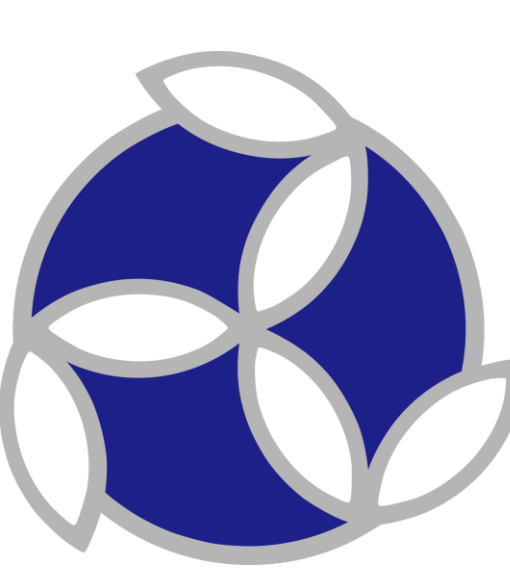




WURCSが持つ詳細な構造情報に基づく グラフ同型判定アルゴリズムの開発

○伊藤俊介¹, 松原正陽¹, 山田一作¹

¹公益財団法人 野口研究所

概要

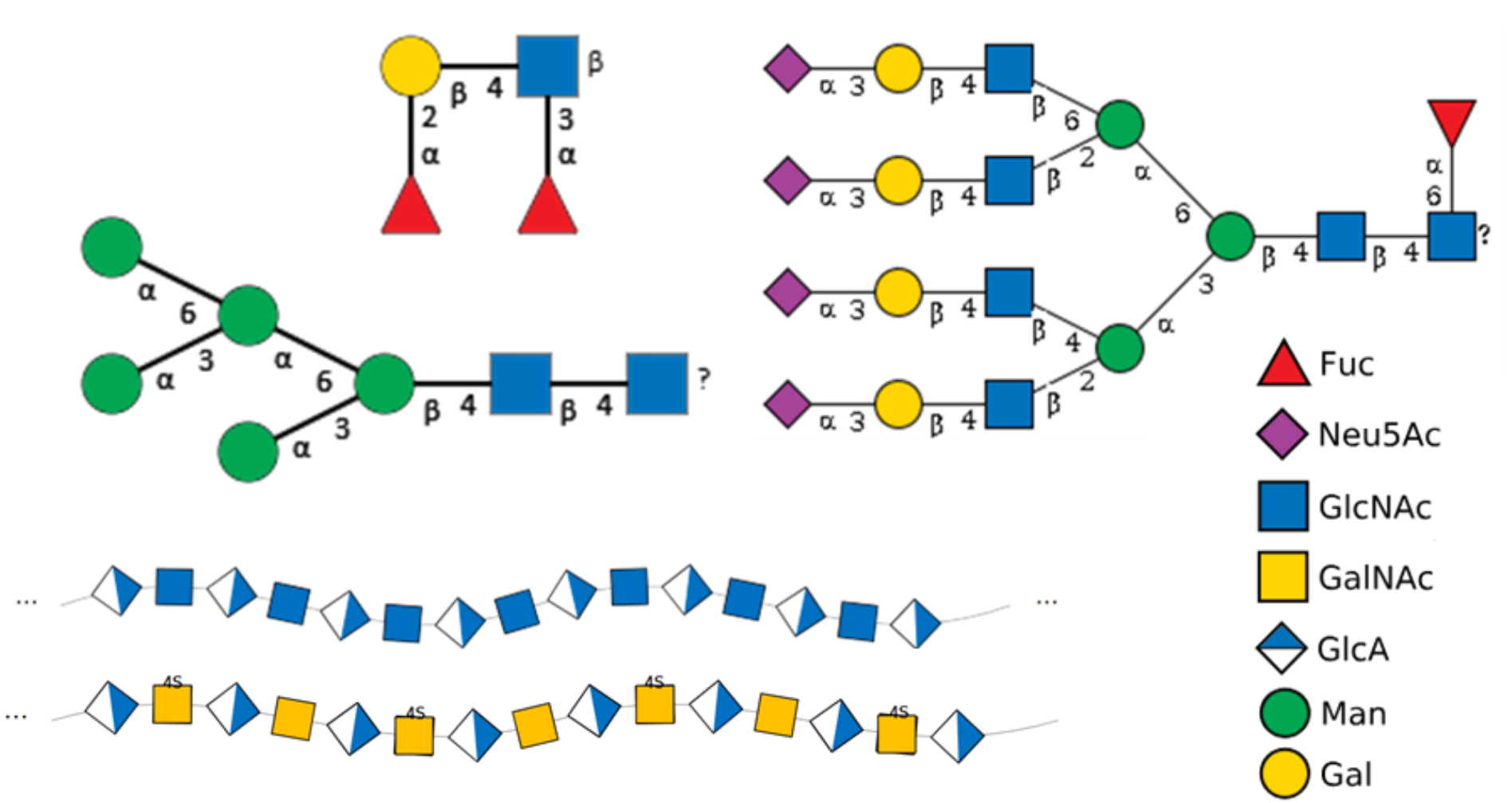
WURCS[1]は複雑な糖鎖構造を一意に記述できる線形表記として国際的に用いられているが、その表現力に完全対応した糖鎖構造検索は未整備であり、既存のRDF系糖鎖構造検索には速度や構造対応の制約がある。そこでJava実装のWURCSFrameworkを基盤に、結合位置や環情報、置換基などを段階的に検査しながら逐次探索型で部分グラフ同型を判定するアルゴリズムを開発した。本アルゴリズムは、単糖やグリコシド結合の同型判定を、包摂関係を含めて判定できることが重要な特徴である。現段階で繰り返し構造を除く特殊表記に対応し、25万件のGlyTouCan全件を1クエリあたり平均30秒で検索することが可能である。

[1] MATSUBARA, Masaaki, et al. WURCS 2.0 update to encapsulate ambiguous carbohydrate structures. Journal of chemical information and modeling, 2017, 57.4: 632-637.

背景

糖鎖とは？

- 糖鎖＝単糖がグリコシド結合で連なった直鎖または分岐した高分子
- 繰り返しや分岐、単糖の種類の多さから、**非常に多様な構造**を持つ



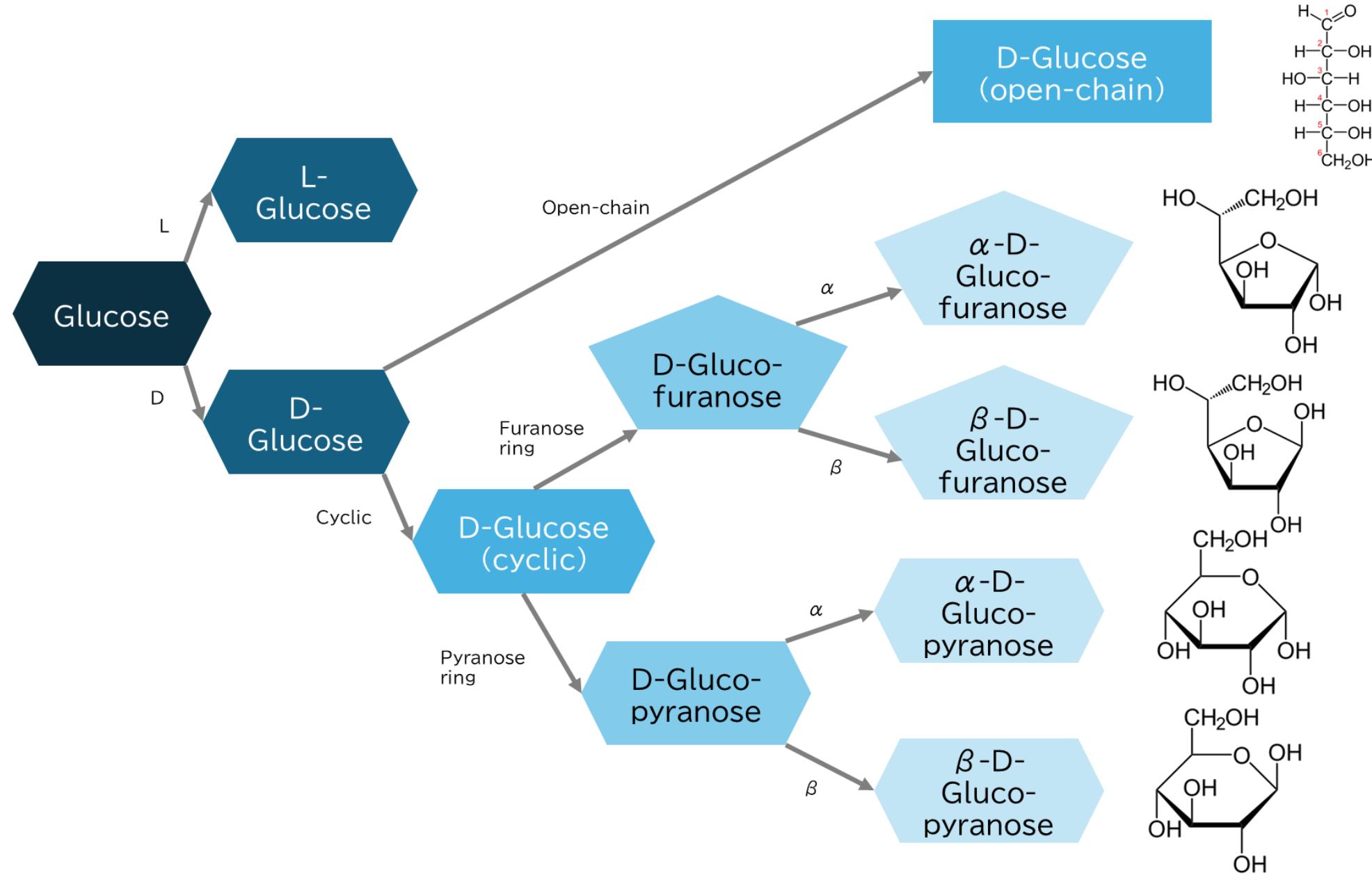
多様な糖鎖構造の例

課題

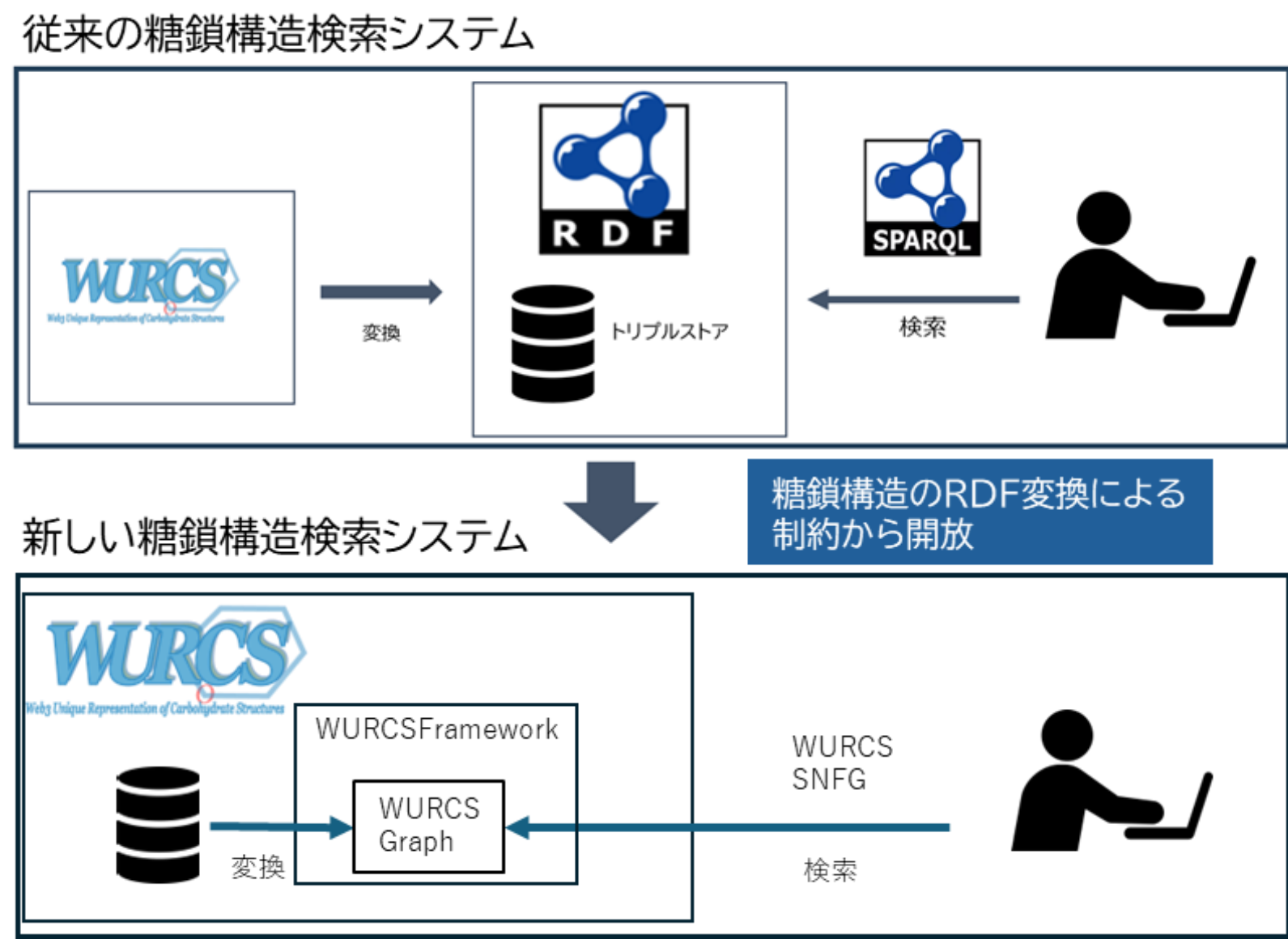
- 従来検索: 検索対象が RDF トリプル化された糖鎖構造であるため、糖鎖構造のRDF変換およびその制約に依存
 - 一部WURCS構造に非対応
 - 特定構造検索時の速度低下

提案

- WURCSを直接グラフとして扱う新規検索システムの開発
 - 検索速度の安定および高速化
 - あらゆる糖鎖構造を扱うことが可能



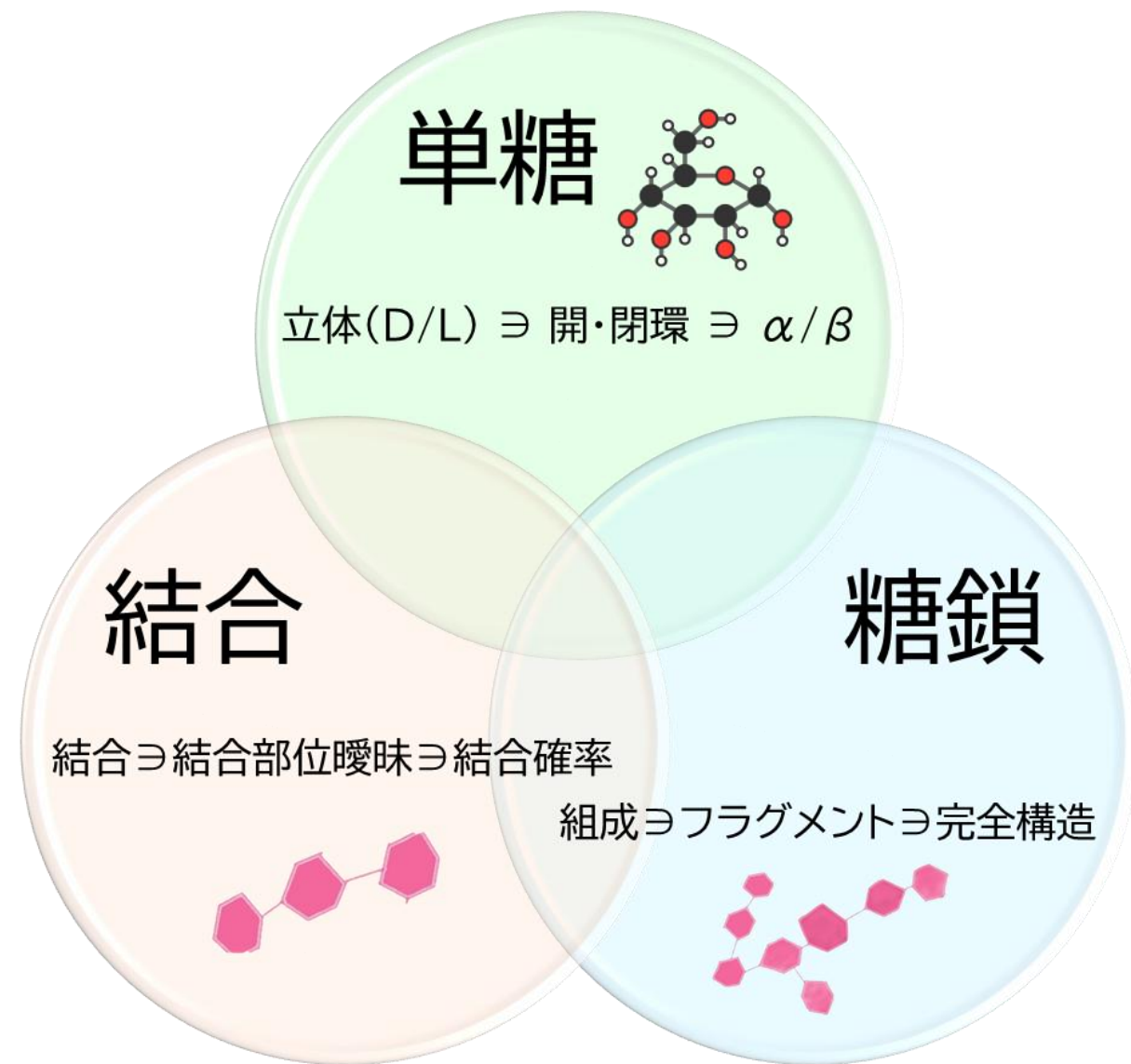
単糖の包摂関係



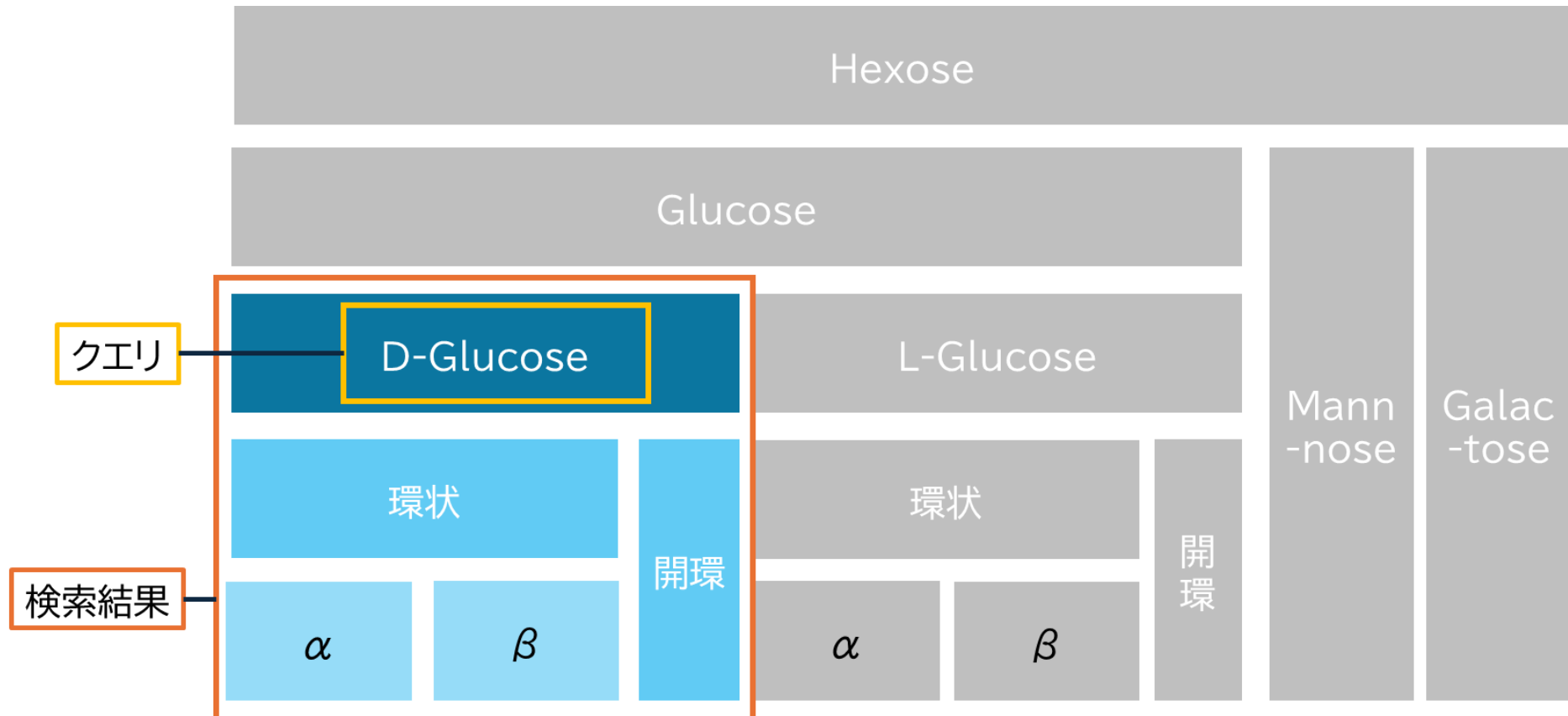
新旧検索システム概要

結果

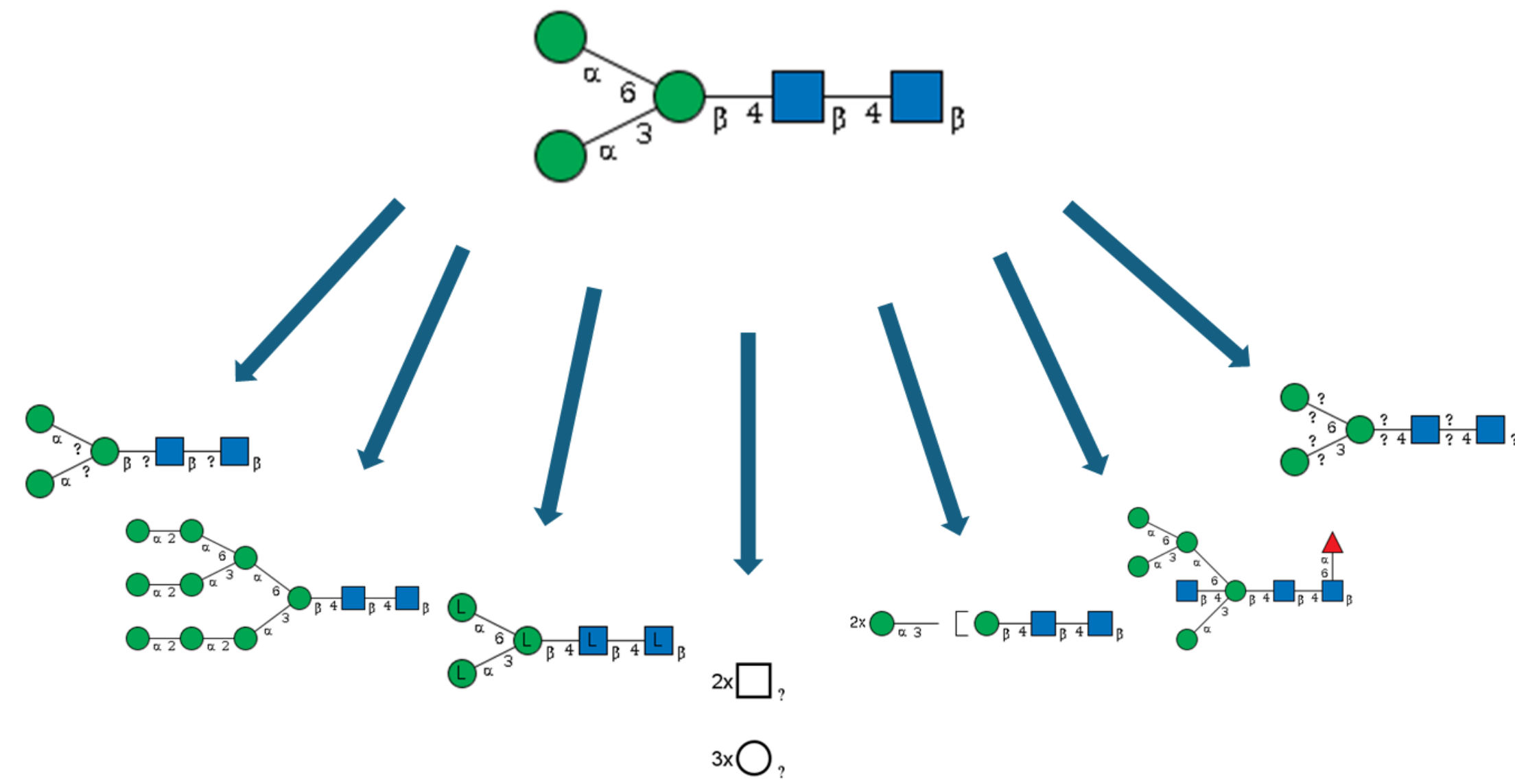
- 検索アルゴリズム: 糖鎖特化型かつ深さ優先探索を基盤とする例外対応付きアルゴリズム
- 部分構造一致によるクエリ包含検索
 - ほぼ全てのWURCS構造に対応
 - 包摂関係に応じた複数候補の同時取得



糖鎖における包摂関係
包摂関係は単糖、結合、糖鎖の各レベルで生じる



包摂関係に応じた複数候補の同時取得



あいまい検索によるクエリからの複数候補取得例

まとめ

- WURCS完全準拠であるため、ほぼ全ての糖鎖構造の検索が可能
- RDF 変換による情報損失や制約から解放され、元の構造表現に基づく柔軟な検索を実現

今後の予定

- リピート構造の対応: 未対応の糖鎖の特殊構造に対応
- 可視化対応: SNFG表記をベースとした一致部分のハイライト表示
- プログラムの最適化: 並列化、インデックス対応などで実用的速度を目指す

謝辞

・本研究は科学技術振興機構(JST)・NBDCにより実施されているライフサイエンスデータベース統合推進事業における「統合化推進プログラム」(JPMJND2204)の助成により実施しました。
・共同研究者および貴重なコメントを頂きましたすべての方に感謝します

ライセンス

Licensed under
a Creative Commons
表示4.0国際ライセンス

