

○下山紘充¹, 松原陽明¹, 山田一作¹

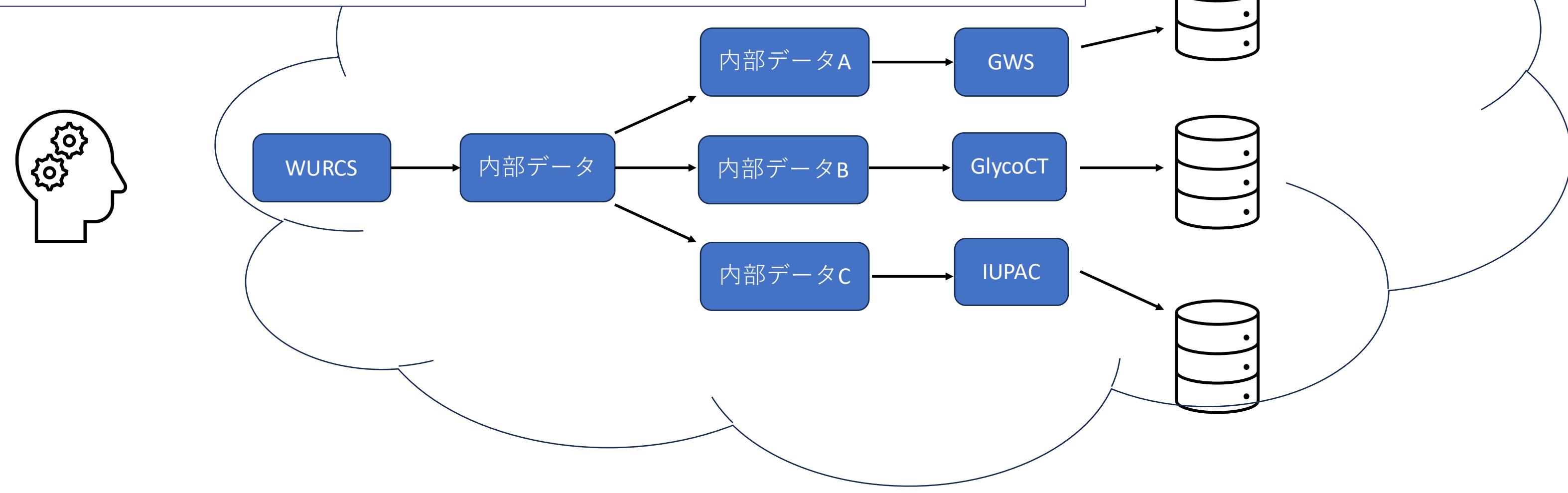
¹公益財団法人 野口研究所

目的・背景 Introduction

糖鎖構造を表現するテキスト表記には様々な形式があり、糖鎖データベース (DB) も目的によって様々な表記を採用している。多様なDBを包括的に活用するためにはこれらの表記を相互に (機械的に) 変換できる仕組みが必要である。また表記ごとに表現できない構造があるため、なるべく情報がこぼれないよう変換したい。そこで内部的にWURCS表記の情報を保持し、適宜変換するプログラムの作成を行っている。WURCSは複雑な構造にも対応できるため、内部データとして適しているという判断である。WURCSを介してDBを包括的に利用する試みの進捗として、WURCSからSNFG (GWS) に変換するプログラムを紹介する。

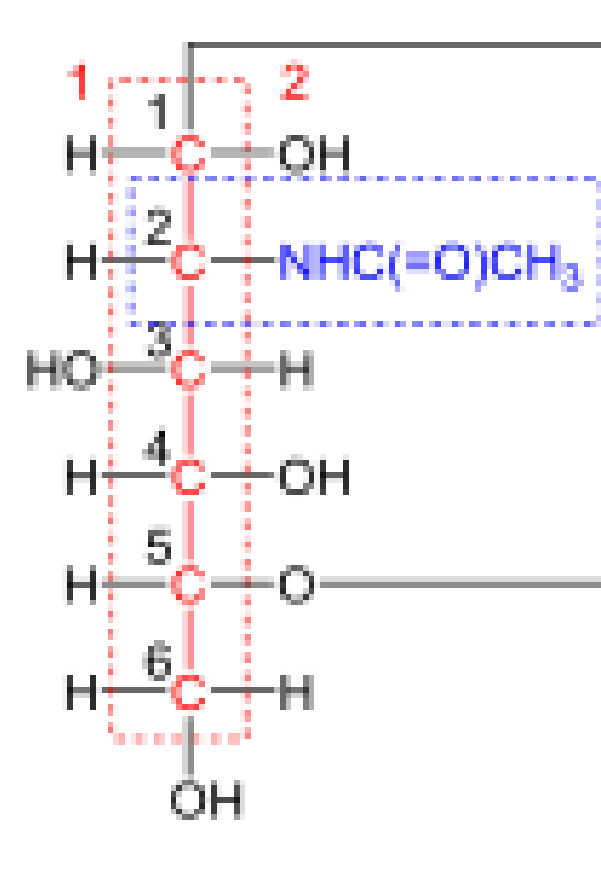
データベースの包括的検索

内部的にWURCSと等価なデータを保持しておき、別の表記に変換する。DBが別々でも、入力をDBに対応する表記に変換し、機械的に検索すれば、人間から見れば1つの表記のデータベースとして包括的に運用できる。難しい対応関係に悩む必要がない。



WURCS[1]

Web3 Unique Representation of Carbohydrate Structures (WURCS)
他のテキスト形式が対応していなかった複雑な糖鎖構造のルールを定義されており、複雑な糖鎖を一意に表現することができる。
(注) ただし、研究者が直接編集するにはWURCSは複雑な形式なので、IUPACや図形であるSNFGなどで検索データの入力を行う予定。



Substituent (置換基)

\rightarrow *NCC/3=O

SkeletonCode (骨格)

\rightarrow a2122h

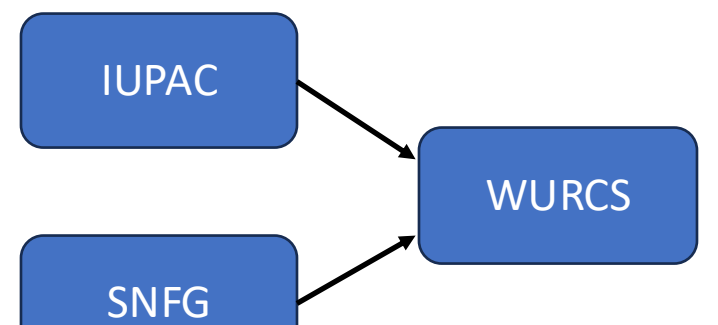
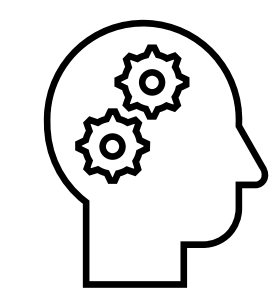
functional group Character

H-C-O o

CH₂OH h

O-C-H 1

H-C-O 2



一旦WURCSにしてしまえば、様々な形式に変換できる。

GlcNAcのフィッシャー投影図とWURCSの関係 ([2]より改変して抜粋)

WURCSの例

N-glycan (全てを一行で書く)

WURCS=2.0/3,5,4/

[a2122h-1b_1-5_2*NCC/3=O]

[a1122h-1a_1-5]

[a1122h-1a_1-5]/

1-1-2-3-3/

a4-b1_b4-c1_3-d1_6-e1

対応

炭素鎖への修飾は「2*NCC/3=O」のように「置換基が付いている炭素の番号*化学式」の形式で書く

1-5はpyranose

1-4はfuranose

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

対応

GWSへの変換例 (GWSからSNFGへはGB4W[3]を使用)

WURCS=2.0/3,5,4/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5]/1-1-2-3-3/a4-b1_b4-c1_c3-d1_c6-e1

\rightarrow freeEnd--?b1GlcNAc--4b1GlcNAc--4b1Man(--3a1Man)--6a1Man

O-glycan core 1

WURCS=2.0/2,2,1/[a2112h-1a_1-5_2*NCC/3=O][a2112h-1b_1-5]/1-2/a3-b1

\rightarrow freeEnd--?a1GalNAc--3b1Gal

Glycosphingolipid

WURCS=2.0/4,4,3/[a2122h-1b_1-5][a1122h-1b_1-5][a2122h-1b_1-5_2*NCC/3=O][a2112h-1b_1-5_2*NCC/3=O]/1-2-3-4/a4-b1_b3-c1_c4-d1

\rightarrow freeEnd--?b1Glc--4b1Man--3b1GlcNAc--4b1GalNAc

GAGs

WURCS=2.0/2,4,3/[a2122h-1b_1-5_2*NCC/3=O][a2122A-1b_1-5]/1-2-1-2/a3-b1_b4-c1_c3-d1

\rightarrow freeEnd--?b1GlcNAc--3b1GlcA--4b1GlcNAc--3b1GlcA

Milk sugar

WURCS=2.0/2,2,1/[a2122h-1x_1-5][a2112h-1b_1-5]/1-2/a4-b1

\rightarrow freeEnd--??1Glc--4b1Gal

変換方法

単糖の変換例: GlcNAc

a2122h-1b_1-5_2*NCC/3=O

2122 (骨格), 1b (α/β), 1-5 (リングサイズ), \leftarrow (鎖長) \rightarrow 3文字コード

2*NCC/3=O \rightarrow 2位のNAc

Glc+NAc=GlcNAc

同様にa2122A; A:acidのように炭素鎖の修飾があればGlc+A=GlcA

1-1-2-3-3/ : 番号は単糖の種類 ([...][...][...] \rightarrow [1番][2番][3番])

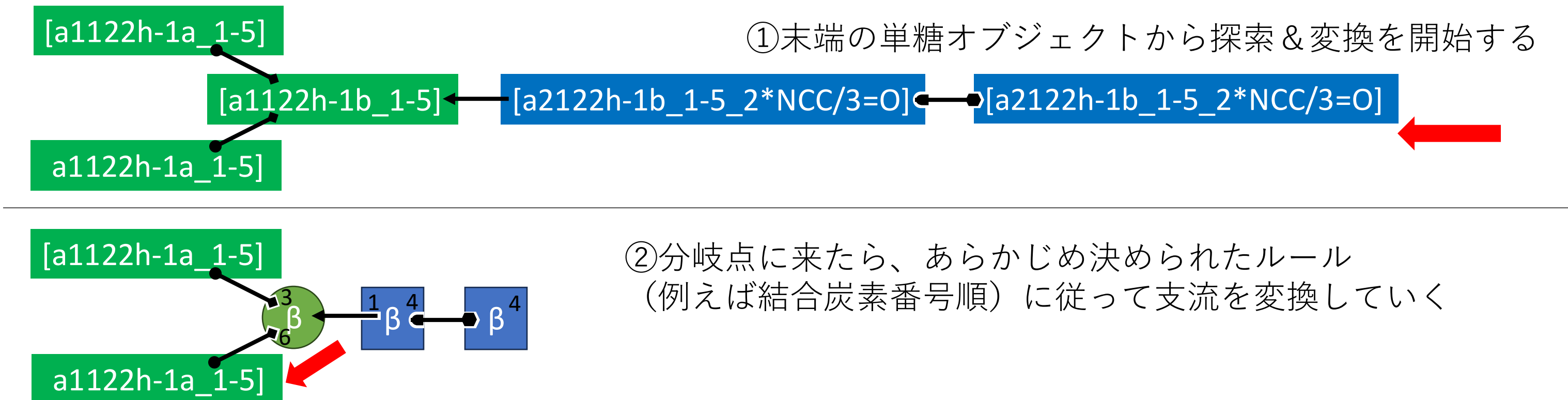
: 1-1-2-3-3はa-b-c-d-eに対応する。

a4-b1_b4-c1_3-d1_6-e1

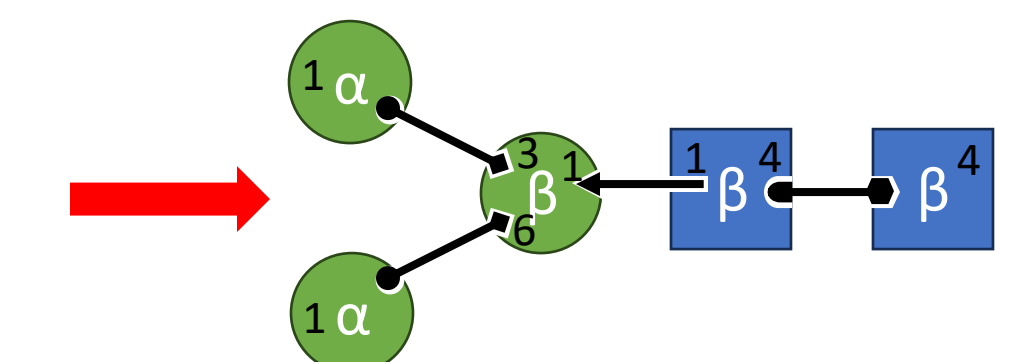
単糖同士の結合の変換: 親側と子側の炭素位置、および、グリコシド結合か否か (間に酸素以外を介するかなど) を持ったオブジェクト

ネットワークの探索

単糖のオブジェクトが以下のように並んでいるとする。
(実際のメモリ上では、SNFGの順にオブジェクトが並んでいるわけではないですが、簡単のため)



③全体を変換し終わると以下のような文字列(GWS)になる
freeEnd--?b1GlcNAc--4b1GlcNAc--4b1Man(--3a1Man)--6a1Man



まとめと今後、参考文献

基本的な糖鎖構造 (N型糖鎖、O型糖鎖、スフィンゴ糖脂質、グリコサミノグリカン、乳糖) についてWURCSからGWSへの変換プログラムを作成した。この機能を実装するにあたり作成した内部データ (オブジェクト) は他の表記に変換する際にも使用する予定である。今後は内部データから変換できる表記をGWS以外 (IUPAC、GlycoCT...) にも増やしていく所存である。また、今回は基本構造についての例を示したが、WURCSでしか書けないような複雑な構造 (繰り返しの構造や円環を形成する構造) などに対応する所存である。

参考文献

[1] Matsubara, Masaaki, et al. *J. Chem. Inf. Model.* 57.4 (2017): 632-637.

[2] 国際標準化糖鎖構造表記法WURCSの開発

[3] GlycanBuilder4Web; <https://gb.glyconavi.org/>

謝辞とライセンス

本研究は本研究は、JST NBDC JPMJND2204 の支援を受けたものです。

