35

糖鎖構造抽出ソフトウェアを用いた化合物データベースとの連携

○松原 正陽[1]、木下 聖子[2]、山田 一作[1]

1. 公益財団法人 野口研究所 糖鎖情報科学研究室、2. 創価大学 糖鎖生命システム融合研究所

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖関連データベースの連携

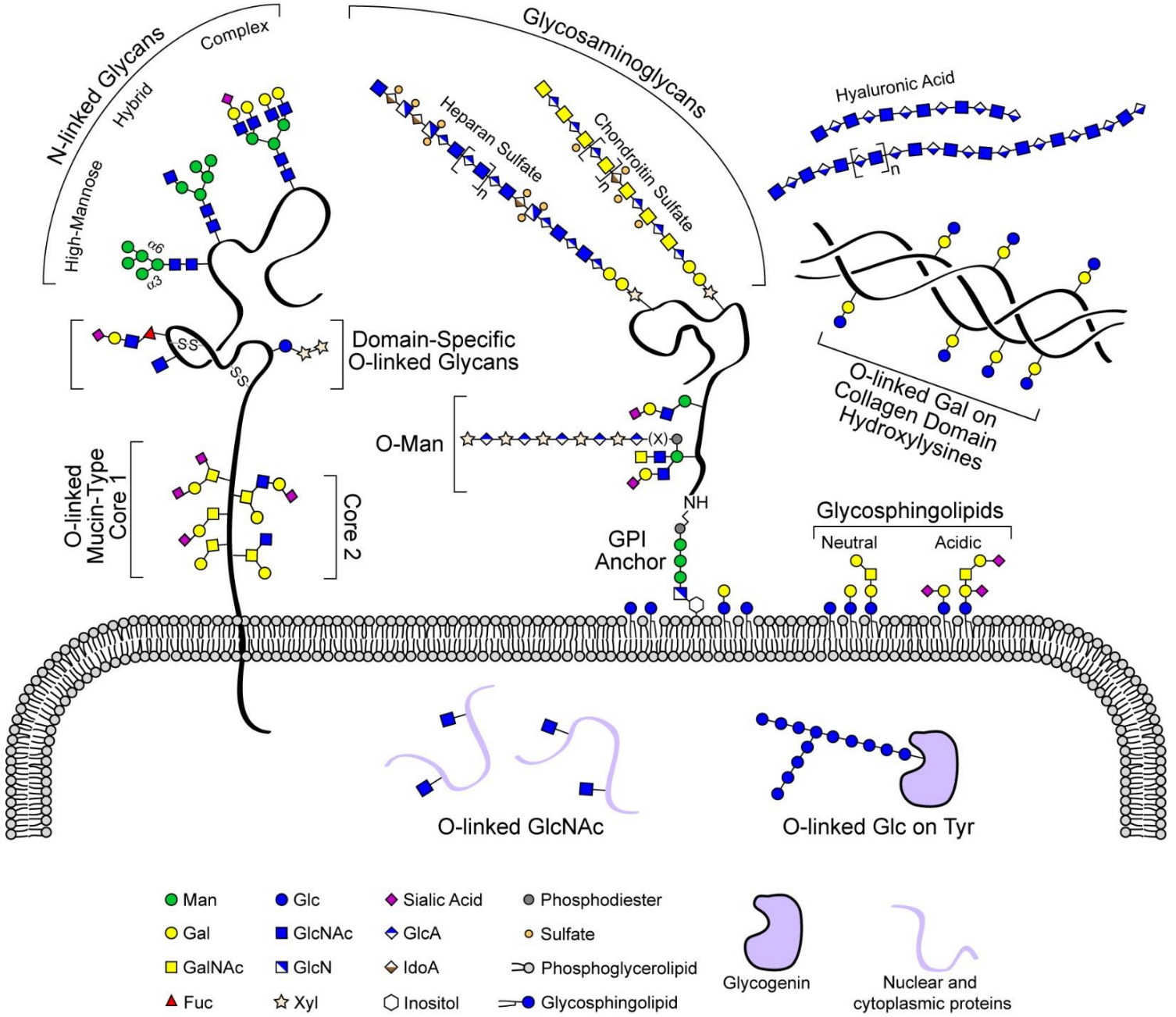
Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖とデータベース

- 糖鎖とは
  - DNA、タンパク質に次ぐ第三の生命鎖
  - タンパク質や脂質を修飾し、複合糖質として様々な形態をとる
  - 修飾する他分子の種類やその形態に応じて様々な機能を担う
  - 糖鎖の組立や他分子への修飾には、様々な糖転移（消化）酵素が関連

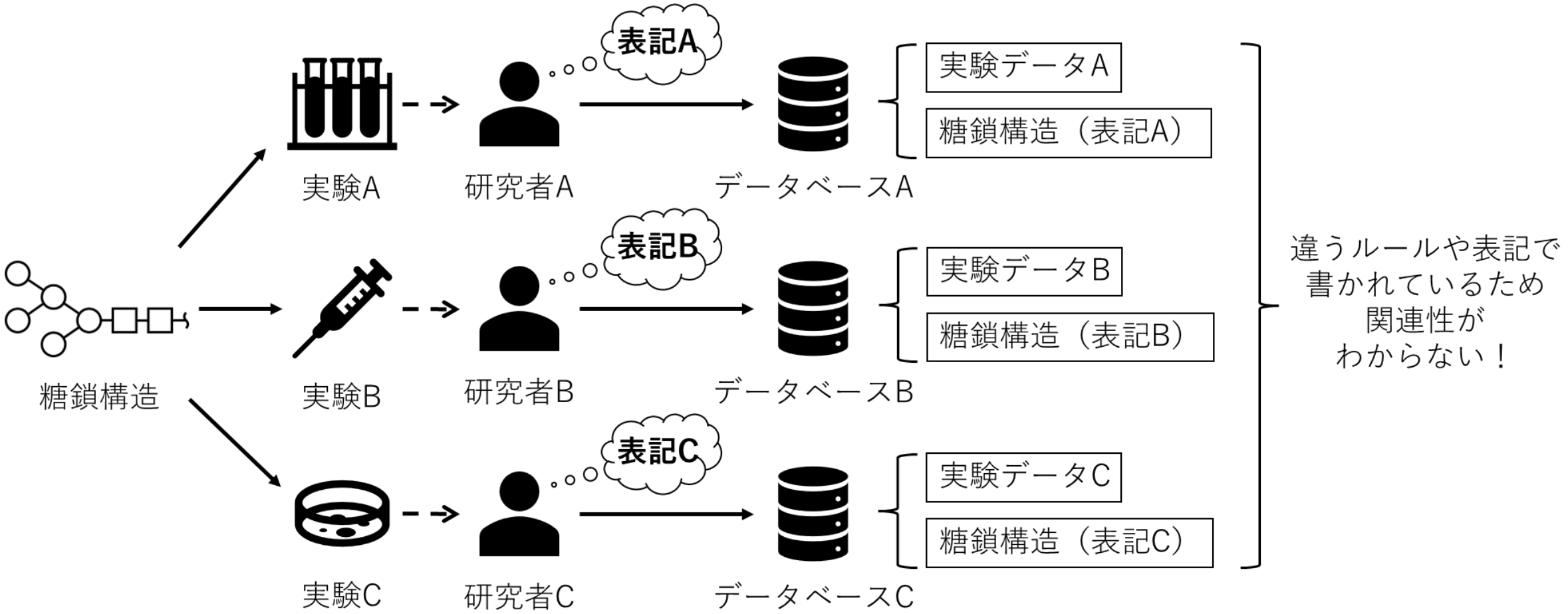
- 糖鎖関連データベースの一例とその活用例
  - 糖転移（消化）酵素のデータベース
    - 糖鎖とその組立に関与する酵素タンパク質との関連
  - 複合糖質の構造解析結果のデータベース
    - 糖鎖と糖鎖が修飾するタンパク質や脂質との関連
  - →DNAやタンパク質、脂質のデータベースとの連携により、糖鎖の発現や生体機能への関与などの理解につながる

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖関連データベース連携における これまでの課題

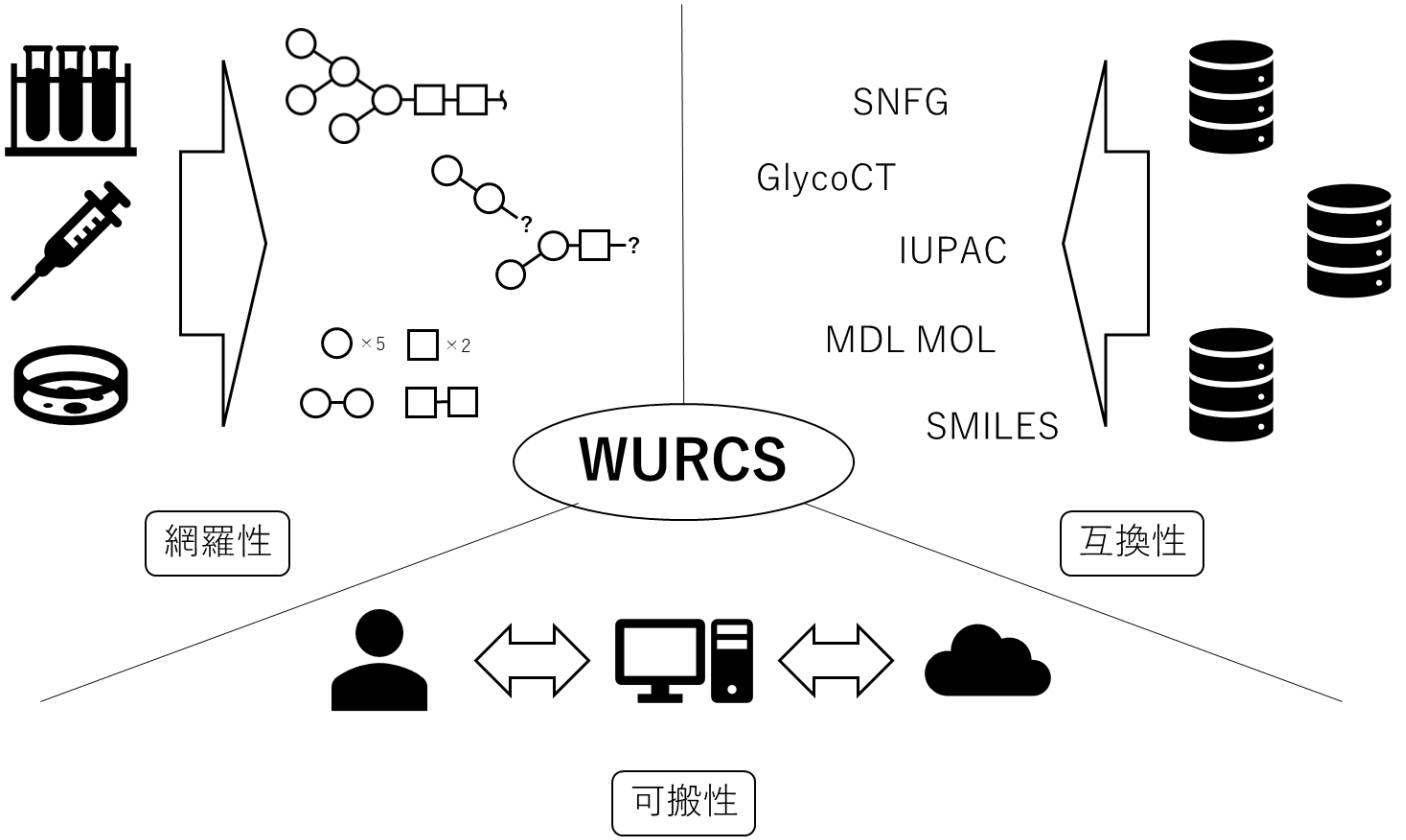


Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖構造表記法WURCSの開発

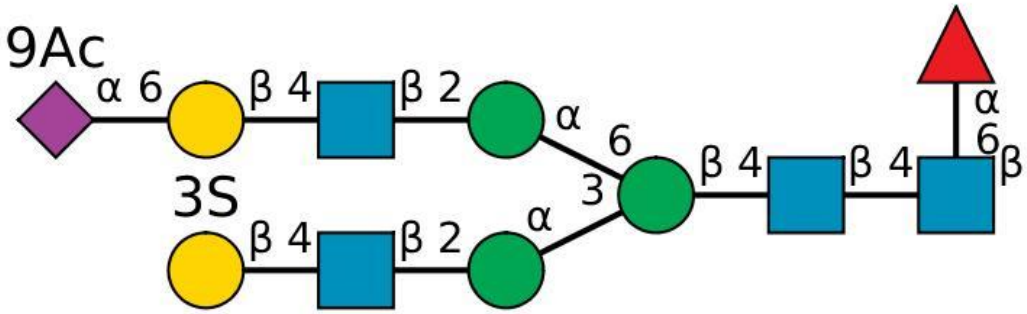
- WURCS
  - Web3.0 Unique Representation for Carbohydrate Structures
  - 糖鎖構造を一意な文字列で表記
  - データベース間で表記の異なる糖鎖構造を統一

- WURCSの特徴
  - 様々な糖鎖構造を**網羅**
  - 他の構造表記形式と**互換**
  - データとして**可搬**



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖構造表記法WURCS



WURCS=2.0/7,11,10/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_3*OSO/3=O/3=O][a2112h-1b][Aad21122h-2a_2-6_5*NCC/3=O_9*OCC/3=O][a1221m-1a_1-5]/1-1-2-3-1-4-3-1-5-6-7/a4-b1_a6-k1_b4-c1_c3-d1_c6-g1_d2-e1_e4-f1_g2-h1_h4-i1_i6-j2

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

GlyTouCanによるデータベース連携

- 糖鎖リポジトリGlyTouCan
  - 糖鎖構造に対し固有のIDを発行
  - 糖鎖構造をWURCSで管理
  - 糖鎖関連DBの糖鎖構造をWURCSに変換し、IDを発行
    →GlyTouCanを中心に、DB間の糖鎖関連情報を相互連携

- 糖鎖科学ポータルGlyCosmos
  - 糖鎖だけでなく、糖鎖に関連する様々な情報を集約、管理

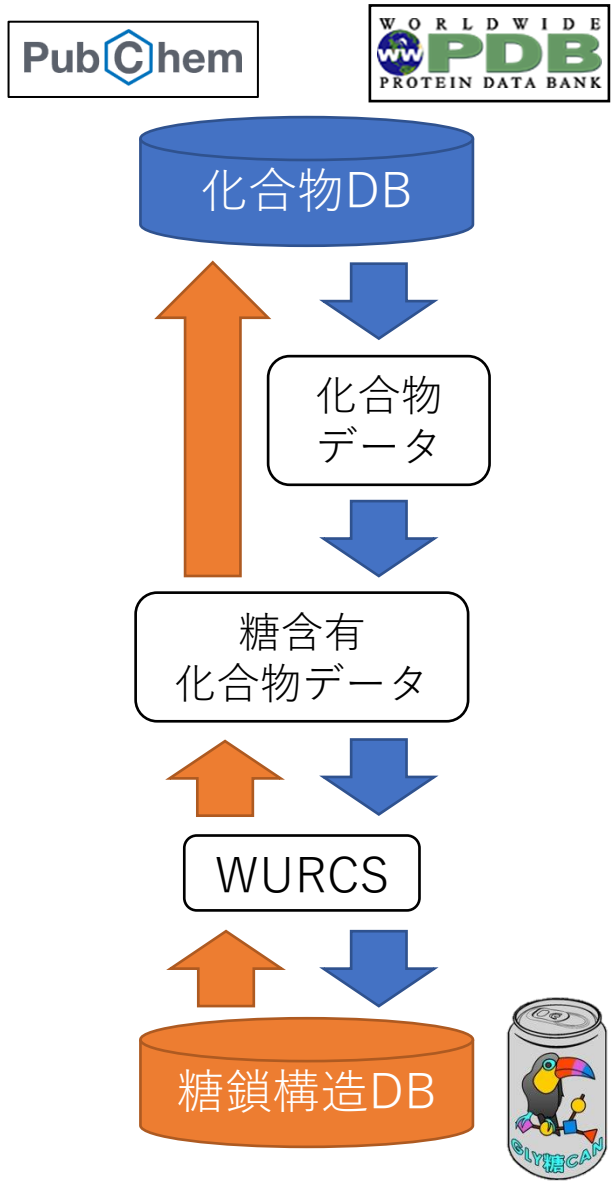
Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

化合物データベースとの連携

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

化合物データベースとの連携

- 化合物データベース
  - 構造を化学構造式で登録
  - 糖や糖鎖を含む化合物も登録されている
- データベース連携の意義
  - 糖鎖構造の化合物としての特性を含む、
    生物学分野の研究だけでは得られない情報が得られる
    - 可能な立体構造、分子間相互作用、合成の可否、など
- データベース連携における課題
  - 糖や糖鎖を含む化合物のみを連携する
  - データベース間で異なる糖鎖構造の表現形式を相互変換
    - 化学構造式 ↔ WURCS
  - 膨大な数の化合物を自動的かつ高速に処理できるシステム

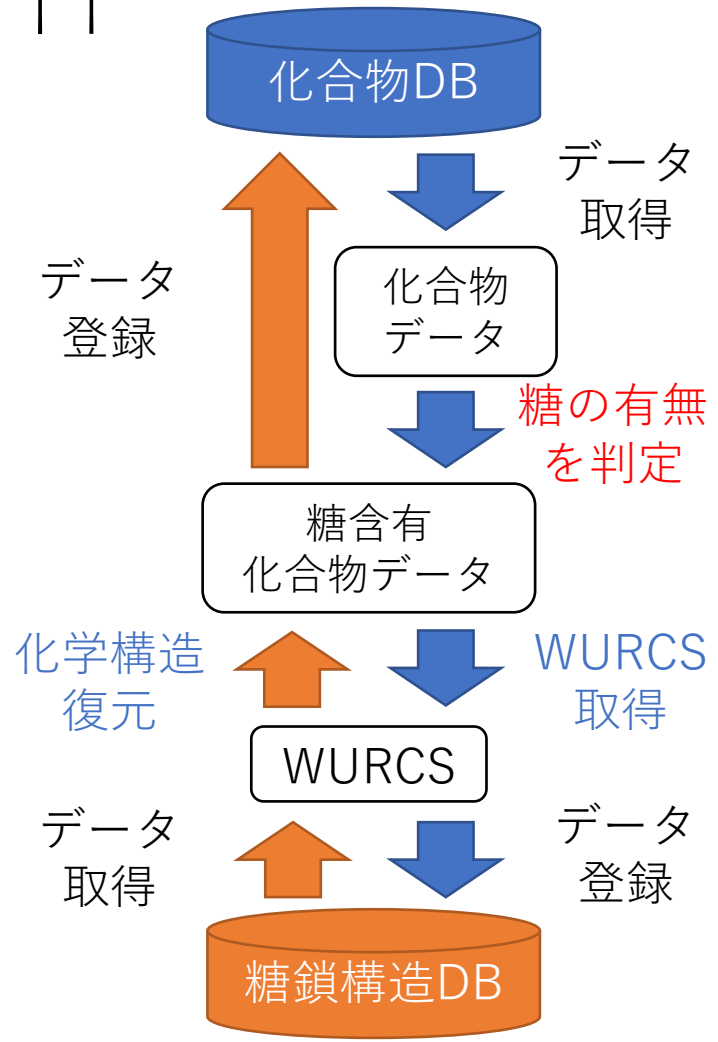


Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖構造抽出ソフトウェアの要件

- WURCS-化学構造相互変換機能
  - 化学構造とWURCSの相互変換のための基本的な機能を実装
  - 化学構造 → WURCS：
    化合物の化学構造の読み込み、糖鎖構造の抽出、WURCSの取得
  - WURCS → 化学構造：
    WURCSから化学構造を復元、化学構造フォーマットで出力

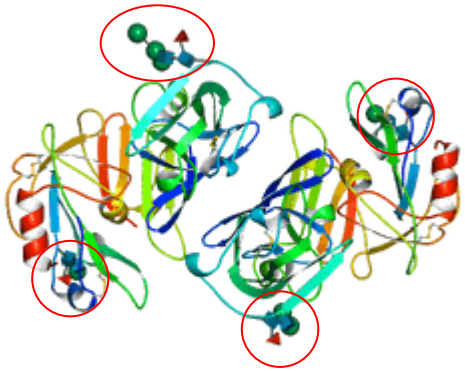
- 糖鎖構造抽出ルールの検討
  - 糖が化合物に含まれるかどうかを適切に判断するための独自のルールを開発



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

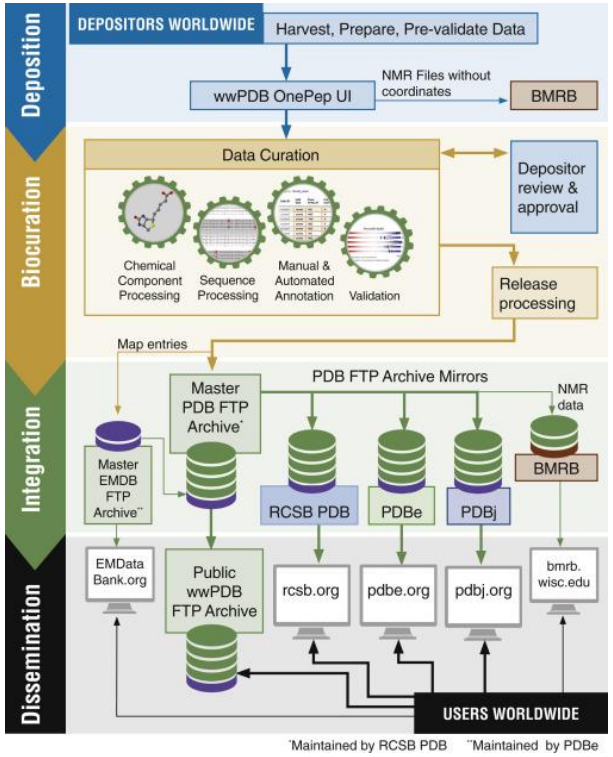
これまでの成果：WURCSによるPDBとの連携

- Carbohydrate Remediationの一環として連携
  - 糖鎖構造用のデータ表現を新たに追加
    - 単糖用のモノマー（CCD）タイプ
    - 枝分かれ構造用のデータデータカテゴリ
    - 糖鎖構造記述子用のデータカテゴリ
  - 糖鎖構造情報の取得が容易に



- PDBデータからWURCSを生成するソフトウェアPDB2Glycanを開発
  - PDBの単糖のモノマーや枝分かれ構造データを元にWURCSを生成
  - PDB2Glycanを組み込んだOneDep[1]システムで、生成されたWURCSを自動登録

- PDBjによるGlyTouCanとの連携が実現
  - 糖鎖エントリーが持つWURCSを使ってGlyTouCanエントリーを連携



Overview of the OneDep System[1]

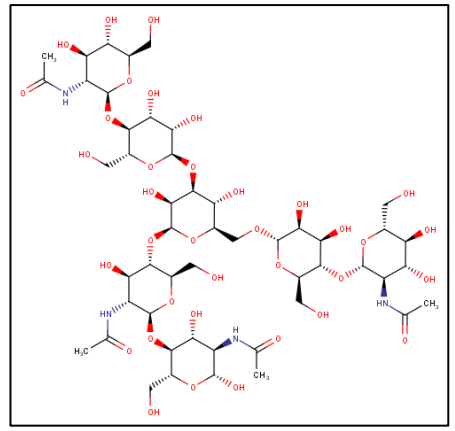
1 Young JY *et al*, OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive, Structure. 2017 Mar 7;25(3):536-545. doi: 10.1016/j.str.2017.01.004. Epub 2017 Feb 9.

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

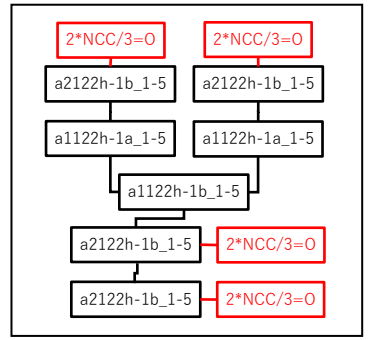
化合物データベースとの連携における課題

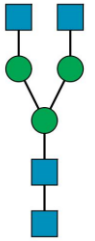
- 糖や糖鎖を含む化合物のみを連携したいが、、、
  - **化合物データベースは**一般的に**糖や糖鎖を明示的に扱わない**
    - PDBにおけるCarbonhydrate Remediationのようなプロジェクトがない限り
  - 少なくとも、化学構造式やその特徴に関する情報は存在
  - →**化学構造式から糖や糖鎖を判定・抽出する仕組みが必要**

- そもそも**糖や糖鎖に対する統一的な判定基準がない**
  - 単糖には様々な修飾を持つものや誘導体が存在
  - 単糖の定義が研究分野によって異なる
  - IUPACは構造に対する命名規則を定義しているが、
    単糖や糖鎖が具体的にどのような構造かまでは言及していない
  - →**適切に糖鎖構造を判定する新たなルールの開発が必要**



化合物（化学構造式）
- 原子
- 結合





糖鎖（WURCS）
- 単糖
- 置換基
- 結合情報

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

糖鎖構造抽出ルールの開発

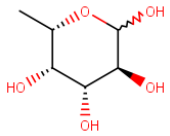
- 基本方針
  - 単糖の標準構造を定義し、その構造から大きく逸脱しないものを単糖と判断する
  - 単糖と判断されなかった構造はすべて置換基とみなす
  - 置換基のうち、糖鎖の一部としての基準を満たさないものは除外する
  - 置換基によって結合している単糖同士は一つの糖鎖の構成要素として扱う
  - 単糖を一つも含まないものは抽出の対象外とする

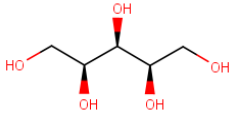
- 糖鎖抽出手順
  - 炭素鎖を単糖骨格にふさわしいかどうか判定
    - 炭素鎖の長さや分岐の数、修飾する原子団の種類や数で判断
  - 単糖骨格以外の原子団が置換基として基準を満たすかどうか判定
    - 特定の原子種を含むかどうかや、原子数、分岐の数等が規定を超えるかで判断
  - 炭素鎖と置換基を組み合わせて糖鎖構造とし、WURCSとして出力

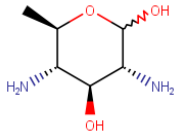
Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

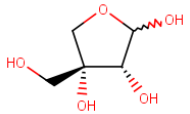
結果：判定の一例

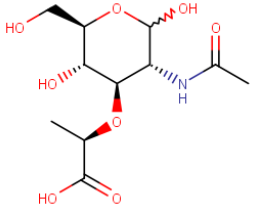
- 単糖と判断されるもの

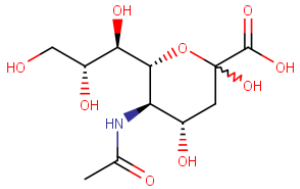

Fucose

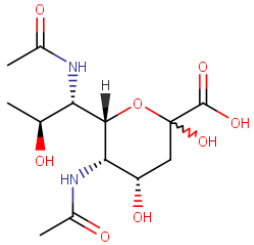

Xylitol


Bacillosamine

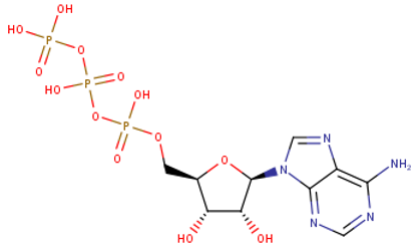

Apiose

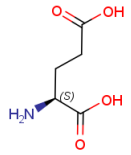

N-acetyl-muramic acid

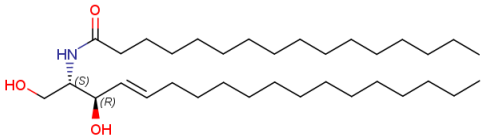

N-acetyl-neuraminic acid

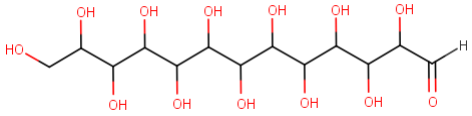

Pseudaminic acid

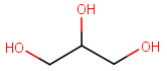
- 単糖と判断されないもの

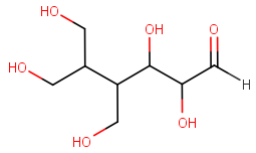

Adenosine triphosphate
(DNA/RNA)

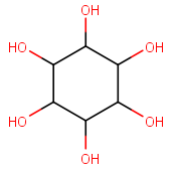

Glutaminic acid
(Amino acids)

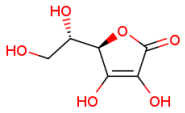

Ceramide
(Lipids)


Too long


Too short


Too many
branches


Carbon cyclic


Too many
modifications

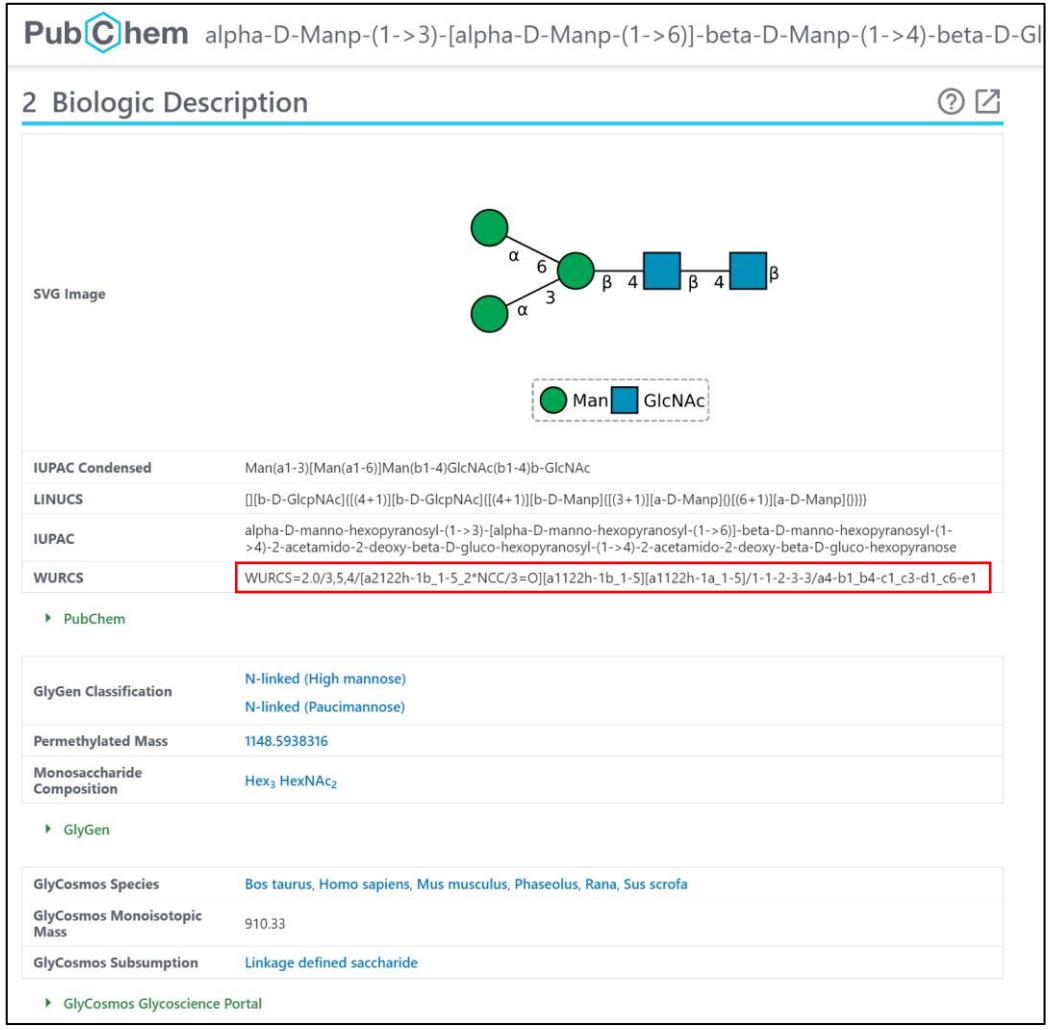
Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

結果：PubChemとのデータベース連携

- 我々が開発した糖鎖構造抽出ソフトウェアをPubChemに提供

- PubChemの化合物から変換されたWURCSが、Biologic Descriptionの一つとして登録
  - 現在、PubChemの化合物エントリーから確認可能

- PubChem-GlyTouCan間の連携によって、糖鎖構造データベースの情報も取得可能に
  - GlyGenやGlyCosmosなどが扱っている生物種や質量の情報など



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）

今後の予定

- 引き続き、PubChemとのデータベース連携を進めていく
  - より多くの化合物を連携するために、今後連携すべき化合物に関する条件をPubChemと協議して決めていく
  - WURCSと一致しなかった化合物については、現在開発中の複合糖質データベースで連携する
- その他の化合物データベースとの連携を検討する
  - 糖脂質データベースLIPID MAPSなど
- MolWURCSの変換機能を改善・修正する
- 糖鎖構造抽出ルールをPubChem以外のデータでも検証し、論文化を進める

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　松原 正陽（公益財団法人 野口研究所 糖鎖情報科学研究室）