijaroen
- ▶ Nozomu Okino
- ▶ Retno Murwanti

jCGGDB GlycoPOD GlycoScience Protocol Online Database

Protocols Authors Titles About Database Login

List of protocols

- ▶ **Isolation & structural analysis of glycan**
 - ▶ Fluorescent labeling of glycan and HPLC analysis
 - ▶ Glycan Analysis by MSn Spectral Matching
- ▶ **N-Glycan**
 - ▶ Release of N-glycan by hydrazinolysis
 - ▶ N-glycanase digestion.
- ▶ **O-Glycan**
 - ▶ Glycosidase digestion of O-glycoproteins and related O-glycans.
 - ▶ Rapid and sensitive analysis of O-linked glycans using an in-line flow glycan-releasing system
- ▶ **Glycolipid**
 - ▶ Thin-layer chromatography (TLC) of glycolipids
 - ▶ Extraction of Glycolipids
- ▶ **Glycosaminoglycan**
 - ▶ Preparation of glycosaminoglycans
 - ▶ HPLC analysis of glycosaminoglycans
 - ▶ Heparin/heparin sulfate oligosaccharides
 - ▶ Chondroitin sulfate oligosaccharides
 - ▶ Keratan sulfate oligosaccharides
 - ▶ Hyaluronic acid oligosaccharides
- ▶ **Glycosyltransferase & related protein**
 - ▶ Enzyme assay of protein-O-mannosylglycan glycosyltransferases
 - ▶ Enzyme assay of sialyltransferase for glycolipids
 - ▶ Enzyme assay of sialyltransferase for oligosaccharides
 - ▶ Enzyme assay of sialidase
 - ▶ Enzyme assay of N-acetylglucosaminyltransferase III (GnT-III)

- [計画と進捗] 平成23年度は、GlycoEpitopeの新システムの開発を行った。Ruby on Railsの枠組に基づいた新システムの採用により、今後の運用・維持における持続可能性を高めることが期待できる。
現在、新システムは、全体のおよそ80%程度の開発が完了しており、年度末までに向けてドラフト版の完成を予定している。
- [改修予定] 平成24年度は、作成された新GlycoEpitopeシステムの動作チェックおよび実際のデータ入力作業からのフィードバックに基づいた調整を予定している。その後、現行システムから新システムへデータの完全移行をもって、新システムとして運用する予定である。
- [体制] GlycoEpitopeの情報入力・管理体制は、現状で4名程度である。専門論文からエピトープ・抗体情報を収集するが、さらに管理者権限を付与された研究者による二重チェックを行い、正確性とクオリティを確保している。平成24年度からは、新たに植物糖鎖を専門とする研究者がグループに参加することが決定しており、この領域の充実を図る。
- [予定] エピトープ数および抗体数は、現在までにそれぞれ161件、578件となっている。平成23年度末までに多少増え、平成24年度も継続してさらに登録数を増やしていく予定である、登録済みの抗体・エピトープに関するコンテンツを充実させるとともに、最終年度末には抗体数で700件となることを目標としている。

共同研究グループ

*National Institute of
Advanced Industrial Science
and Technology*

野口研究所

AIST

共同研究グループ(3) 野口研

(1) 研究開発題目 グライコナビデータベースの開発

(2) 研究開発の目的および内容

糖鎖大量合成や糖鎖研究において、化学構造の明確な糖鎖標準品は重要な意味を持っている。しかし、糖鎖標準品の合成方法や同定方法は複雑であるため、糖鎖合成に精通した合成化学者でなければ合成することが困難であることも多い。そこで、合成反応、NMRスペクトル、TLC、精製法のデータベースの開発を実施し、糖鎖標準品の合成をサポートする合成反応データベースシステムを開発する。

(3) 本年度の研究開発実施項目・概要

本研究開発では、全研究開発期間を通じ糖鎖合成化学を専門とする研究者と連携し、研究リソースとして欠かすことのできない、糖鎖標品の合成法と合成物の品質を確認するためのシステム開発とデータ整備を行う。本年度においては、糖鎖合成化学を専門とする研究者との連携体制を整備する。さらに、NMRスペクトル、TLC、精製法データベース登録システムを開発する。また、NMR、TLC、精製法データ整理のための管理方法の検討および管理ツールの開発を行うとともに、合成反応、NMRスペクトル・TLC・精製法データベースのデータ整理・登録更新を行う。NMRスペクトルデータベース構築に関しては、理研と協力して実施する。また、合成反応のデータベースからJCGGDBの「糖鎖合成のための統合データベース」へデータを移行する際のAPIの改良も行う。

- 糖鎖合成化学を専門とする研究者との連携体制
- 登録システム・管理ツール開発、データ整理・登録更新
 - NMRスペクトル(理研・野口研):管理ツール開発・データ整理
 - 合成反応データベース:管理ツール改良・登録更新
 - TLC・精製法データベース:管理ツール開発・データ整理
- 「糖鎖合成のための統合データベース」データ移行APIの改良

糖鎖合成化学を専門とする研究者との連携体制を整備

糖質合成データベースワーキンググループを結成
Carbohydrate Synthesis Database Working Group(CSDWG)

- 東北大学 正田先生
- 東北大学 小林先生
- 東北大学 野口先生
- 東海大学 稲津先生
- 東海大学 北条先生
- 東京工業大学 高橋先生
- 東京工業大学 田中先生
- 新潟薬科大学 小島先生
- 岐阜大学 木曾先生
- 岐阜大学 石田先生
- 岐阜大学 安藤先生
- 大阪大学 深瀬先生

•野口研究所 研究員

CSDWG

データ提供
糖鎖有機合成反応、NMRデータ等(文献…)
GlycoChemPOD
化学合成プロトコル集データベースへの要望…
システムに対する要望



野口研

NMR等提供データの整理・解析・登録
糖鎖合成反応文献の整理・登録
データベース表示構造式の検討…

糖鎖NMRデータベース開発(理研・野口研)

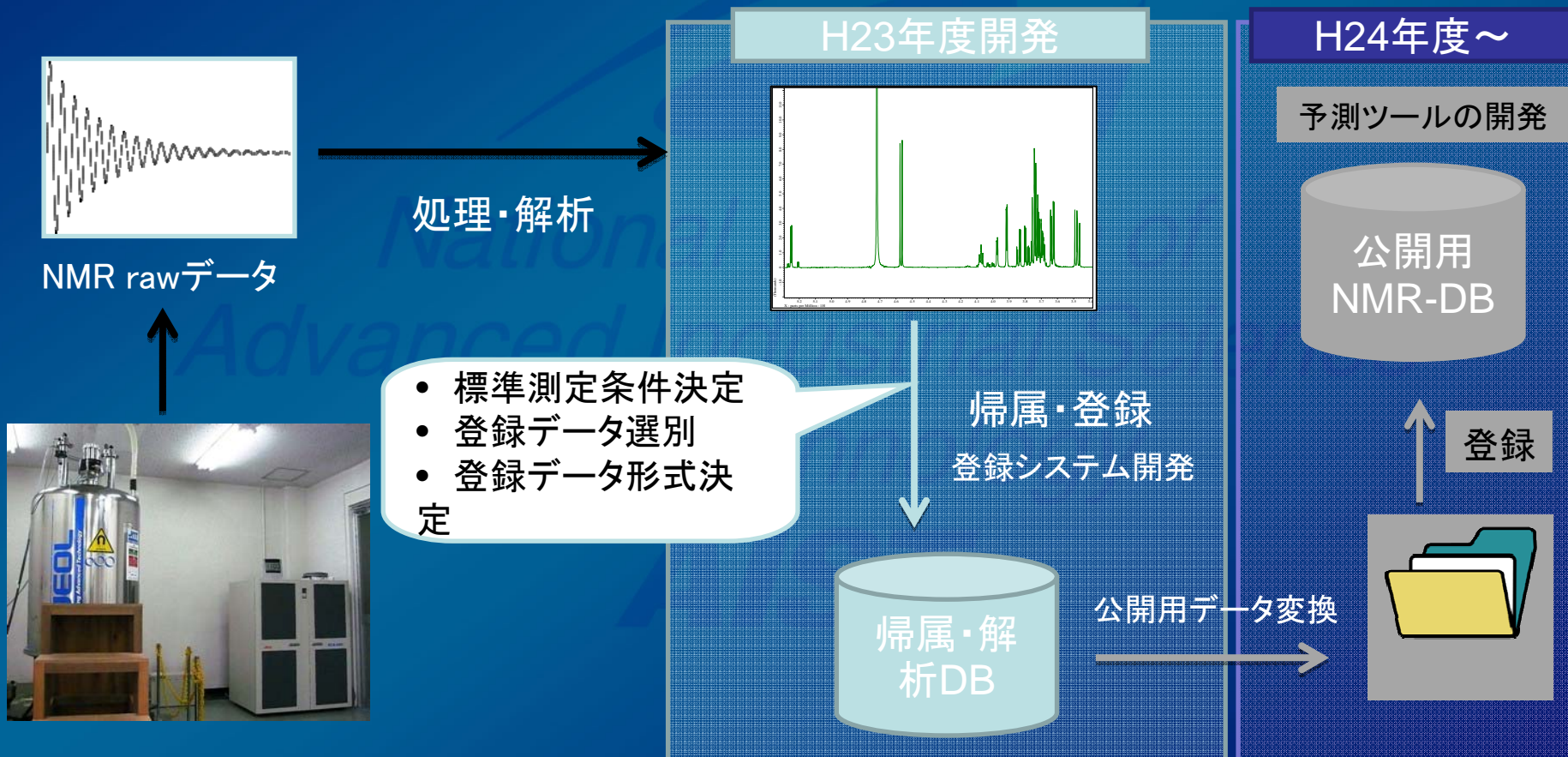
～糖鎖構造生物学・糖鎖合成化学～

糖鎖NMR(核磁気共鳴)データベースの特徴

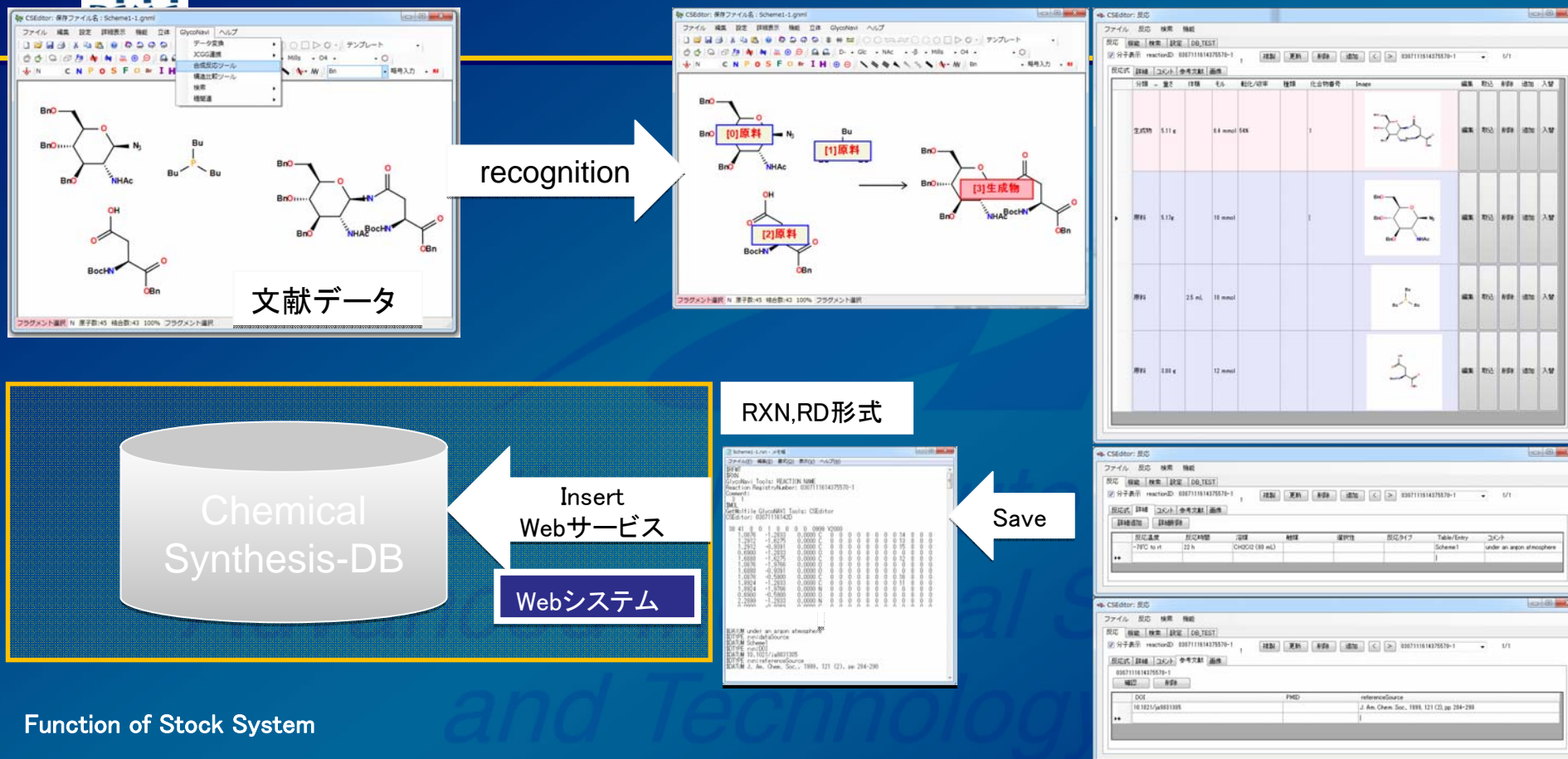
- 高分解能(600MHz)
- 統一測定条件
- スペクトルの完全帰属

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



合成反応データ登録システムの改良



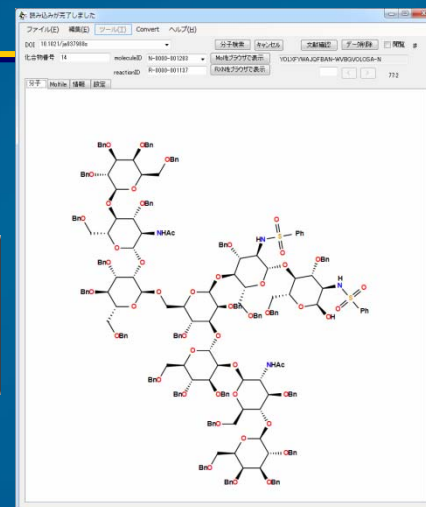
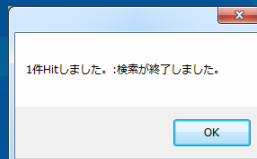
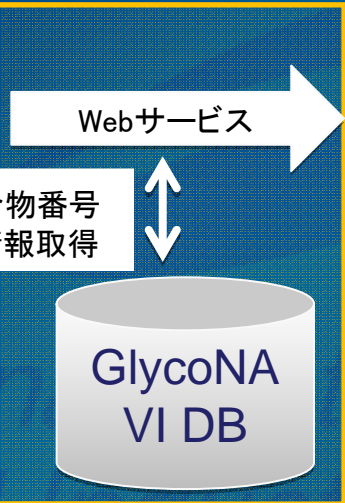
Function of Stock System

- Edit of chemical structure using CSEditor
- Registration of reaction information to local system
- Retrieval of reaction information
- Edit of reaction information (temperature, reaction time, etc...)
- Enter its comments, usage, and yield for each compound

反応式	試料	試料	試料	試料	試料	試料	試料	試料	試料
生成物	5.11 g	8 mmol	54%						
原料	5.17g	10 mmol							
原料	25 mL	10 mmol							
原料	1.01 g	12 mmol							

保存データ
XMLを利用した
データ構造

AIST LC・精製法・NMR等データ登録システムの開発



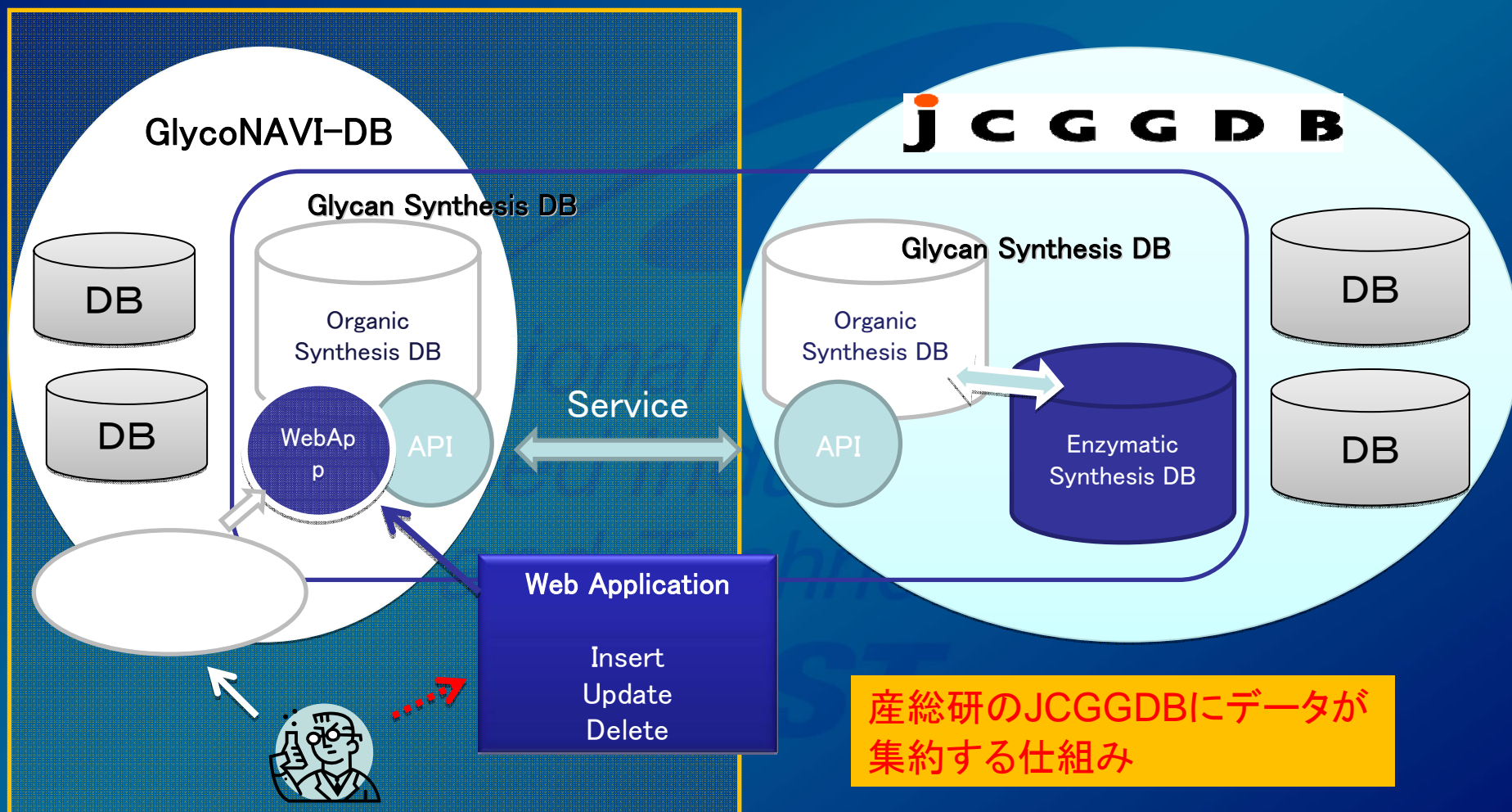
文献データ



保存データ XMLを利用したデータ構造

糖鎖合成統合データベース開発

「糖鎖合成のための統合データベース」データ連携APIの改良



研究開発の達成度～データ登録整理～

第1年次研究開発（平成23年度）

研究開発の達成度は、1)NMRスペクトル、TLC、精製法データベース登録システムの開発完了
2)合成反応データベースについては、**合成反応を1500件**整理・登録 3)**NMRスペクトル・TLC・精製法のデータを各50件**整理することを判断基準とする。

	達成度(2012/2)
1) データベース登録システムの開発	80%
2) 合成反応整理登録更新	73%
- 合成反応データ:	
・ 合成反応整理登録(1089件)	
3) 分子情報データ整理	
- NMRデータ:	
・ 実験データ整理・帰属(18件)	36%
・ 文献データ整理(200件)	(400%)
- TLCデータ:	
・ 文献整理・登録(40件)	80%
- 精製法データ:	
・ 文献整理・登録(59件)	118%

共同研究グループ

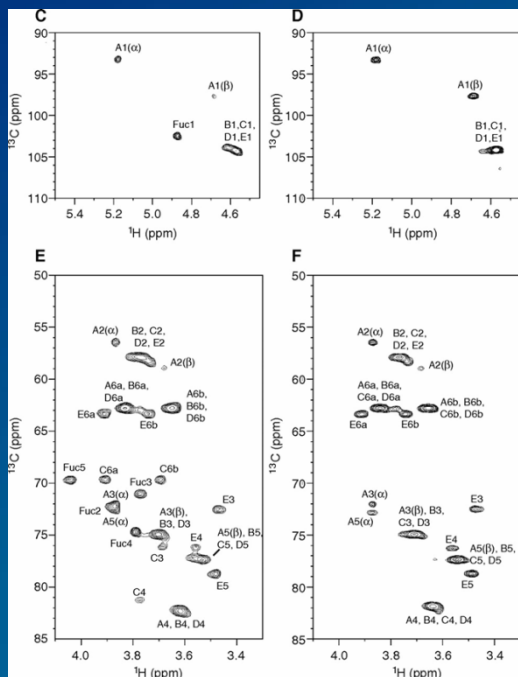
*National Institute of
Advanced Industrial Science
and Technology*
理化学研究所
AIST

RIKEN Glycan NMR Database

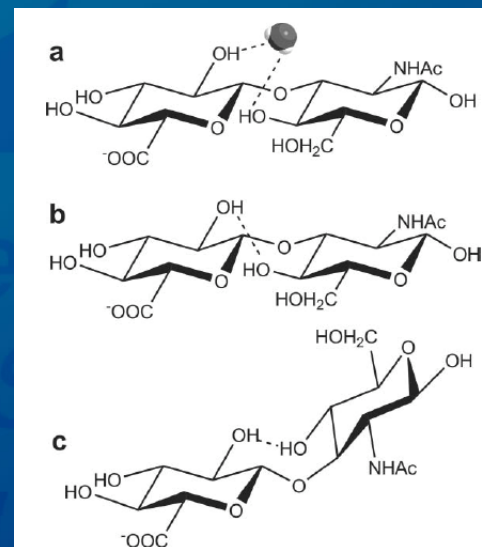
NMR Chemical shift Data
Chemical shift 経験則

Prediction of chemical shift
(protein-free and -bound state)

Collaboration with Dr. Sugita's group



NMRデータの蓄積
構造データの蓄積
計算機シミュレーション



Prediction of glycan conformation
Prediction of glycan-protein
interaction

糖鎖のNMRデータベース構築のコンセプト

既存の糖鎖NMR化学シフトデータベースの問題点

SUGABASE:

<http://www.boc.chem.uu.nl:8081/sugabase/sugabase.html>Glycosciences de: <http://www.glycosciences.de/index.php>

測定条件がまちまち
データベースが更新されていない
情報が部分的である

新しく目指す糖鎖NMR化学シフトデータベース

同一の測定条件・同一の化学シフト基準物質とする
完全帰属と付加情報 ($^3J_{\text{HH}}$ 等)

化学シフトと立体構造の相関解析等に適用可能

NMR測定・完全帰属をしたもの(11月23日現在)

(glucose disaccharides)

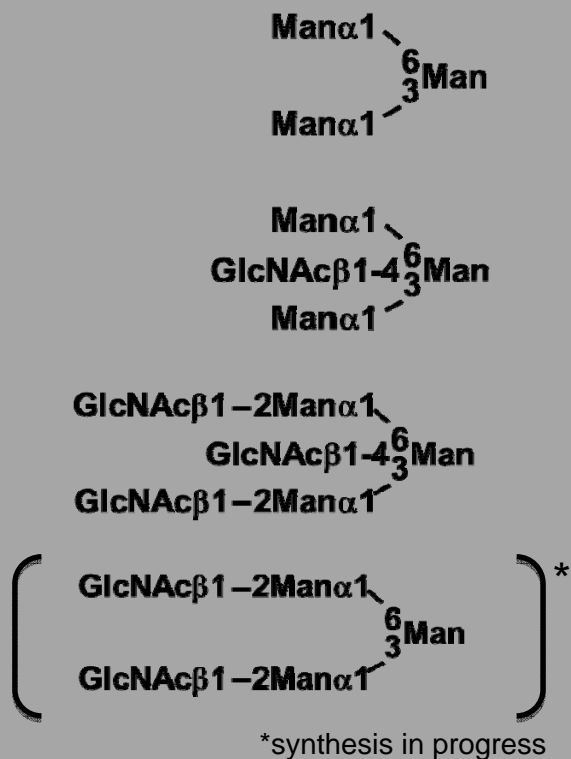
8種類

- Glc α 1-2Glc
- Glc α 1-3Glc
- Glc α 1-4Glc
- Glc α 1-6Glc

- Glc β 1-2Glc
- Glc β 1-3Glc
- Glc β 1-4Glc
- Glc β 1-6Glc

(bisecting GlcNAc form)

3種類



(carbohydrate oligomers)

18種類

- (Glc β 1-3)_n
n=3-7

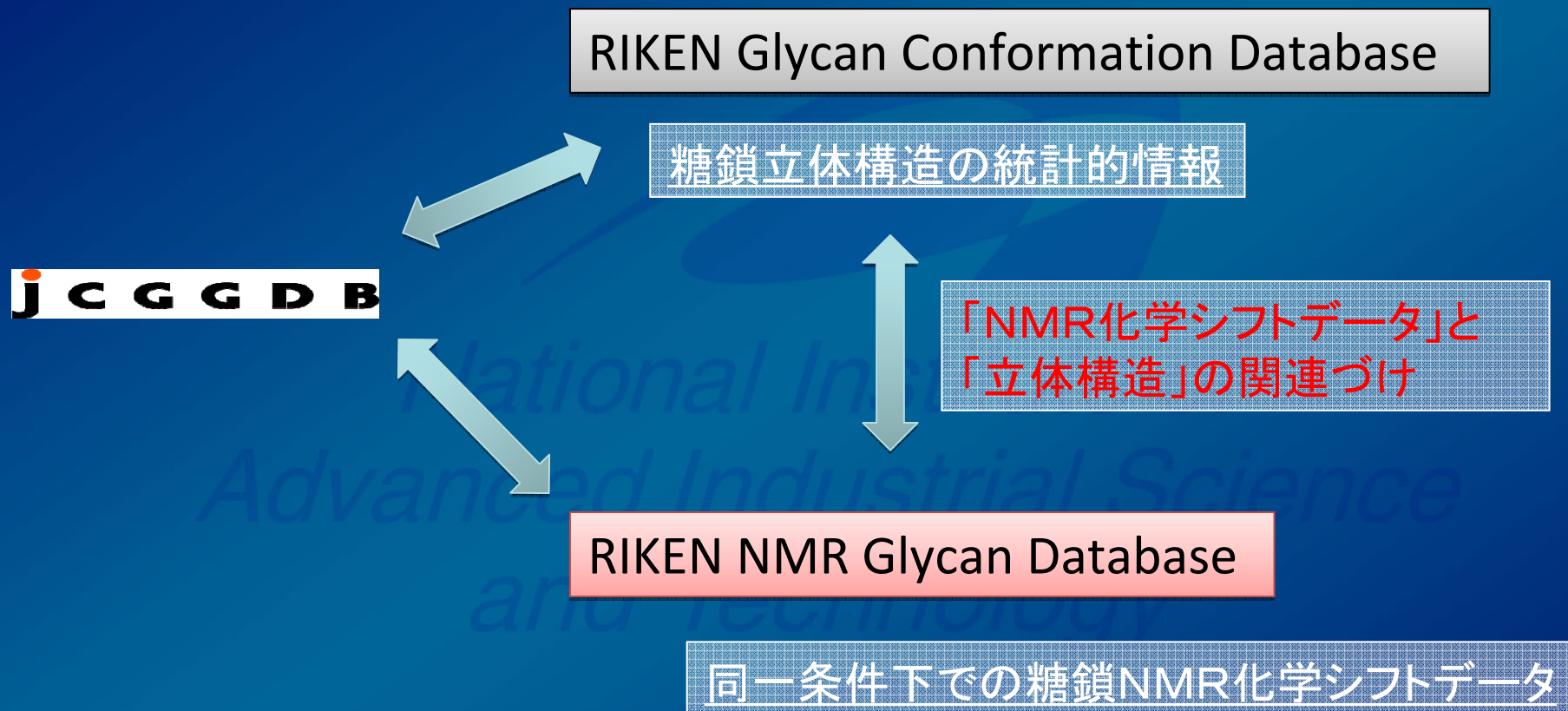
- (Glc α 1-4)_n
n=2-7

- (NeuAc α 2-8)_n
n=2-6

- (GlcNAc β 1-4)_n
n=3,5

これまでに29種

Relationships among RIKEN Glycan-Related Databases



まとめ

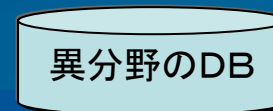


JCGGDBの機能強化と拡張

連携強化



1~3年目



異分野のDB

DBCLS
RDF化
セマンティックの利用

実験プロトコル

生物学系



化学系



糖鎖有機&酵素合成DB

- ・合成のプロトコルを取りまとめる
- ・合成化合物(リソース)の提供

1~3年目

合成物の確認
(研究支援)

(一例として)

糖鎖NMRDB

- ・統一した条件下で糖鎖のNMRシグナル

1~2年目

今までの文科省・統合DBプロジェクトの

継続

- ・DBとサーバの管理
- ・JCGGDBのデータ更新
- ・新規DBの構築

拡張

- ・DB間連携・有機的統合
- ・研究者主導型の蓄積型DBにするための改良
- ・webやAPIの高機能化
- ・DBの知名度向上

研究支援ツールの構築



NMRシグナルの構造への帰属ツールの開発など

2~3年目

統合化支援

更なる新規&既存DBの統合(国内)
未公開データをDB化支援(国内)
海外DBとの連携

糖鎖関連のDB

1~3年目