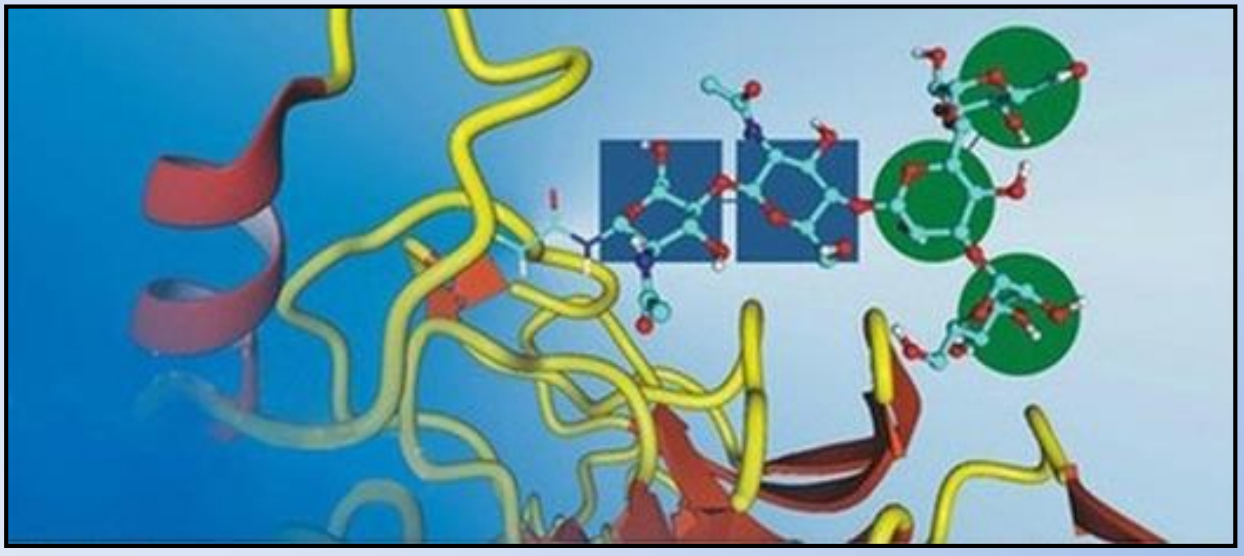
Институт
органической химии
им. Н.Д. Зелинского
РАН, Москва

Группа гликоинформатики

лаборатории химии углеводов и биоцидов ИОХ

д.х.н. Филипп Тоукач

вед.н.с. ИОХ РАН, профессор НИУ ВШЭ



http://toukach.ru/rus/nmr.htm

Гликоинформатика

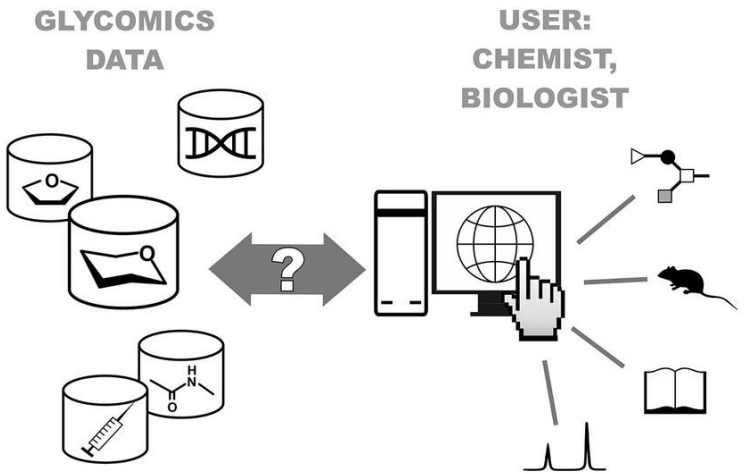
- **Роли углеводов:** гликозилирование белков, антигены микроорганизмов, межклеточные контакты, клеточная стенка, биоактивные гликозиды.

- Интерес растет, объём данных увеличивается, объекты сложнее и вариативнее, чем белки и ДНК =>
  => **тяжело ориентироваться в этом океане информации**

Задача:

обеспечить исследователей углеводов

всей мощью информационных технологий



Как это сделать?

биоорганическая химия – это не только синтез и экспериментальный анализ

- **Базы данных – легкий доступ к знаниям**

  Какие природные структуры похожи на заданные? Какие их фрагменты специфичны для заданных биологических видов? Где они опубликованы, в привязке к каким таксонам, болезням, и т.д.? Какие ферменты их синтезируют и с какой достоверностью это показано? На какие гликоэпитопы реагируют антитела?

- **Моделирование свойств молекул**

  Молекулярная геометрия, спектры, биоактивность, …

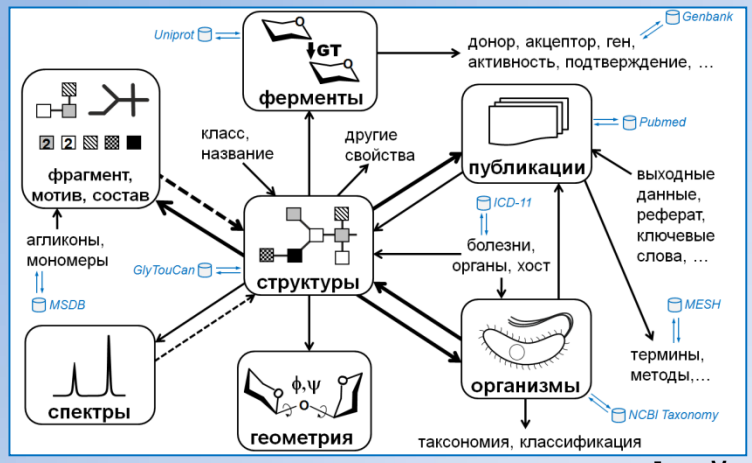
- **Предсказание структуры по наблюдаемым свойствам**
- **Предсказание свойств организмов**

  Кластеризация на основании гликомов, поиск схожести и различий таксонов, хемотаксономическая классификация

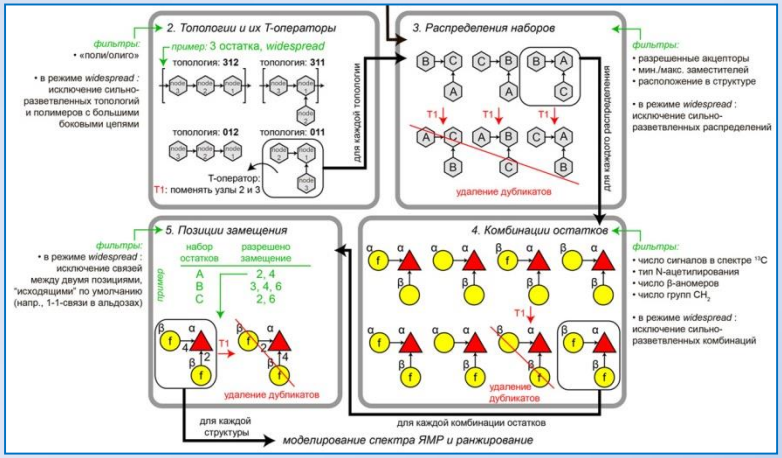
- **Идентификация и визуализация биомолекул**

Направления работы

Nucleic Acids Res, IF=17



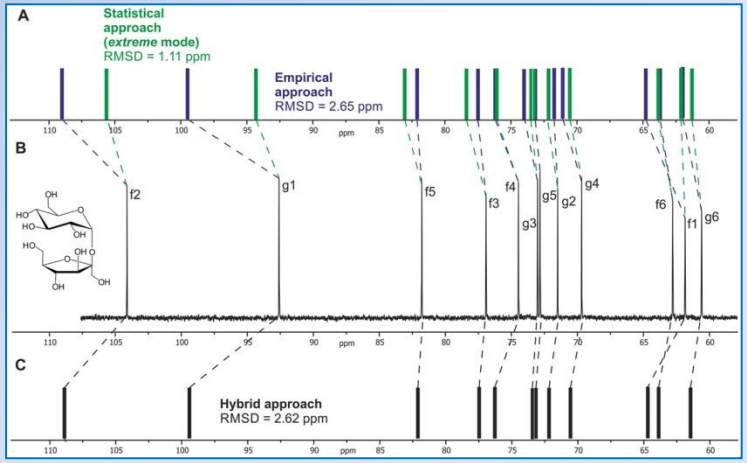
архитектура, сервисы, интерфейсы



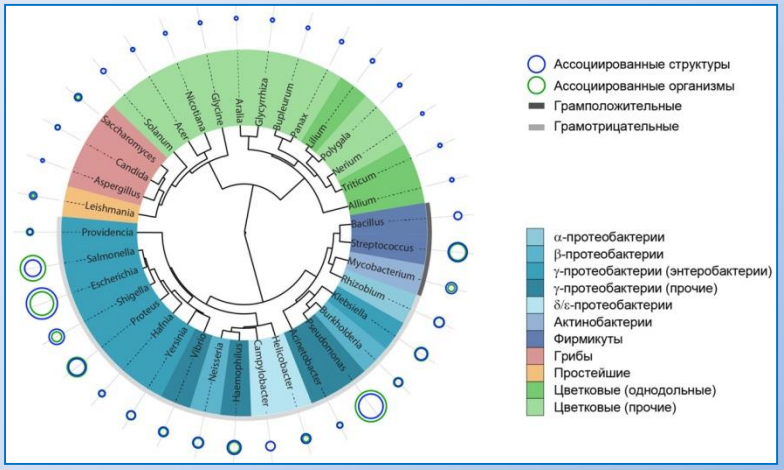
Bioinformatics, IF=7

хемоинформатика, алгоритмы

Chem Soc Rev, IF=55



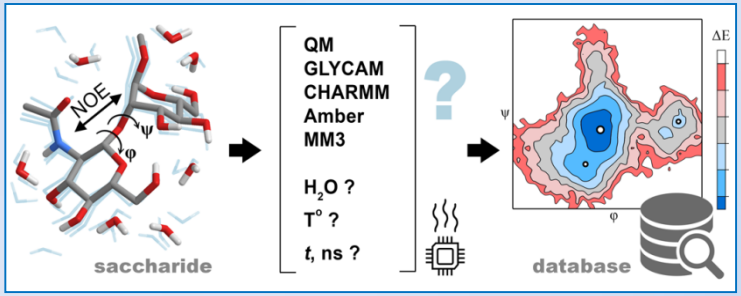
ЯМР-симуляция и предсказание структуры



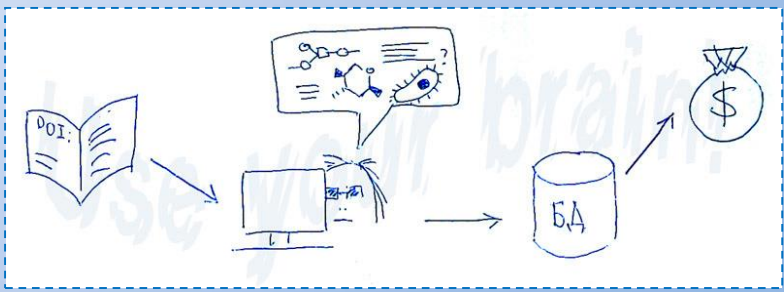
J Biol Databases Curat, IF=4

углеводная фенетика и таксономия

Int J Mol Sci, IF=6

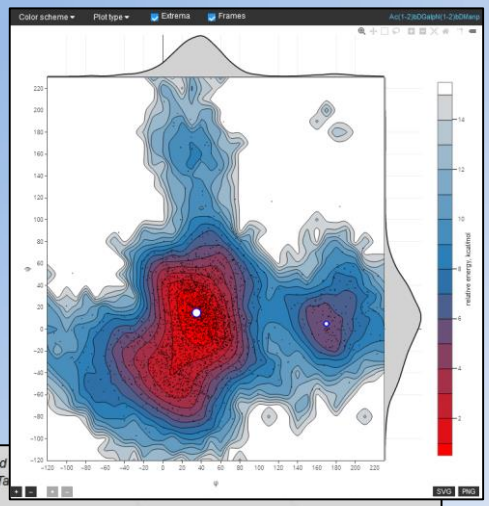


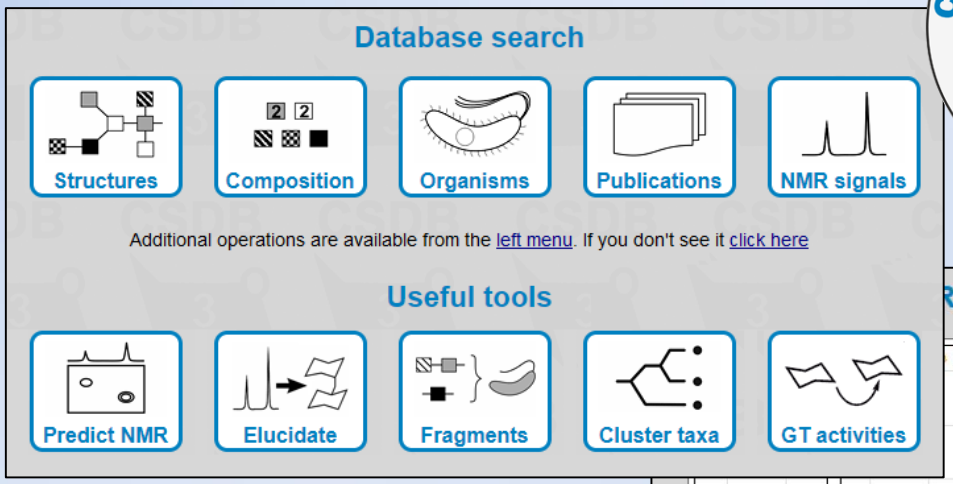
молекулярная динамика и моделирование

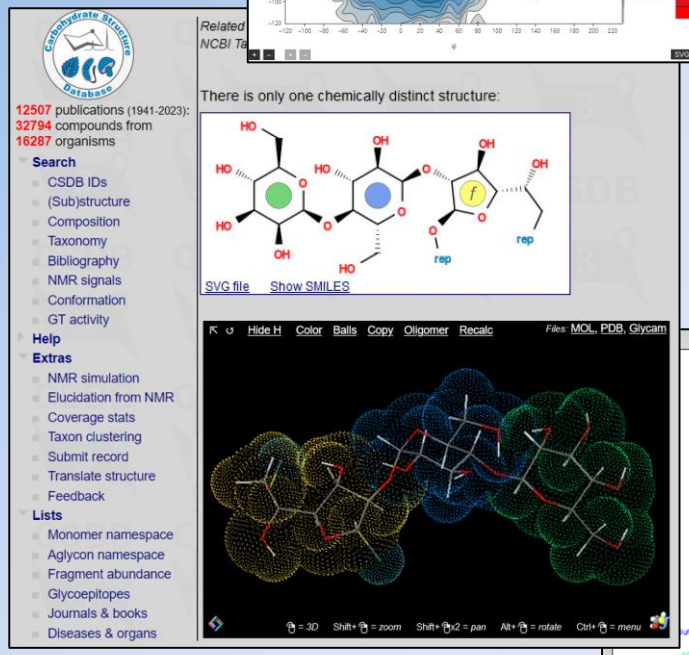


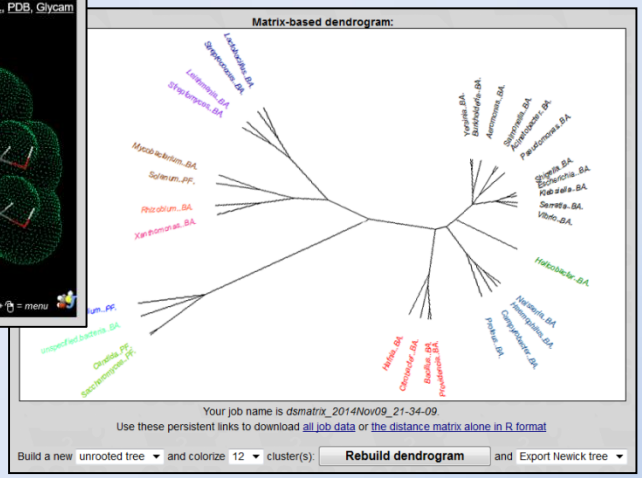
автоматизация аннотирования + ИИ

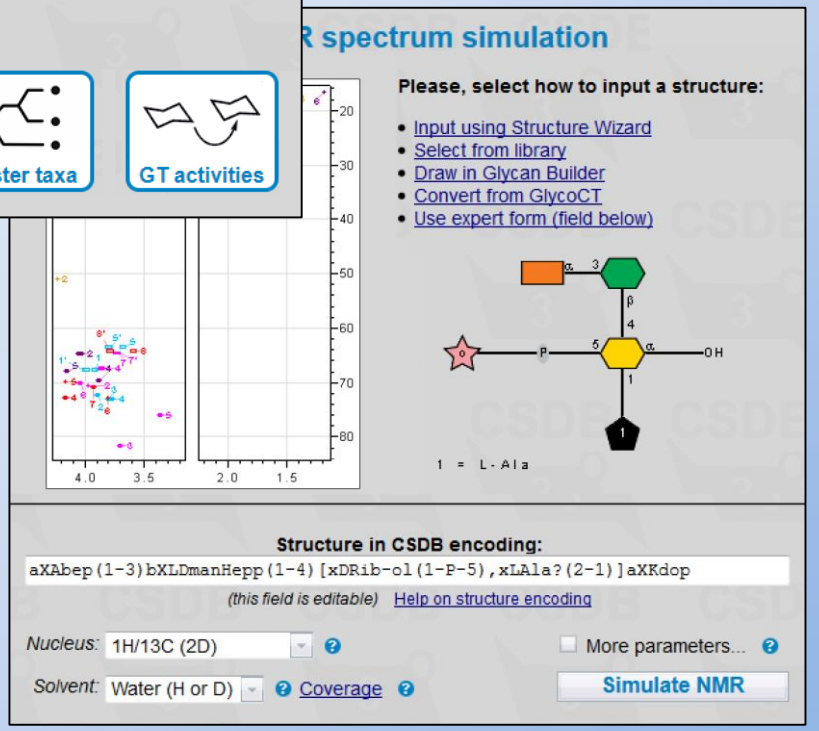
Carbohydrate Structure Database







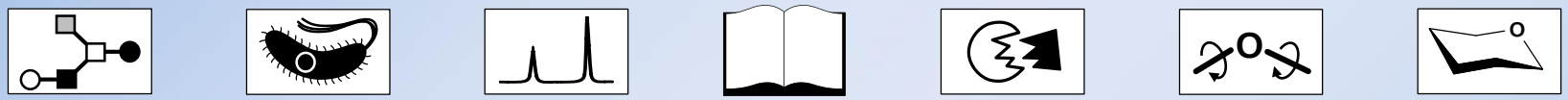




http://csdb.glycoscience.ru

Участие в проекте

База данных природных углеводов
Платформа для сервисов гликоинформатики



- **Интересная работа** с отчётностью по результату и свободным графиком
- **Публикации** в высокорейтинговых журналах (2-3 соавтора)
- **Получение опыта** работы в большом проекте со множеством взаимосвязей, требующим понимания контекста всей области знания
- Возможность как проявить инициативу, так и работать по техзаданию.
- Участие в конференциях, наработка научной репутации, диплом и диссертация
- Финансирование 2004-2022

Российский Фонд Фундаментальных Исследований

Международный Научно-Технический Центр

Комиссия по грантам при президенте РФ

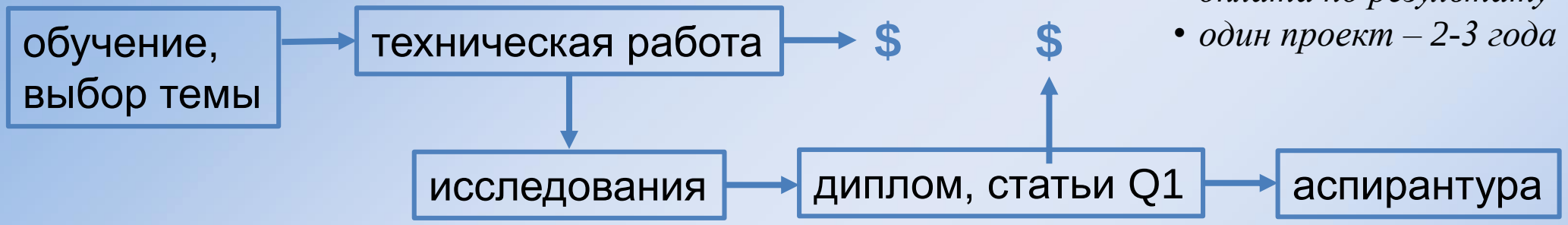
Немецкий Центр Исследования Рака

Фонд Содействия Отечественной Науке

Российский Научный Фонд

Чем можно заняться?

- *все на компьютере*
- *оплата по результату*
- *один проект – 2-3 года*



- молекулярные расчеты (автоматизация), генерирование конформационных карт
- нечеткое сравнение структур* и поиск структур с похожими мотивами
- симуляция масс-спектров, интеграция с программами их анализа
- предсказание стерических свойств и NOE (на основании конформаций фрагментов)
- автоматический сбор неструктурированных данных из других баз и Интернет (data-mining, аннотирование и верификация), алгоритмы выявления ошибок
- статистические исследования структур, выявление корреляций с таксономией, кластеризация организмов
- выявление углеводных компонентов в молекулярных графах* (напр., PDB)
- сравнение и предсказание спектров ЯМР (оптимизация)
- распознавание образов (иллюстраций в статьях), семантический анализ SNFG* и WURCS*
- другое (Ваши идеи?)

** программирование, предсказуемо, относительно быстро*

От вас потребуется

- **Интерес к информационным технологиям**

  программирование желательно, но не обязательно;
  умение и привычка пользоваться компьютером для решения самых разных задач - обязательно

- **Базовые знания органической химии и английского языка**
- **10-20 часов в неделю** (из них очно ~2) с перерывом на сессии
- **Самостоятельность в работе**
- **Ответственный подход к взятым обязательствам**
- **Готовность работать** много и интенсивно ради выдающихся результатов ☺

- **специальных знаний не требуется**

структурную химию углеводов, теоретическую информатику, интерпретацию экспериментальных данных, биохимический софт - изучите в процессе

Контакты

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

Группа : http://vk.com/glyco

Обзорная лекция «Гликоинформатика»: http://toukach.ru/rus/glycoinf.htm

Идеолог, разработчик, руководитель:
Филипп Тоукач
http://toukach.ru/rus/
email: phyl@toukach.ru
телефон & whatsapp: +7 916 1724710 (13:00-22:00)

Публикации по теме проекта: http://csdb.glycoscience.ru/help/credits.html

Институт органической химии им Н.Д. Зелинского РАН,
Ленинский пр-т 47, к. 432

эта презентация: http://toukach.ru/files/invi_2023.pdf

видео:

Хемоинформатика – конкретные задачи:

- автоматизация: получать SMILES из ChEBI ID + сравнивать его со сгенерированным для эпитопа
- конвертер CSDB Linear -> InChi using rdkit.Chem.inchi Python
- конвертер CSDB Linear -> bigsmiles Python
- распознавание остатков в SMILES Python|PHP
- распознавание остатков (и конфигураций) в PDB / MOL Python|PHP
- parse SNFG images to CSDB Linear PHP
- сравнение структурных дескрипторов (типа MACCS) в контексте пригодности к гликанам
- генератор структурных дескрипторов (из parsed structure) PHP
- генерирование формул Хеворта из SMILES или CSDB Linear Python|PHP
- парсинг формул Хеворта (Haworth) Python|PHP
- изучение применимости HELM к углеводам
- конвертер CSDB Linear -> HELM (combine smiles of residues) PHP
- заменить в конф. модуле MMFF на MM3 Python-RDKit

- подцеплять углы из БД в начальную геометрию перед генерацией MOL Python+PHP
- распознавание SNFG нейросеть (?)
- распознавание иллюстраций в статьях нейросеть