



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

○木下聖子^{1,2)} 青木信行²⁾ 新町大輔²⁾ 藤田晶大²⁾ 土屋伸一郎²⁾
藤田典昭¹⁾ 鈴木芳典¹⁾ ソロビヨワイエレナ¹⁾ 鹿内俊秀¹⁾ 松原正陽³⁾
山田一作³⁾ 奥田修二郎⁴⁾ 川壽敏祐⁵⁾ 成松久¹⁾

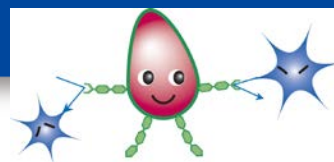
¹⁾産業技術総合研究所 糖鎖創薬技術研究センター、²⁾創価大学、³⁾野口研究所、⁴⁾新潟大学、⁵⁾立命館大学

トーゴーの日シンポジウム2014

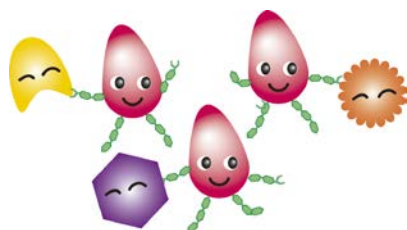
2014年10月5日



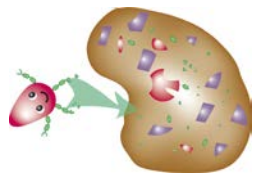
水に溶けやすくなる



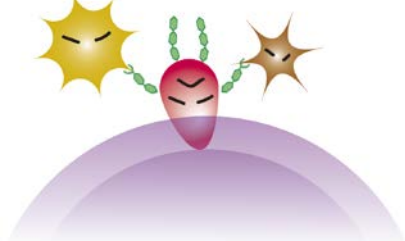
タンパク質分解酵素から守る



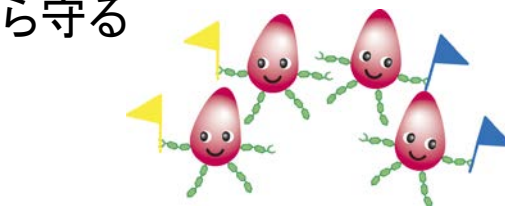
レクチンやホルモン、増殖因子と結合する



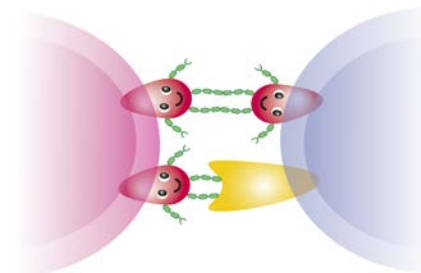
糖タンパク質の代謝を決定する



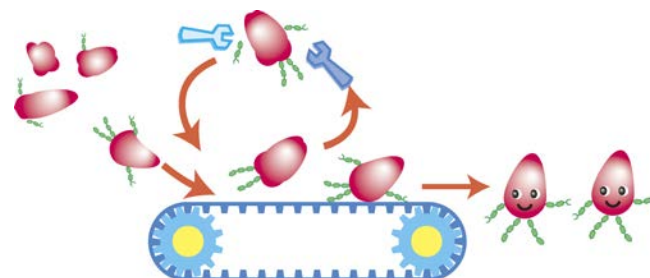
病原菌やウイルスの受容体になってしまう



行き先を明確にする



細胞と細胞を接着させる



タンパク質の品質管理

糖鎖の機能

糖鎖関連データベース1989



糖鎖関連データベース 2014



本プロジェクトの目標

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
 - 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)

JCGGDBについて



JCGGDB Alliance

産総研・糖鎖センターのDB

GlycoGene Inhibitors database

Pathway database

GlycoGene DNA microarray database

Glycosidase database

- データの拡充
- 検索機能強化
- 統合化推進

Glycotechnology Patents database

JCGGDB Report

TuMa Rdb

JMSDB

GlycoPOD

野口研究所

名古屋大学・産総研

Knockout Mouse DB

創価大学

FlyGlyco DB

立命館大学

名古屋市立大学

KEGG

京都大学

生化学工業

Glycan Synthesis DB

野口研・産総研

LipidBank構築委員会

理研・システム糖鎖

N-glycan Conformation DB

Worm database

九州大学

Advanced Search

Keyword
Family Classification

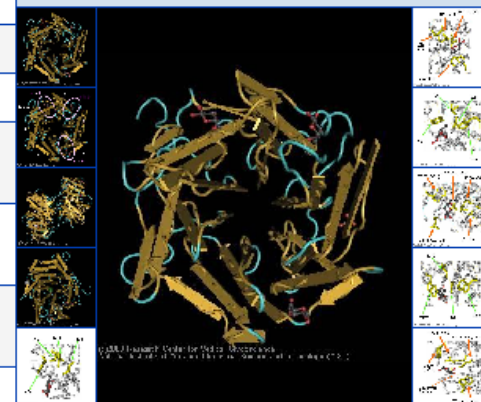
Keyword

AAL (LFF-00003f)

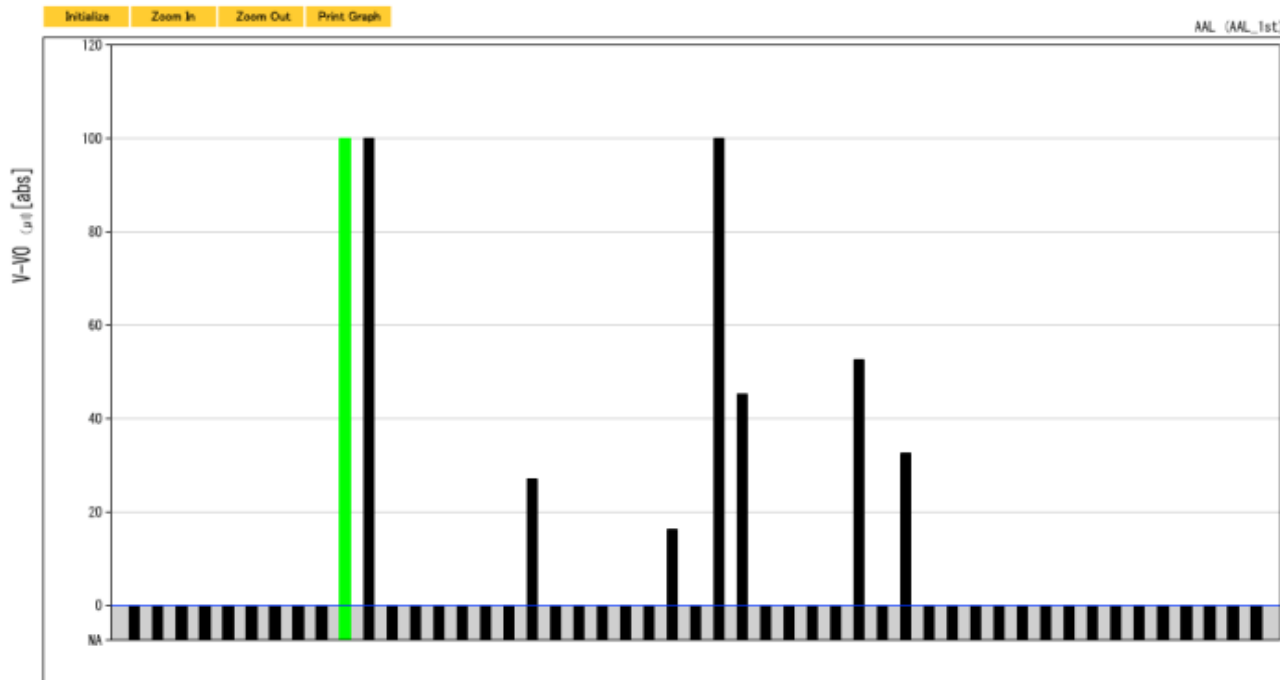
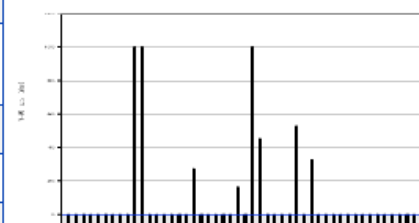
[Print]

Lectin ID	LFF-00003f
Lectin Name	AAL
Lectin Family	Fucose-binding Lectin

Molecular Structure



Interaction Graph [Viewer] [GlycanList]

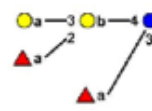


Selected glycan



418

Clicked glycan



723

1-parameter

```
riraqr ggdnpwtggs sqnvigeakl
lgskwit gqgsvgvkv gnsklaalq
rtagasf gstvpgtgig ataigpgrlr
iaaaisw gstpnrivyw qkgreelyea
ididsvf fqasgvslqq wqwisgkqws
```

Reference

GlycoProtDB (AIST)



Tissues
(Liver, brain, lung, kidney, testis, stomach, colon, heart, skeletal muscle, & serum)

Protein digestion

Peptide mixtures

Lectin column

ConA, RCA120, AAL, WGA, SSA, and/or Amide80(HILIC).

Glycopeptide subsets

GlycoProt ID: GP_MMU_0000025

IGOT-LC/M

Q-TOF Ultima (LTQ-Orbitrap Vc)

Mascot search

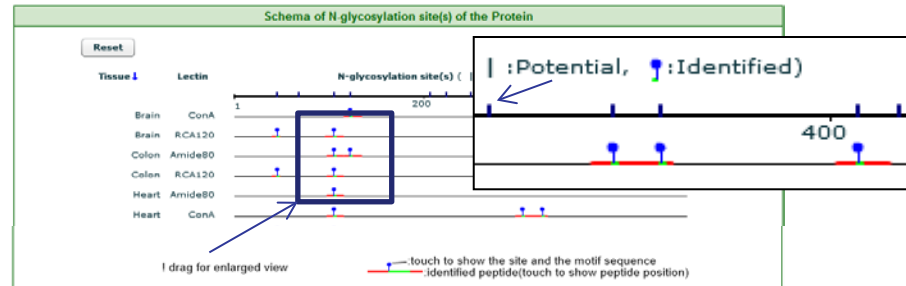
Now, going on small intestine, pancreas, T-cells, B-cells, amniotic fluid, embryo tissues....



Peptide sequence, modification
AAAANGSAAAK : 0.00040000
AAANVTAAR : 0.00040000
(4: glycosyl)

Protein Name lysosome membrane protein 2 precursor [Mus musculus]
Protein Accession Number NP_031670
Protein GI 6680878
Gene Symbol Scarb2
Gene Name scavenger receptor class B, member 2
Gene ID 12492

▼Glycosylation Sites



▼Sequence / Glycosylation site(s)

(Potential Sequon, Identified Site)

```

1  NRSQPTAS TLELLLVTE VLLVAVFQ KAVDQTEKH NVLQETVNF NENKFFLPV YIQFFVPTI NPERILQEI FLLEEVGPTV YVELLNKAI
101 QFCHLSEVTS AVTRKAVYFK QVQVQVNV ELLEINTEL LTVVLAGLI LRELIEMAL KAVQGLPFI HYVRELLACHY KQRLLELVI FQSDVDFPC
201 LFFVNSHNDI GEVFLVIGD NYVAFSKIVE NNGETLMMN YTDTCNMNS TGDGPHIEL SKREVLVLFK SMLGSHVIT FQSFNVEGL DAFYFVFAE
301 ILLQENAG PCIEFGNCHD SOVLEICK NGAPIMSFP HFYQADEKVF SAIKGRHFK EKHESFVDIN ELTGILRGA KRFGINTYVR KLDGFVETD
401 IRTNVFVNVY HVEVLIDKE TANQLKEVIN TLLVTVNIFV IIMALGVFFK LVFTNLACRG QGSDGECTAD KRAPLIRT
    
```

■ N-glycosylation sites (Identified and Potential) Identified only

Potential sites		Identified Peptide * 1				Brain		Colon		Heart		Kidney	
Position	Sequon	Position	Sequence	Unique or Shared	ConA	RCA120	Amide80	RCA120	Amide80	ConA	RCA120	ConA	RCA120
45	NGT	40-40	NMVLDQTK	Unique		Y		Y			Y		Y
	NVT	40-45	NVLDQTK	Unique									
105	NGT	96-115	NKANQFGEQTTSAVTK	Unique		Y							Y
		90-115	ANQFGEQTTSAVTK	Unique				Y	Y	Y	Y	Y	Y
122	NQS	116-134	KYVFERQVYQDPVQLR	Unique	Y		Y	Y	Y	Y	Y	Y	Y

GeneSymbols

Caenorhabditis elegans : 701

Mus musculus : 1830

: 937 (GalT^{-/-} KO mouse)

Homo sapiens : 447 (HCC)

Human-derived cultured cell : 842 (Supernatants)

説明

- GlycoProtDBとは
- データの収集法
- Database Construction
- 参考文献
- WEBサービス
- お問い合わせ

GlycoProtDB検索

キーワード

ID(完全一致検索)
絞り込み検索

jcgddb@m.aist.go.jp

(独)産業技術総合研究所
糖鎖工学研究センター
グライコプロテオーム解析チーム

〒305-8568
茨城県つくば市福田 1-1-1
中央第2事業所 OSL

技術開発機構

研究情報公開データベース

Peptide: KNNMEITWTPM^NNATSAFGPNLR

Peptide: KNNMEITWTPM^NNATSAFGPNLR

Included by

GI	Symbol
35215309	Nfasc
237858630	
237858632	
237858634	

Protein GI: 237858632

Protein (or Gene) Name: neurofascin
Protein Accession Number: NP_0011537
Gene Symbol: Nfasc
Gene ID: 269116

▼ Sequence / Motifs

```

1   MARQQAPPVW HIALILFLLS LGGAIEIPMD LTQPFTITKQ SVKDHIVDPR DNILIECEAK GNPAPSFHWT RNSRFFNIAK DPRVSMRRRS GTLVIDFRSG
101  GRPEEYEGEY QCFARNKFGI ALSNRIRLQV SKSPLWPKEN LDPVVVQEGA PLTLQCNPPP GLPSPIVFWM SSMPEPITQD KRVSQGHNGD LYFSNVMLQD
201  MQTDYSCNAR FHFHTTIQQK NPFTLKVLTN NPYNDSSLRN HPDIYSARGV AERTPSFMYP QGTSSSQMVL RGMDDLLECI ASGVPTPDIA WYKGGDLPS
301  NKAKFENFNK ALRITNVSEE DSGEYFCLAS NKMGSIRHTI SVRVKAAPYW LDEPKNLILA PGEDGRLVCR ANGNPKPTVQ WMVNGEPLQS APPNPNREVA
401  GDTILFRDTQ ISSRAVYQCNTSNEHGYLLA NAFVSVLDVP PRMLSARNQL IRVILYNRTR LDCPFPGSPI PTLRWFKNQG GSNLDGGNYH VVENGSLEIK
501  MIRKEDQGIY TCVATNILGK AENQVRLEVK DPTRIYRMPD DQVAKRGTTV QLECRVKHDP SLKLTIVSWLK DDEPLYIGNR MKKEDDSLTI FGVAERDQGS
601  YTCMASTELD QDLAKAYLTV LGRPDRPRDL ELTDLAERSV RLTIWPGDDN NSPITDVVVQ FEEDQFQPGV WHDHSRFPQS VNSAVLHLSF YVNYQFRVIA
701  VNEVGS SHPS LPSERYRTSG APPESNPSDV KEGGTRKNNM EITWTPMNNAT SAFGPNLRYI VKWRRRRETR TWNNVTWVGS RYVVGQTPVY VPYEIRVQAE
801  NDFGKGPEPD TIIGYSGEDY PRAAPTEVKI RVLNSTAISL QWNRVYSDTV QGQLREYRAY YWRESSLLKN LWVSQKRQQA SFFGDRPRGV VARLFFYSNY
901  KLEMVVVNGR GDGPRSETKE FITPEGVPSA PRRFRVRQPN LETINLEWDH PEHPNGILIG YILRYVPENTKLGKQMVEN FSPNTKFSV QRADPVSRYR
1001 FSLSARTQVQ SGEAATEESP APPEANPTA AYTNNQADIA TQGFIGLMC AIALLVLLIL IVCFIKRSRG GKYPVREKDD VPLGPEPKPE EDGSPDYSDS
1101 DNKPLQGSQT SLDGTRKQQE SDDSLVDYGE GEGEQVFNEDG SFIGQYTVKK DKEETEGNES SEATSPVNAI YSLA
  
```

■ N-glycosylation sites Identified and Potential Identified only

Potential sites		Peptide fragment*1		AAL	ConA	RCA120
Position	Sequon	Position	Sequence	Unique or Shared	Brain	Brain
234	NDS					
316	NVS	314-332	IT ^N VSEEDS ^N GEYFCLASNK	Shared (4)	•	•
420	NTS	415-442	AVYQC ^N TSNEHGYLLANAFVSVLDVPPR	Shared (4)	•	•
457	NRT					
494	NGS					
650	NNS					
748	NAT	737-758	KNNMEITWTPM ^N NATSAFGPNLR	Shared (4)	•	•
		737-758	KNNMEITWTPM ^N NATSAFGPNLR	Shared (4)	•	•
		737-758	KNNMEITWTPM ^N NATSAFGPNLR	Shared (4)	•	•
		737-758	KNNMEITWTPM ^N NATSAFGPNLR	Shared (4)	•	•
774	NVT	770-781	ETW ^N NVTWVWGSR	Shared (4)	•	•
834	NST					
989	NGT	965-972	YVPFN ^N GTK	Shared (4)	•	•
980	NFS					
984	NQT	980-987	YVPFN ^N GTK	Shared (4)	•	•
		976-987	QMVENFSP ^N NQTK	Shared (4)	•	•
		976-987	QMVENFSP ^N NQTK	Shared (4)	•	•
1158	NES					

*1: _:Motif's potential site ■:Asn (glycosylation+180) ■:PyroGln (N terminus) ■:Met (oxidation) ■:Deamination (NtermCys)

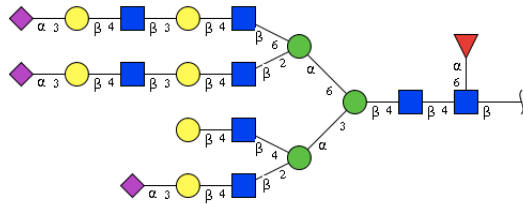
JCGGDBのセマンティックウェブ化

	オントロジー開発	RDF化	スタンザ(UI)開発
GlycoProtDB: 糖タンパク質の糖鎖結合部位情報	完了	完了	完了
LfDB: レクチンの糖鎖結合親和性データ	完了	完了	開発中
GlycoEpitope: 糖鎖抗原とそれを認識する抗体	完了	完了	開発中
GDGDB: Glyco- Disease Genes DB	完了	完了	完了
PacDB: Pathogen adherence DB	開発中	開発中	開発中
GlycoSite: 糖タンパク質の糖鎖情報	開発中	開発中	—

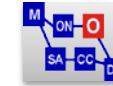
国際糖鎖構造リポジトリシステムの構築と 運用



この糖鎖についての情報は？



糖鎖関連データベース



検索方法:

描画が難しい... 書き方が統一していない...

一つの糖鎖構造についての情報を一カ所に統合できないか？

糖鎖構造や関連情報をデータベースに追加するのにどうしたら良いか？

(slide by Will York@CCRC)

glyspace

[http://en.wikipedia.org/wiki/Namespace_\(computer_science\)](http://en.wikipedia.org/wiki/Namespace_(computer_science))

A **namespace** (sometimes also called a **name scope**) is an abstract container or environment created to hold a logical grouping of unique identifiers or symbols (i.e., names).

- Stable and maintained over the long term (NCBI)
- Limited in scope (structures and IDs only)
- Minimally Curated
- Searchable
- Organized

GYCO22のACGG-DB会議 (2013年6月、中国にて)

Glycobiology vol. 23 no. 12 pp. 1422–1424, 2013
doi:10.1093/glycob/cwt084

Glyco-Forum section



The Fifth ACGG-DB Meeting Report: Towards an International Glycan Structure Repository

Kiyoko F Aoki-Kinoshita¹, Hiromichi Sawaki², Hyun Joo An³, Matthew Campbell⁴, Qichen Cao⁵, Richard Cummings⁶, Daniel K Hsu⁷, Masaki Kato⁸, Toshisuke Kawasaki⁹, Kay-Hooi Khoo⁷, Jaehan Kim³, Daniel Kolarich¹⁰, Xianyu Li⁵, Mingqi Liu¹¹, Masaaki Matsubara¹², Shujiro Okuda^{9,13}, Nicolle H Packer⁴, René Ranzinger¹⁴, Huali Shen¹¹, Toshihide Shikanai², Daisuke Shinmachi², Philip Toukach¹⁵, Issaku Yamada¹², Yoshiki Yamaguchi⁸, Pengyuan Yang¹¹, Wantao Ying⁵, Jong Shin Yoo¹⁶, Yan Zhang¹⁷, Yang Zhang¹¹, and Hisashi Narimatsu²

repository would be that: (a) it would function as a central location where glycan structures are registered, similar to DDBJ/EMBL/GenBank of the International Nucleotide Sequence Database Collaboration (INSDC) for nucleotide sequences, (b) it would provide unique identifiers for every glycan structure (including ambiguous structures and even monosaccharide compositions) such that glycan IDs can be uniquely identified from any resource by using these identifiers, and (c) researchers who publish newly identified structures can use the identifiers to link their structures with relevant information in their publications.

The next issue was to determine the data content, or scope,

糖鎖関連データベース

UniCarbKB

GlycomeDB

Plant & Fungal Carbohydrate Structure Database

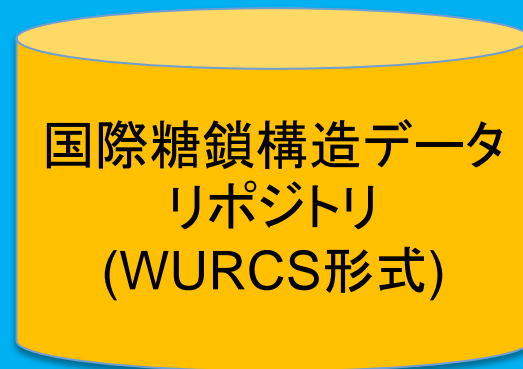
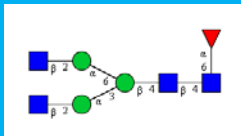
9-1-9-C-0-SCIENCES.DE

JCGGDB DB

M ON-O SA-CC DB



糖鎖ID + 構造データ



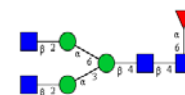
国際糖鎖構造データリポジトリ (WURCS形式)

⑤ データ共有

糖鎖研究の知識共有

① 登録

② 糖鎖ID発行



研究者

研究論文

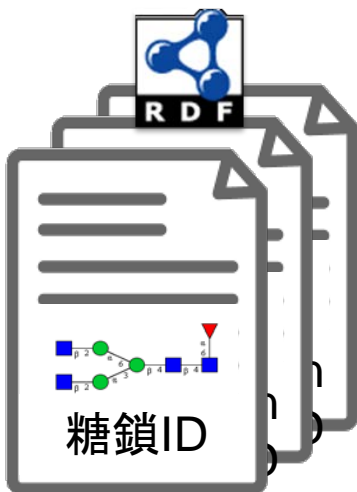
③ 投稿



糖鎖ID

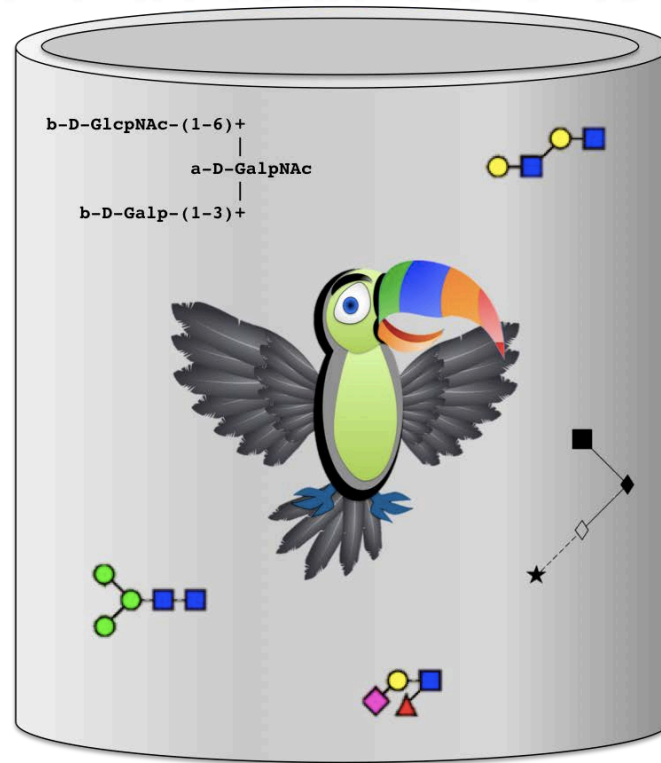
学術雑誌

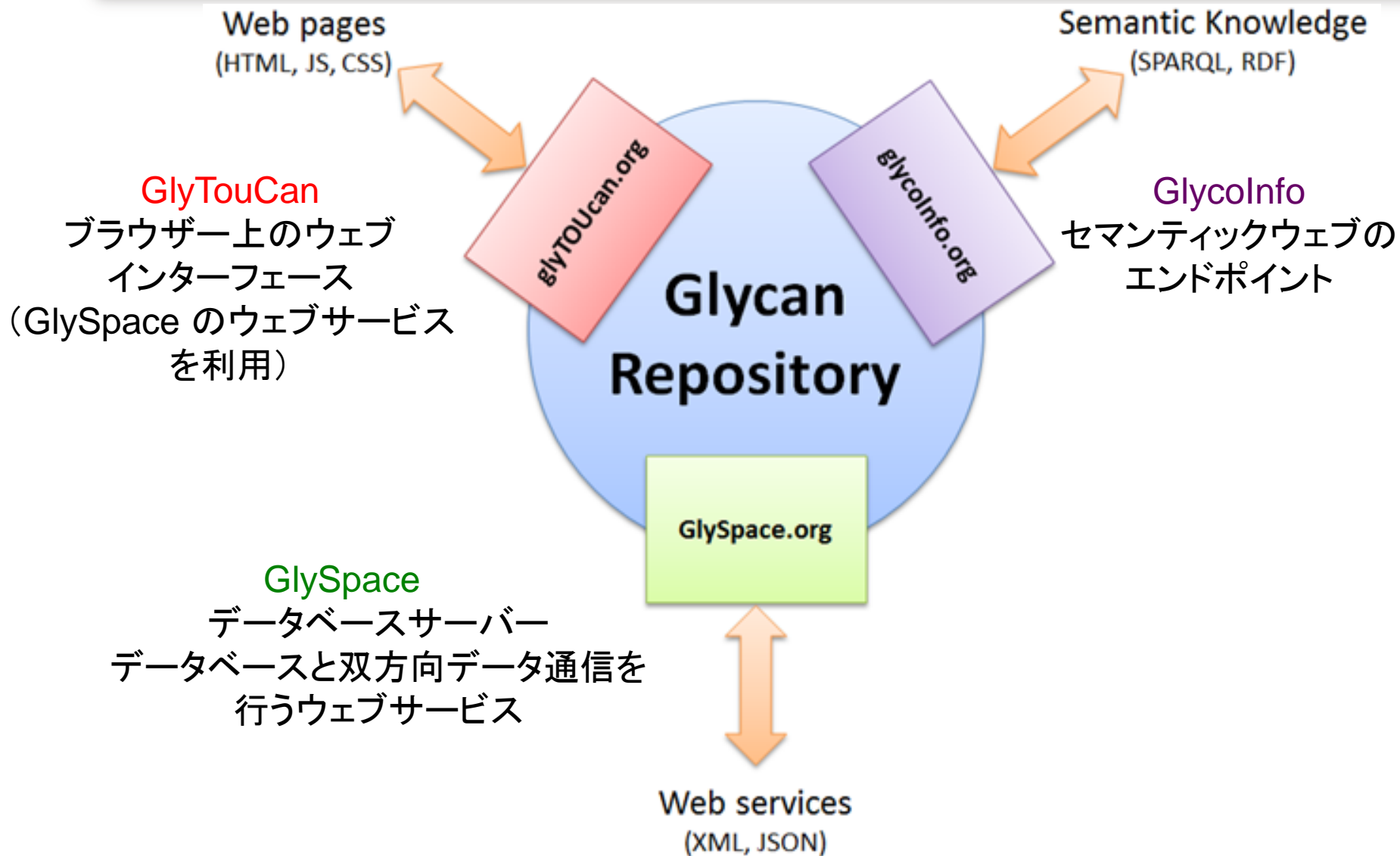
④ アノテーション情報のキュレーション



糖鎖リポジトリ”GlyTouCan”の構想

糖鎖 → Glycan → 糖缶 → TouCan → GlyTouCan





Search ▾

View All ▾

Accession Number

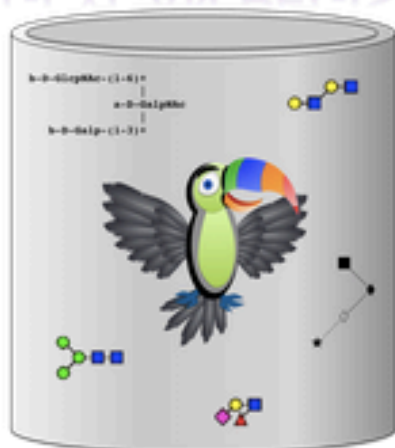
Search

Preferences

SignIn

SignUp

THE GLYCAN REGISTRY



GlyTouCan is the web interface for 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 including glycosidic linkage configuration, can be registered as long as there are no inconsistencies in the structure.

Users can search for glycan structures and motifs that have been registered into this repository. Registered users can additionally register new glycan structures to obtain unique IDs for each structure, which can be used in publications and other databases upon approval.

糖鎖構造の登録画面

Registration ▾

Search ▾

Glycan List

Accession Number

Search

Preference

Glycan String

File Upload

Glycan Registration

Input your glycan structure(s) below in GlycoCT condensed format.

```
RES
1b:a-dgal-HEX-1:5
2s:N-acetyl
3b:b-dgal-HEX-1:5
4r:r1
5b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d
6s:n-acetyl
7r:r2
8b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d
9s:n-acetyl
```

Submit

RES

1b:a-dgal-HEX-1:5

2s:N-acetyl

3b:b-dgal-HEX-1:5

4r:r1

5b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d

6s:n-acetyl

7r:r2

8b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d

9s:n-acetyl

LIN

1:1d(2+1)2n

2:1o(3+1)3d

3:3o(3+1)4n

4:4n(3+2)5d

5:5d(5+1)6n

糖鎖構造登録確認画面

tration ▾

Search ▾

Glycan List

Accession Number

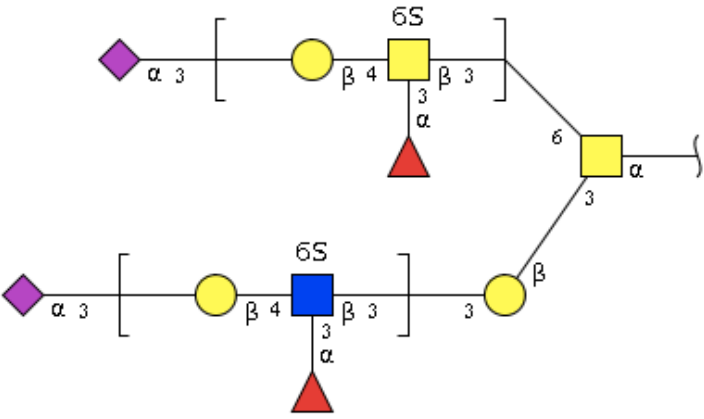
Search

Preference

Pr

Confirmation

The following structure(s) will be registered upon clicking the submit button.

	Structure	Image
No.1	<pre>RES 1b:a-dgal-HEX-1:5 2s:N-acetyl 3b:b-dgal-HEX-1:5 4r:r1 5b:a-dgro-dgal-NON- 2:6 1:a 2:keto 3:d 6s:n-acetyl 7r:r2 8b:a-dgro-dgal-NON-</pre>	

submit

Glycan LIST

Range of the Mass



1326 Glycans

LIST • STRUCTURE

sorted by: [mass](#); then by... • grouped as sorted

1 • 2 • 3 ... 67 Next »

Number of Fuc



Number of Gal



Number of GalNAc



Number of Glc



Number of GlcNAc



Number of Man

Tag

- ambiguous (152)
- Epitope (1159)
- full (1326)
- Glycosphingolipid (285)
- N-Glycan (889)
- O-Glycan (152)
- Subclass (51)

Motif

- Blood group A (43)
- Blood group B (2)
- Blood group H (157)
- Galalpha1-3Gal epitope (178)
- Glycosphingolipid Arthro series (16)

1.

Accession Number : [G67988RO](#)

Image : 

Mass : 342.116211546

Motif : Glycosphingolipid Gala series

Classification : Glycosphingolipid and full

Contributor : Administrator

Contribution time : Sun, Jul 27, 2014, 11:40 pm

2.

Accession Number : [G39461CL](#)

Image : 

Mass : 383.142760647

Preferences

Image notation : IUPAC

Change the graphical representation of glycan

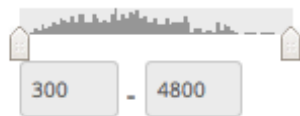
Select image notation

- ✓ CFG symbols
- CFG greyscale symbols
- Oxford symbols
- Oxford colorscale symbols
- CFG and Oxford
- IUPAC

閲覧(Oxford 形式)

Glycan List

Range of the Mass



1326 Glycans

LIST • STRUCTURE

sorted by: [mass](#); [then by...](#) • grouped as sorted

1 • 2 • 3 ... 67 Next »

Tag

- [ambiguous](#) (152)
- [Epitope](#) (1159)
- [full](#) (1326)
- [Glycosphingolipid](#) (285)
- [N-Glycan](#) (889)
- [O-Glycan](#) (152)
- [Subclass](#) (51)

Motif

- [Blood group A](#) (43)
- [Blood group B](#) (2)
- [Blood group H](#) (157)
- [Galalpha1-3Gal epitope](#) (178)
- [Glycosphingolipid Arthro series](#) (16)
- [Glycosphingolipid](#)

1.

Accession Number : [G67988RO](#)

Image : 

Mass : 342.116211546

Motif : Glycosphingolipid Gala series

Classification : full and Glycosphingolipid

Contributor : Administrator

Contribution time : Sun, Jul 27, 2014, 11:40 pm

2.

Accession Number : [G39461CL](#)

Image : 

Number of Fuc



Number of Gal



Number of GalNAc



Number of Glc



Number of GlcNAc



Number of Man



検索方法

- 糖鎖構造の検索:
 - ID番号より (accession number search)
 - 構造より (完全一致、部分構造一致)
 - 単糖組成より
 - モチーフより

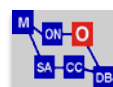
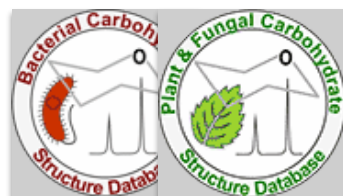


SHORT REPORT

Open Access

Introducing glycomics data into the Semantic Web

Kiyoko F Aoki-Kinoshita¹, Jerven Bolleman², Matthew P Campbell³, Shin Kawano⁴, Jin-Dong Kim⁴, Thomas Lütteke⁵, Masaaki Matsubara⁶, Shujiro Okuda^{7,8}, Rene Ranzinger⁹, Hiromichi Sawaki¹⁰, Toshihide Shikanai¹⁰, Daisuke Shinmachi¹⁰, Yoshinori Suzuki¹⁰, Philip Toukach¹¹, Issaku Yamada⁶, Nicolle H Packer³ and Hisashi Narimatsu^{10*}



WURCS: The Web3 Unique Representation of Carbohydrate Structures

Kenichi Tanaka,^{†,‡,▽} Kiyoko F. Aoki-Kinoshita,[§] Masaaki Kotera,^{||} Hiromichi Sawaki,[†] Shinichiro Tsuchiya,[§] Noriaki Fujita,[†] Toshihide Shikanai,[†] Masaki Kato,[⊥] Shin Kawano,[#] Issaku Yamada,^{*‡} and Hisashi Narimatsu[†]

[†]Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8568, Japan

[‡]The Noguchi Institute, Itabashi, Tokyo 173-0003, Japan

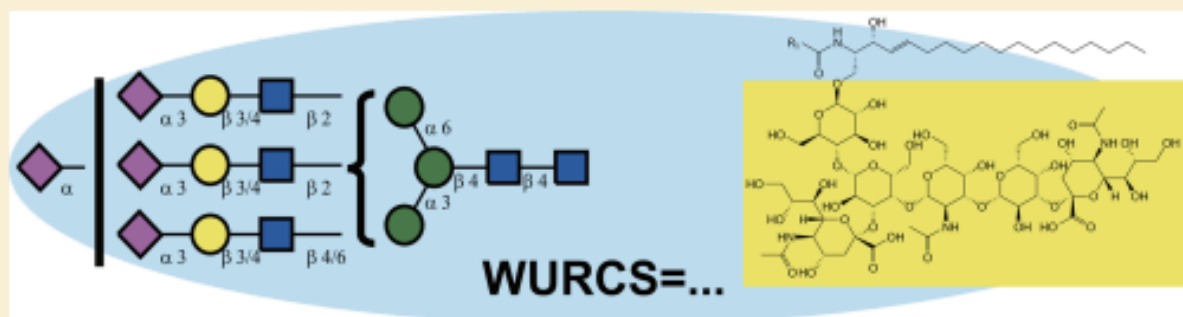
[§]Department of Bioinformatics, Faculty of Engineering, Soka University, Hachioji, Tokyo 192-8577, Japan

^{||}Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152-8550, Japan

[⊥]Structural Glycobiology Team, RIKEN Global Research Cluster, Wako, Saitama 351-0198, Japan

[#]Database Center for Life Science, Research Organization of Information and Systems, Kashiwa, Chiba 277-0871, Japan

Supporting Information



ABSTRACT: In recent years, the Semantic Web has become the focus of life science database development as a means to link life science data in an effective and efficient manner. In order for carbohydrate data to be applied to this new technology, there are two requirements for carbohydrate data representations: (1) a linear notation which can be used as a URI (Uniform Resource Identifier) if needed and (2) a unique notation such that any published glycan structure can be represented distinctively. This latter requirement includes the possible representation of nonstandard monosaccharide units as a part of the glycan structure, as well as compositions, repeating units, and ambiguous structures where linkages/linkage positions are unidentified. Therefore, we have developed the Web3 Unique Representation of Carbohydrate Structures (WURCS) as a new linear notation for representing carbohydrates for the Semantic Web.

糖鎖関連データベース

UniCarbKB

GlycomeDB

Carbohydrate Structure Database
Plant & Fungal Carbohydrate Structure Database

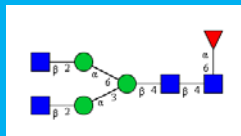
9-1-9-C-0-SCIENCES.DE

JCGGDB DB

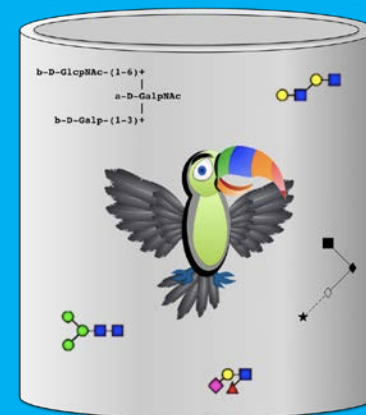
M ON-O
SA-CC DB



糖鎖ID + 構造データ



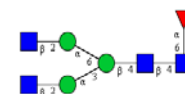
⑤ データ共有



① 登録

糖鎖研究の知識共有

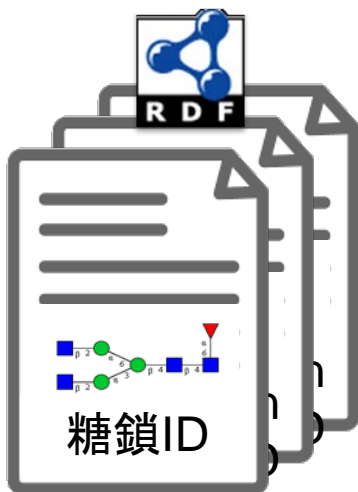
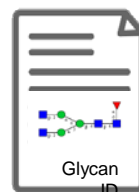
② 糖鎖ID発行



研究者

研究論文

③ 投稿



糖鎖ID

学術雑誌

④ アノテーション情報のキュレーション



まとめ

- 国際糖鎖構造リポジトリ GlyTouCanが公開された
- 今後の予定：
 - GlyTouCanの機能の改善：
 - GlycoCT以外の形式による登録
 - 描画ツールによる登録・検索
 - 言語の選択(日本語、中国語、ロシア語、ドイツ語など)
 - GlyTouCanのデータの自動RDF化
 - UniProt(タンパク質)やPubChem(化合物)データとのリンク
 - JCGGDBのRDFデータをGlycoInfoのエンドポイントに公開
 - ジャーナルや海外データベースとの連携
- セマンティックウェブを通して、遺伝子やタンパク質などの関連情報との統合が可能になる

糖鎖関連データベース 2017?

