



## CSDB. Биоинформатика делает шаг в сторону углеводов!

Современная химия, и тем более биология и медицина – это не только синтез и практическое изучение новых соединений, но и инструментальные аналитические методы и компьютерные технологии. Молекулярные расчеты, моделирование и сложная обработка данных начинают играть ведущую роль в в биологических и химических исследованиях.

В отличие от других наук о молекулярной основе жизни, гликомика – наука об углеводах – только делает первые шаги в сторону использования подходов, связанных с информационными технологиями.

База данных природных углеводов (Carbohydrate Structure Database, CSDB) – информационный проект мирового уровня, активно развивающийся с 2005 года в Лаборатории химии углеводов ИОХ. Его целью является привнесение в гликомику степени информатизации, сравнимой с существующей в геномике и протеомике. Полнота покрытия по бактериальным углеводам, беспрецедентное качество данных и изощренные алгоритмы их обработки делают CSDB востребованным ресурсом, не имеющим аналогов.

На основе этого проекта существует множество сопутствующих сервисов – от конформационных расчетов олигосахаридов и предсказания химической и пространственной структуры антигенов по спектрам ЯМР до автоматизации поиска структур – кандидатов в антибактериальные вакцины.

За последние 10 лет наш коллектив включал от 5 до 20 человек, работавших по свободному графику, в том числе удаленно. Финансовую поддержку обеспечивали несколько научных фондов России, США и Германии. CSDB – некоммерческий проект, свободно доступный пользователям через Интернет. Все участвовавшие студенты поступили в аспирантуру и к защите диплома имели несколько публикаций в авторитетных научных журналах.

Если разработка алгоритмов, системное проектирование и программирование биохимического софта не менее интересно Вам, чем работа в химической лаборатории, приглашаем Вас присоединиться к CSDB. Если Вы не имеете опыта работы в компьютерной химии (но имеете к ней интерес) – не страшно, научим ☺.

*Филипп Тоукач, идеолог и основной разработчик CSDB*

### Возможные области, в которых Вы смогли бы реализовать себя в рамках проекта:

- связь существующих языков описания углеводов с атомными моделями и дальнейшая интеграция с другими сервисами, использующими эти модели (молекулярные расчеты)
- предсказание спектров ЯМР, предсказание структуры по спектрам, оптимизация ресурсоемких алгоритмов и организация их удаленного многопоточного выполнения
- интеграция с программами анализа структур по масс-спектрам
- генерирование и обработка конформационных карт, предсказание стерических свойств и NOE
- автоматический сбор неструктурированных данных из других БД (PubMed, MESH) и Интернет, создание инструментов data-mining, аннотирования и верификации
- статистические исследования структурных свойств (остатки, связи, размер звена, терминаторы, ацетилирование и т.д.), выявление корреляций с таксономией, кластеризация организмов
- взаимодействие с другими проектами гликомики (NCBI taxonomy, PubMed, JCCGDB, GlyTOUcan, ICD-11), создание API
- редакторы молекул, визуализация структур и другая функциональность (Ваши идеи?)



### Дополнительная информация и контакты:

Сайт проекта: <http://csdb.glycoscience.ru/>

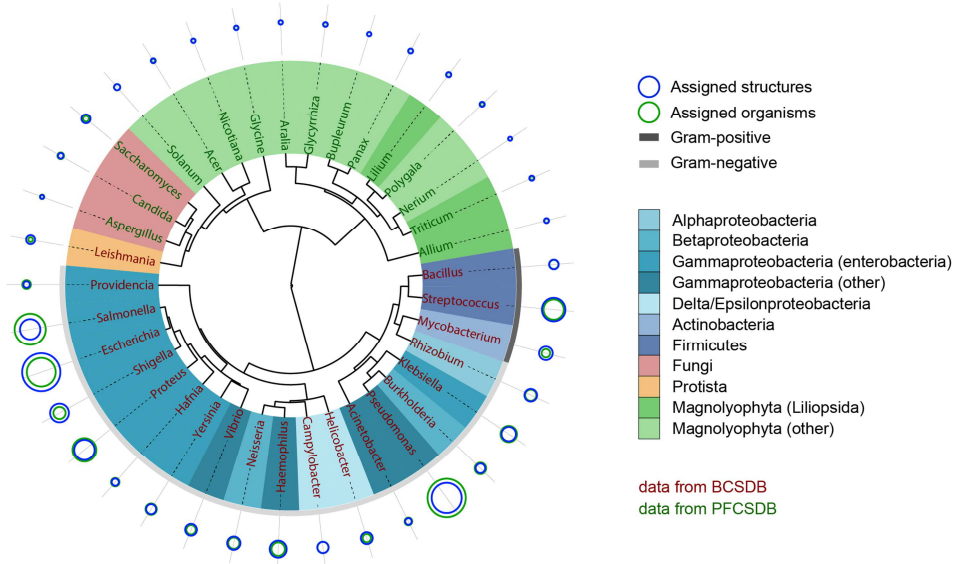
Лекция «Гликоинформатика»: <http://toukach.ru/rus/glycoinf.htm>

Или обращайтесь лично ко мне: Ф. Тоукач, д.х.н., с.н.с., доцент, выпускник лицей 1303 и ВХК РАН.

<http://toukach.ru/rus/> email: [phyl@toukach.ru](mailto:phyl@toukach.ru) тел.: +7 916 1724710 (12:00-22:00)

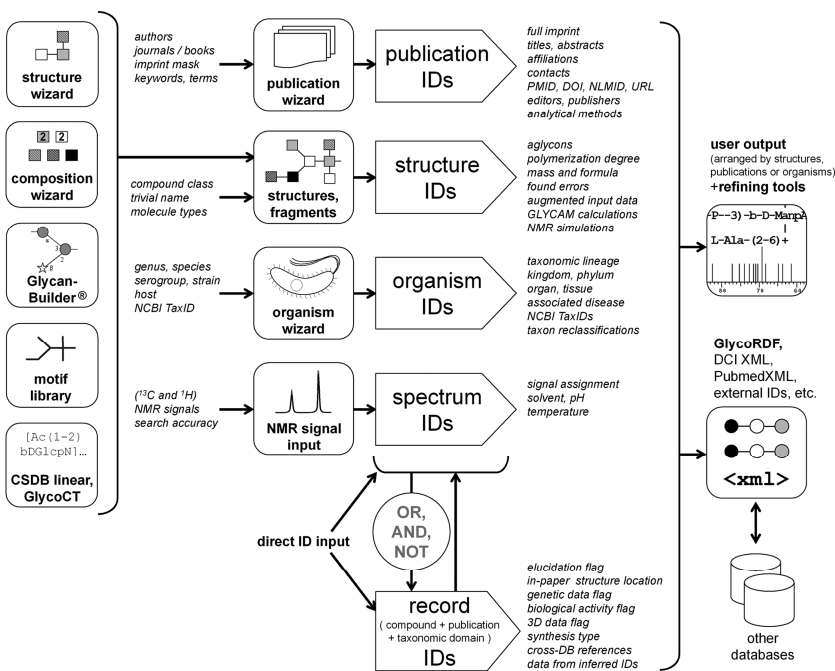
Ссылки на публикации по теме проекта: <http://csdb.glycoscience.ru/help/credits.html>

В качестве дополнения – некоторые картинки из наших статей:



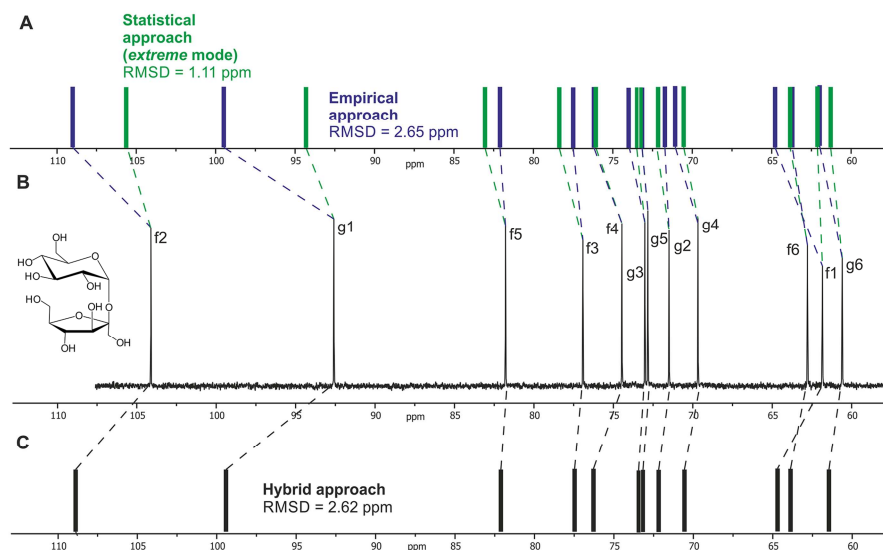
Фенетическое дерево наиболее изученных родов, построенное на основании анализа их гликомов

Toukach PhV, Kondakova AN, Egorova KS  
Database 2015  
импакт 4.5



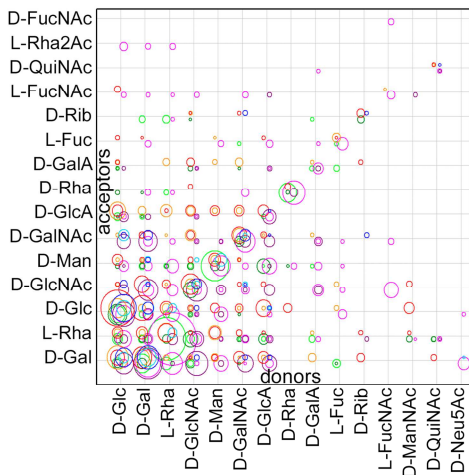
Взаимодействие данных о природных углеводах в проекте CSDB

Toukach PhV, Egorova KS  
Nucleic Acid Research Database Issue 2016  
импакт 9.2

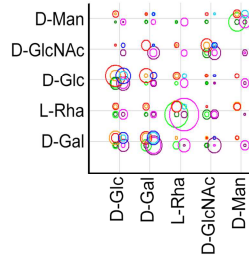


Предсказание спектров ЯМР<sup>13</sup>C сахарозы сервисом CSDB / GODESS

Toukach PhV, Ananikov VP  
Chemical Society Reviews 2013  
импакт 30.2

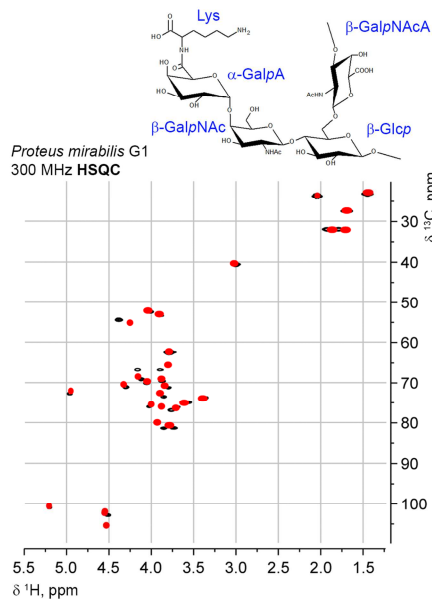
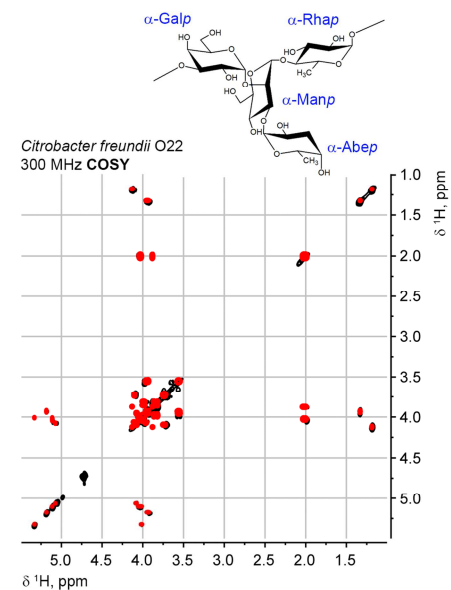


lower left corner demultiplied:



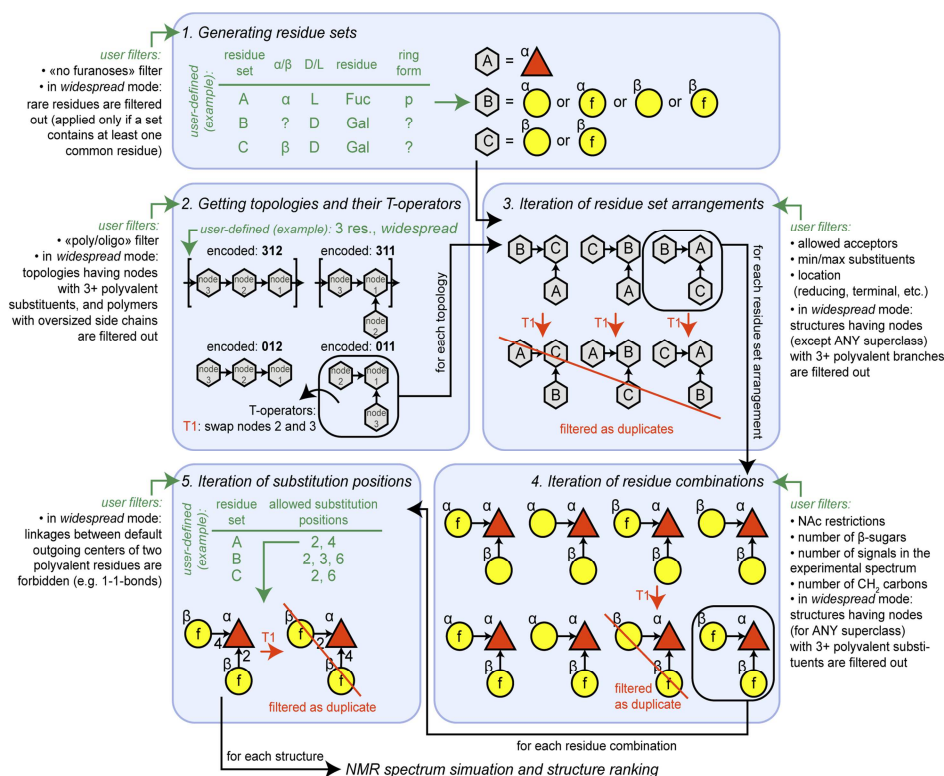
Распределение субстратов и реагентов гликозилтрансфераз в бактериях на основании анализа их гликомов

Herget S, Toukach PhV, Ranzinger R, et al. *BMC Structural Biology* 2008 импакт 2.0



Сравнение экспериментальных и предсказанных спектров двух бактериальных O-антигенов

Капаев RR, Toukach PhV *J Chemical Information and Modeling* 2016 импакт 4.1



Алгоритм генерирования многообразия углеводных структур по указанным структурным ограничениям

Капаев RR, Toukach PhV *Bioinformatics* 2017 импакт 7.3

Destination format: SMILES & 3D Convert

**Atomic structure**

There are 4 chemically distinct structures. Please select:

1. P-5([LIP(1-3)]xDRib-ol(1-2)[?Ala?](1-3))aDFuc3N(1-//LIP  
 2. P-5([LIP(1-3)]xDRib-ol(1-2)[?Ala?](1-3))aDFuc3N(1-//LIP  
 3. P-5([LIP(1-3)]xDRib-ol(1-2)[?Ala?](1-3))aDFuc3N(1-//LIP  
 4. P-5([LIP(1-3)]xDRib-ol(1-5)[?Ala?](1-3))aDFuc3N(1-//LIP

SVG file Show SMILES

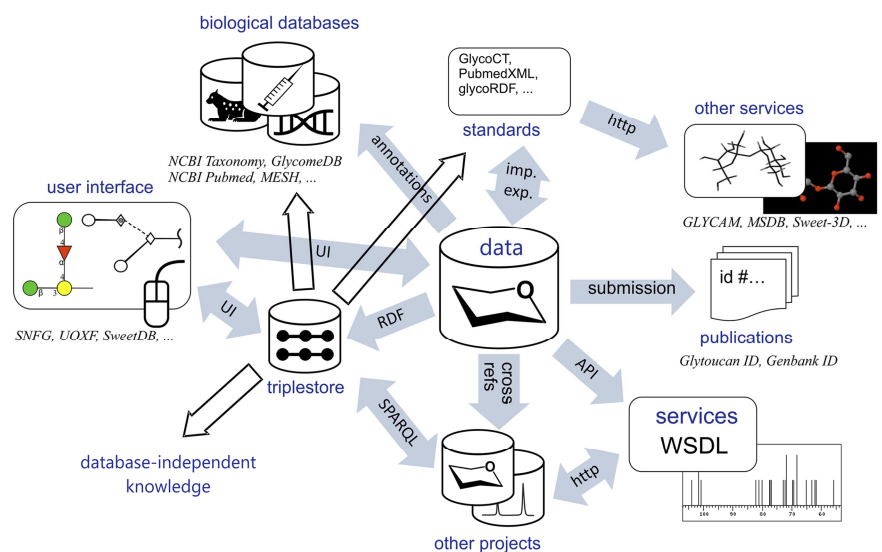
X1 = LIP (mix C12.0 + C14.0)

Please select one of 2 sterically distinct structures:

Get MOL Spin rotate zoom move

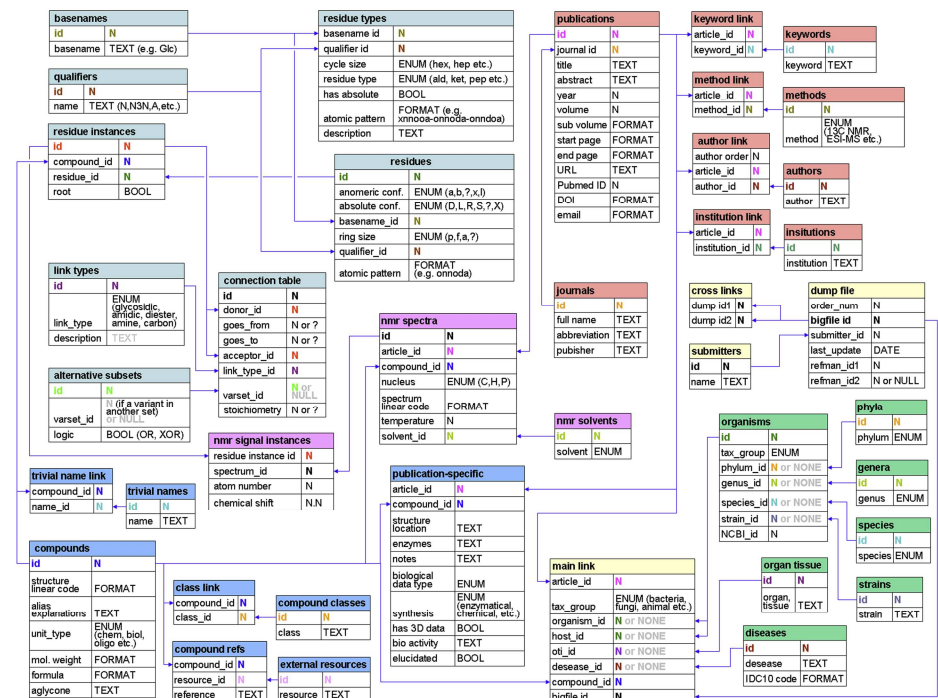
Трансляция символической углеводной нотации в SMILES и атомные координаты

Chernyshov IYu, Toukach PhV  
*Bioinformatics* 2018  
 импакт 7.3



Взаимодействие данных и сервисов гликоинформатики

Toukach PhV, Egorova KS  
*Angewandte Chemie* 2018  
 импакт 12.0



Архитектура базы данных природных углеводов

Toukach PhV  
*J Chemical Information and Modeling* 2011  
 импакт 4.3