日本糖鎖科学コンソーシアムデータベース（JCGGDB）

○ソロビヨワ・イェレナ[1)]、藤田典昭[1)]、鹿内俊秀[1)]、鈴木芳典[1)]、木下聖子[1),2)]、成松久[1)]
[1)]産業技術総合研究所 糖鎖技術研究グループ　[2)]創価大学 理工学部

概要

産総研・糖鎖技術研究グループは文部科学省委託研究開発事業やJST/NBDCの「ライフサイエンスデータベース統合化推進プログラム」に参加し、糖鎖関連の多種多様なデータベースを構築し、糖鎖関連データベースの統合化を推進してきた。2010年から公開されている糖鎖科学コンソーシアムのデータベース（JCGGDB; http://jcggdb.jp）を活用している。我々は、SKOS、OWL、RDFなどのセマンティックWeb技術を用いて、それらのデータベースのコンテンツのRDF化とオントロジーの開発を行っている。 現在、RDFの形式で公開されているデータベースはタンパク質の糖鎖修飾位置のデータベース（GlycoProtDB）、レクチンと糖鎖の相互作用のデータベース（LfDB）、糖鎖遺伝子データベース（GGDB）、糖鎖関連の遺伝性疾患とその原因遺伝子のデータベース（GDGDB）、糖鎖を認識する病原体の相互作用のデータベース（PACDB）である。また、実験で得られた糖鎖修飾位置についているグライコフォームの情報をGlycoProtDBデータベースの拡張として公開している。我々は、RDF化したJCGGDBのデータベース群と開発したオントロジーをAsian Community of Glycoscience and Glycotechnology（ACGG）の情報リソースとしてhttp://acgg.asia/db/から提供している。

結論

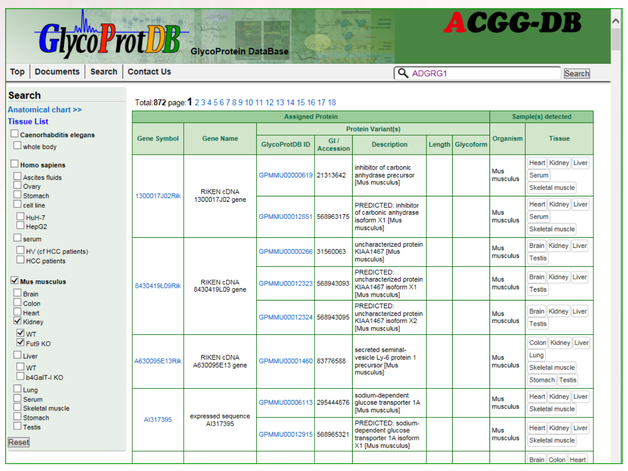
利用者により分かりやすい情報提供を行うためには、糖鎖研究側からの観点から糖鎖に関する様々な情報を分類し階層化しオントロジーを構築した。作成したオントロジーを利用して糖鎖関連のデータをRDF化した。既存のJCGGDBのデータをRDF化することで、共通する用語（構造情報、遺伝子名、化合物名など）を利用して外部のオントロジーとの連結が可能になった。 糖鎖関連データベースに蓄積されている情報の理解をより深めるために、データの検索や階層構造の閲覧、詳細表示ができるユーザインタフェースの開発と提供を行い公開した。

RDF化された情報は糖鎖関連の分野で研究する研究者の知識の理解と整理を容易いものとし、セマンティックweb化した内外のデータと結びつくことで新しい知識発見に役立つと考えられる。

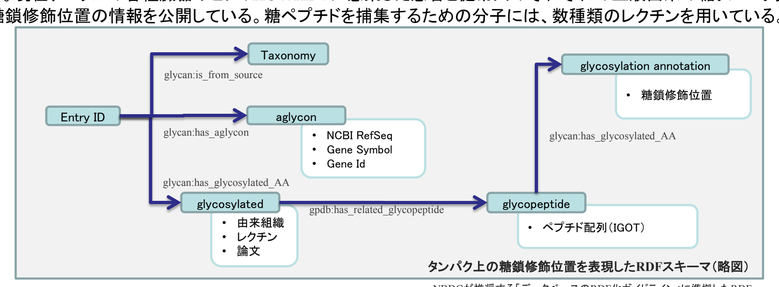
Acknowledgments

This work has been supported by National Bioscience Database Center (NBDC) of Japan Science and Technology Agency (JST).

GlycoProtDB: GlycoProtein DataBase http://acgg.asia/db/gpdb　糖タンパク質の糖鎖修飾位置とグライコフォームの情報

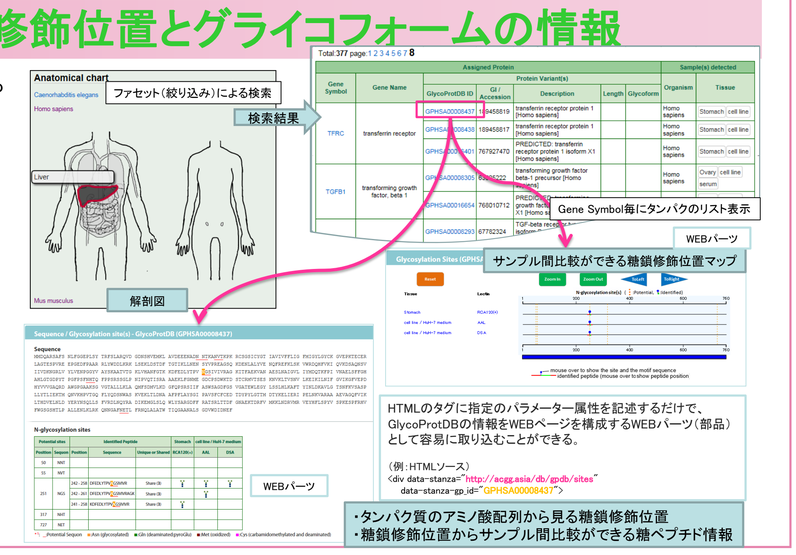


本データベースに格納されているデータは、糖鎖付加部位を同定できる方法（IGOT法）から得られた実験データである。現在、マウスの各種臓器やヒトのHCV/HBVに感染した患者と健常人のそれぞれの血液由来の糖タンパク質の糖鎖修飾位置の情報を公開している。糖ペプチドを捕集するための分子には、数種類のレクチンを用いている。

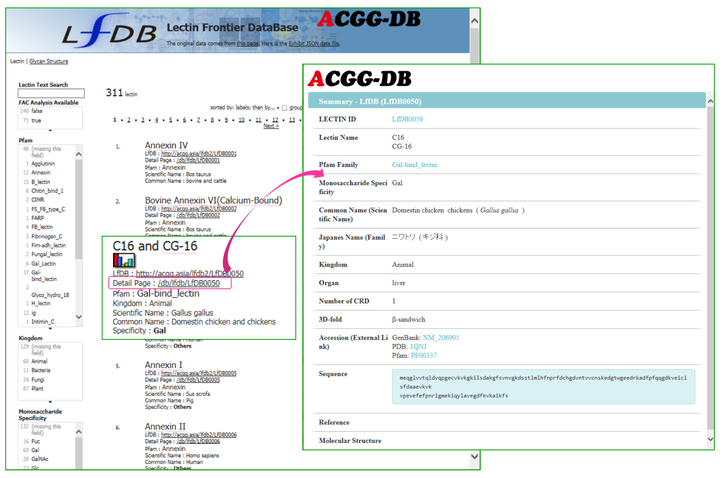


NBDCが推奨する「データベースのRDF化ガイドライン」に準拠したRDF

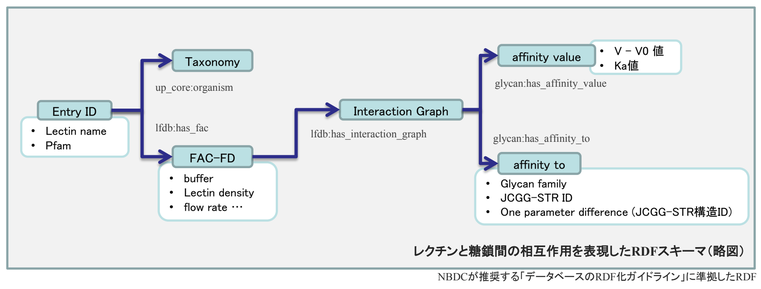
Taxonomy	IGOT sample	IGOT peptide	Genes	gi	N-Glycosylation sites
C.elegans	3	624	701	986	1,808
Mus musculus	32	2,135	1,878	6,437	15,398
Homo sapiens	22	941	531	1,828	3,451



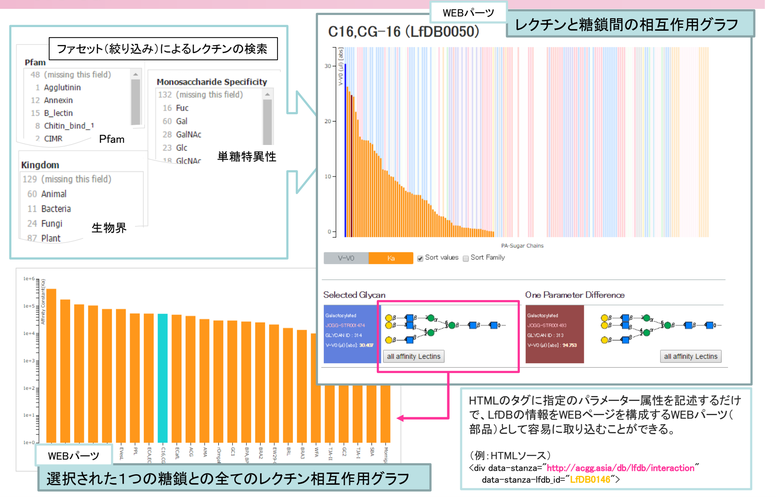
LfDB: Lectin Frontier DataBase http://acgg.asia/db/lfdb　レクチンと糖鎖の相互作用の情報



本データベースには、蛍光検出を用いた自動化フロンタルアフィニティークロマトグラフィーシステム(FAC-FD)を用いて取得した各種レクチンとピリジルアミノ化糖鎖間の結合定数を含む定量的相互作用データを格納している。（現在、311件のレクチン情報、71件のFACデータ）



NBDCが推奨する「データベースのRDF化ガイドライン」に準拠したRDF



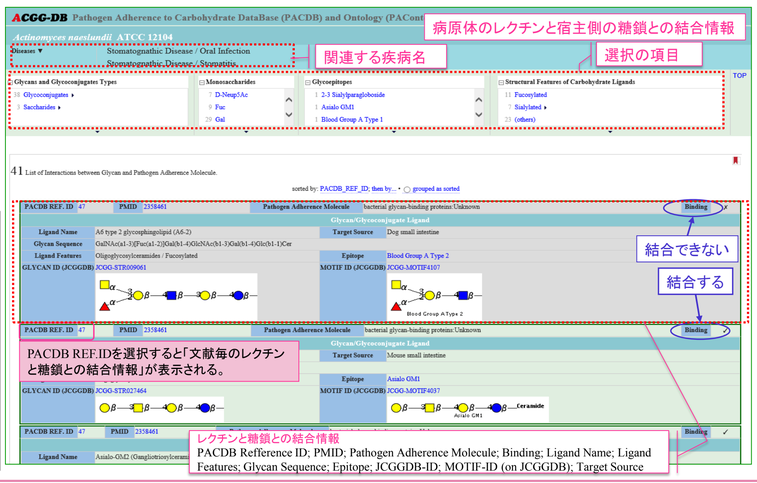
PACDB: Pathogen Adherence to Carbohydrate DataBase http://acgg.asia/db/diseases/pacdb　糖鎖と病原体との相互作用に関する情報

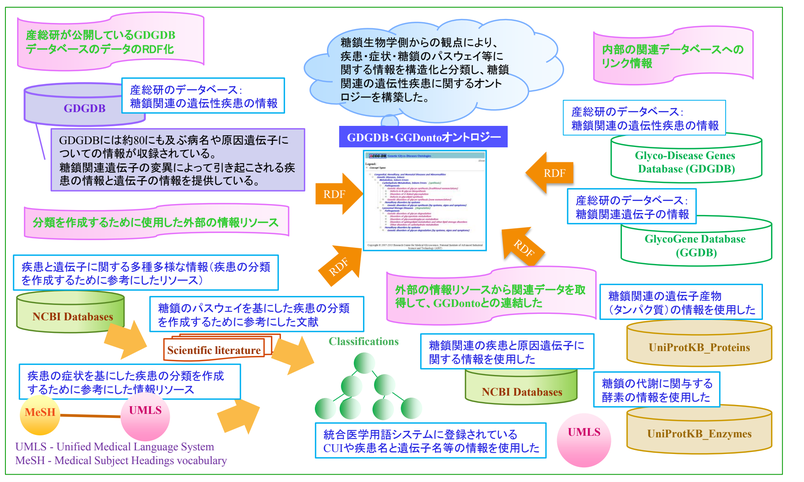
PAConto: PACDBデータのRDF表現及び糖鎖関連の感染症に関するオントロジー

本データベースには、97の疾患名と127の病原体（370の株）についての情報が214の学術論文から抽出され、約1700件の糖鎖と病原体の相互作用のデータが収録されている。

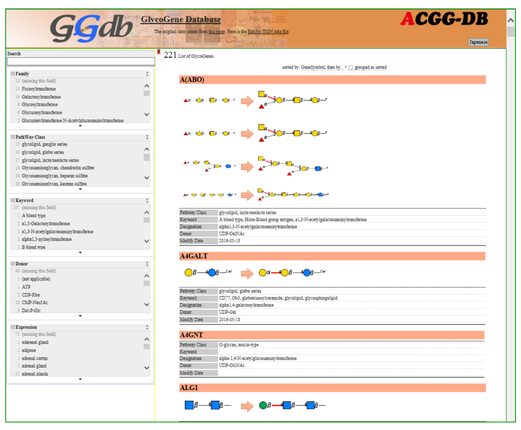
感染症を引き起こす病原体のタンパク質（レクチン）とそのレクチンが結合する宿主（主にヒト）側の糖タンパク質や糖脂質（糖鎖リガンド）の情報が提供されている。

特徴として、実験で結合しない糖鎖構造の結果も収集しているので、実験していないものとの区別ができる。





GGDB: GlycoGene DataBase http://acgg.asia/db/ggdb　糖転移酵素の基質特異性の情報



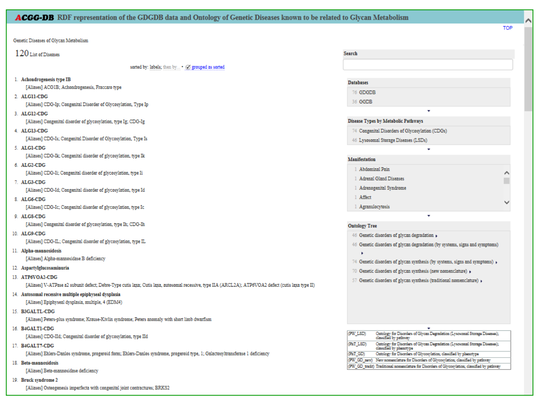
本データベースの特徴

- GlycoGeneの遺伝子情報
- 糖鎖合成遺伝子の基質特異性の情報
- ヒトの遺伝子をベースにした糖鎖関連のオーソログ遺伝子
- 遺伝子ライブラリーリソースの提供

現在、約190のヒト糖鎖関連遺伝子（糖転移酵素・糖ヌクレオチド輸送体・硫酸基転移酵素・糖ヌクレオチド合成酵素など）が同定されて、アノテーション付けされている。
多くの遺伝子リストから目的とする対象を抽出するため、ファミリー・ドナー基質・合成する糖鎖の種類などの項目を利用してファセット検索で絞り込めるユーザインタフェースを開発した。

詳細はポスター（No.33）を参照ください。

GDGDB: Glyco-Disease Genes DataBase http://acgg.asia/db/diseases/gdgdb　遺伝性疾患の関連情報



GDGDB RDF & GGDonto ontology
GGDonto:糖鎖関連の遺伝性疾患に関するオントロジー

本オントロジーの特徴

- 参考文献を基にして、糖鎖関連の遺伝子性疾患に関連した「疾患の病態に基づいた分類」と「原因遺伝子の欠陥によって生じる糖鎖代謝の異常に基づいた分類」を2つ作成し、オントロジーを構築した。
- 糖鎖関連の遺伝子性疾患についての情報は多様な医学関連の情報リソースに掲載されているため、本オントロジーと外部の情報リソースとのリレーションを多く持つ。

詳細はポスター（No.33）を参照ください。

Licensed under a Creative Commons表示4.0国際 (c)2016 産業技術総合研究所