



VII Съезд биохимиков России  
ГЛИКОБИОЛОГИЯ



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

# Филипп Тоукач

## Трехмерные структуры в базе данных природных углеводов (CSDB)

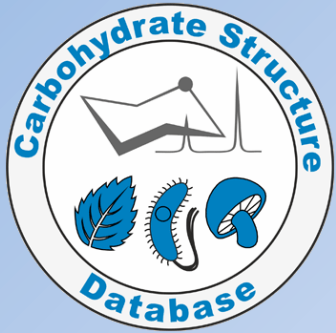


[http://toukach.ru/rus/CSDB\\_3D.htm](http://toukach.ru/rus/CSDB_3D.htm)

# Carbohydrate Structure Database

2

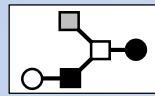
прокариоты, грибы, простейшие, растения



**CSDB**

База данных природных углеводов

Платформа для сервисов гликоинформатики



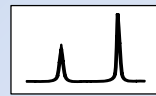
29K

первичные  
структуры



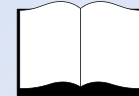
15K

таксономия



18K

спектры  
ЯМР



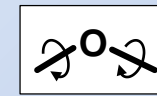
11K

библио-  
графия



2K

гликозил-  
трансферазы



3K

геометрия



3K

мономеры

- ежегодные обновления
- курируемые данные
- полное покрытие

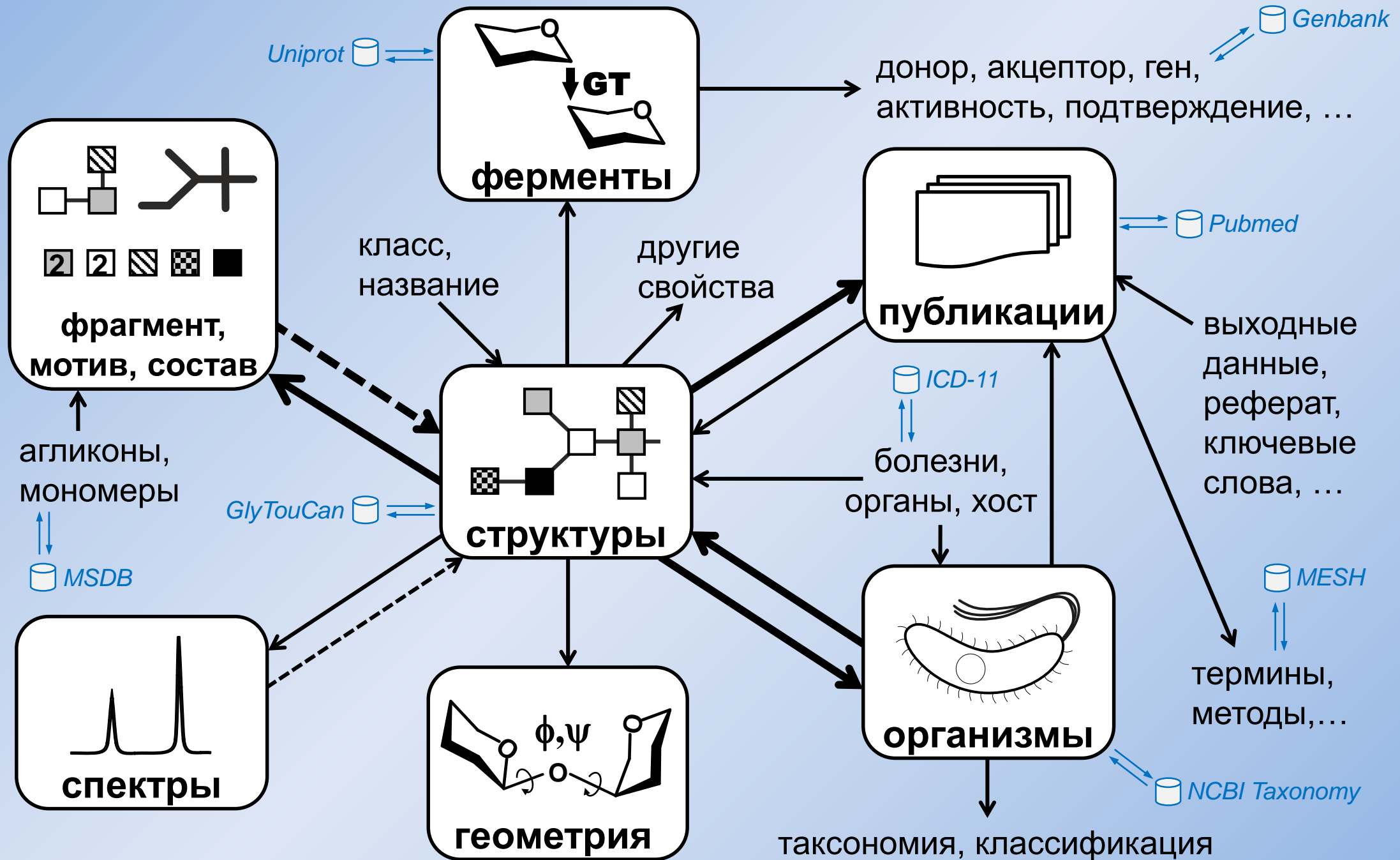
(микроорганизмы – до 2021, растения – до 2001)

- инструменты анализа данных
- анализ ЯМР / предсказание
- интеграция с другими базами

<http://csdb.glycoscience.ru>

- свободный доступ
- детальный «хелп»
- примеры решения задач

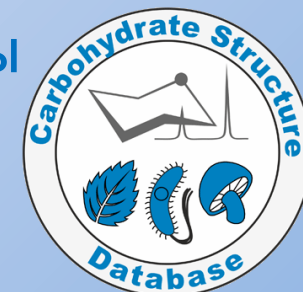
# Пути данных в CSDB



# Доступ к 3D-структурам

4

- Большинство баз = 3D-структуры гликанов **млекопитающих**  
(как часть гликопротеинов)
- <5% статей содержат 3D-данные
- Структуры симулированы в **разных** условиях
- **Гликополимеры** полностью отсутствуют
- Каждая симуляция превращается в отдельное исследование
- Нам нужен инструмент «из коробки», для не-информатиков
  - стандартизованные модели структур в растворе
  - автоматическая генерация, обширные пред-расчеты
  - экспорт в атомарные форматы



# Генератор начальных геометрий

CSDB/SNFG structure editor

Popular Small sugars Hexoses Higher sugars Alditols Aliphatic acids Other acids Superclasses

Novice Expert Insert Replace Oligo Poly Ac Am Cm Cho Fo Me Et Pr ETN Allyl Bz P S Pyr NH2

search residues search modifications

online редактор

Chemical repeating unit; n=10

`-3)aLFucp(1-6)[Subst(7-3)xDRib-ol(1-P-4)]?DGlc(1-?) [Ac(1-2)]bDGalfN(1- // Subst = chrysin = SMILES O=C2cc(c1cccc1)oc3c{7}c(0)c{5}c(0)c23`

- режимы «новичка» и «эксперта»
- все поли- и олигомерные топологии
- 600+ моносахаридов и других остатков
- SMILES для атипичных компонентов
- все типы связей (включая хелатные и C-C)
- поддержка неопределенностей, повторов, вариативности, суперклассов

Previews Refresh

Hi-res image

RES  
1r:r1  
REP  
REP1:5o(3+1)2d=-1--1  
RES  
2b:b-dgal-HEX-1:4  
3s:n-acetyl

«ЖИВОЙ»  
ЭКСПОРТ

Subst-(7-3)-D-Rib-ol-(1--P-4)--+  
|  
-3)-a-L-Fucp-(1-6)-D-Glcp-(1-?) -b-D-GalfNAc-(1-

Subst = chrysin = SMILES O=C2cc(c1cccc1)oc3c{7}c(0)c{5}c(0)c23

There are 3 chemically distinct structures. Please, select:

1. -3)aLFucp(1-6)[Subst(7-3)xDRib-ol(1-P-4)]?DGlc(1-3)[Ac(1-2)]bDGalfN(1- //
2. -3)aLFucp(1-6)[Subst(7-3)xDRib-ol(1-P-4)]?DGlc(1-5)[Ac(1-2)]bDGalfN(1- //
3. -3)aLFucp(1-6)[Subst(7-3)xDRib-ol(1-P-4)]?DGlc(1-6)[Ac(1-2)]bDGalfN(1- //

There are 2 sterically distinct structures. Please, select:

1. -3)aLFucp(1-6)[Subst(7-3)xDRib-ol(1-P-4)]aDGlc(1-6)[Ac(1-2)]bDGalfN(1- //
2. -3)aLFucp(1-6)[Subst(7-3)xDRib-ol(1-P-4)]bDGlc(1-6)[Ac(1-2)]bDGalfN(1- //

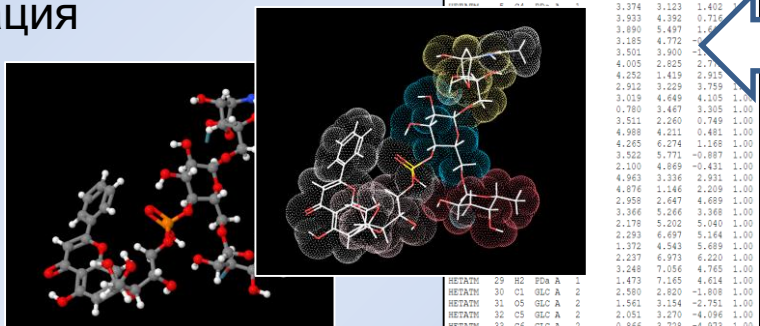
Show H Decolor Spheres Copy Files: MOL, PDB, Glycam

Files: MOL, PDB, Glycam

SMILES code:  
[ \* ] O [ C @ H ] 1 O [ C @ H ] ( ( [ C @ H ] ( O ) C O C 2 O [ C @ H ] ( C O [ C @ H ] 3 O [ C @ H ] ( C ) [ C @ H ] ( O ) [ C @ H ] ( [ \* ] ) [ C @ H ] 3 O ) [ C @ H ] ( O P ( = O ) ( O ) O C [ C @ H ] ( O ) [ C @ H ] ( O c 3 c c ( O ) c 4 c ( = O ) c c ( - c 5 c c c c 5 ) o c 4 c 3 ) [ C @ H ] ( O ) C O ) [ C @ H ] ( O ) [ C @ H ] 2 O ) [ C @ H ] ( O ) [ C @ H ] 1 N C ( C ) = O

3D Shift+ = zoom Shift+ x2 = pan Alt+ = rotate Ctrl+ = menu

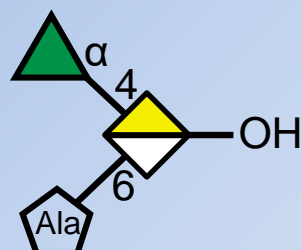
- online-моделер
- экспорт на углеводных и химических языках
- экспорт атомных координат
- визуализация
- цвета SNFG



# Анализ конформаций

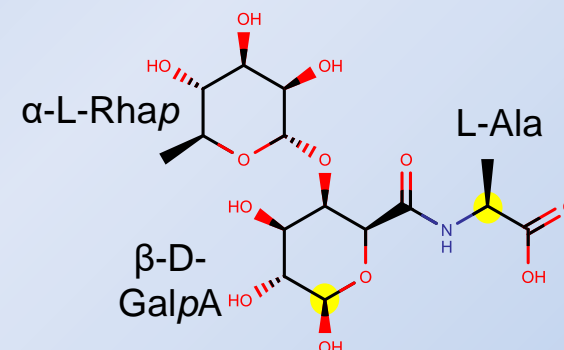
aLRhap(1-4)[x?Ala?(2-6)]?DGalpA

структура  
(возможно, неполная)

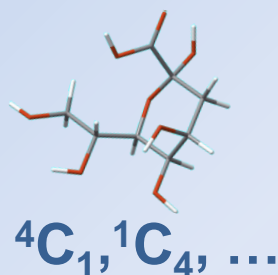
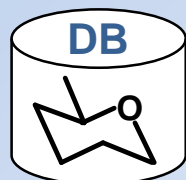


другие варианты  
( $\alpha$ -GalA, D-Ala, etc.)

SMILES

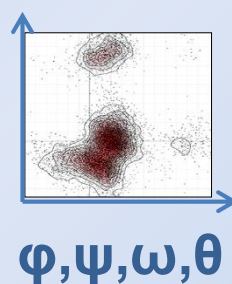
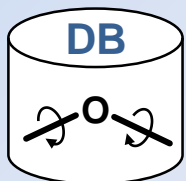


заселенные  
состояния  
циклов  
~1000 остатков

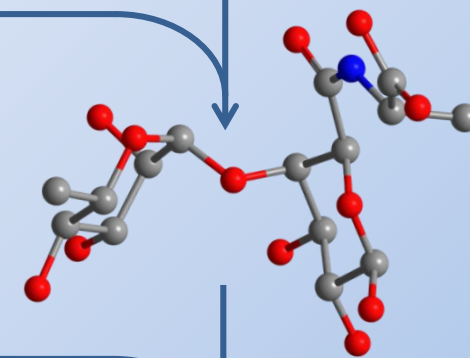


«креслификация»  
пираноз

заселенные  
торсионные  
углы связей

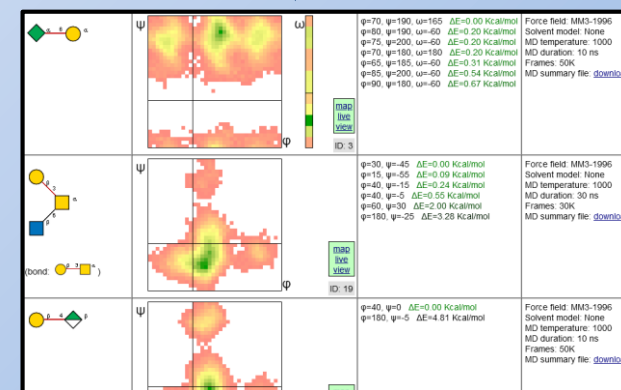


выбор минимумов  
ММ-оптимизация



мол. динамика  
300К, 100нс,  
явная  $H_2O$

конформеры  
+ энергии



# Поиск конформаций

## Search for conformation maps

Use the following criteria alone or in any combination to search for conformation maps.

**Conformation ID:**

**Model bond:** Use selectors

or type dimeric fragment in CSDB encoding

Strict modification search

**Model size:** Filter by target structure size

**Force field:** Filter by MD method

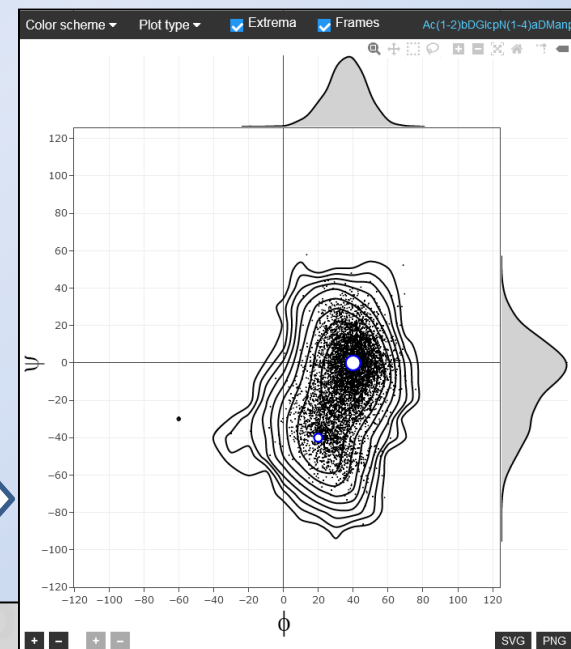
**Temperature:** Filter by MD temperature

**Solvent model:** Filter by solvent model

[Home](#) [Help](#)

explorer:

анализ и экспорт карт энергий и заселенности, поиск экстремумов



## CSDB conformation data search

73 conformation maps have been found.

Model structure	Conformation map	Energy minima	Details
		φ=40, ψ=0 ΔE=0.00 Kcal/mol φ=20, ψ=-40 ΔE=0.53 Kcal/mol	Force field: MM3-2000 Solvent model: Tip3P MD temperature: 300 MD duration: 100 ns Frames: 50K MD summary file: <a href="#">download</a>
		φ=40, ψ=175, ω=-60 ΔE=0.00 Kcal/mol φ=30, ψ=165, ω=-60 ΔE=0.00 Kcal/mol φ=30, ψ=165, ω=180 ΔE=0.10 Kcal/mol φ=40, ψ=185, ω=-60 ΔE=0.32 Kcal/mol φ=25, ψ=150, ω=180 ΔE=0.44 Kcal/mol φ=40, ψ=185, ω=165 ΔE=0.71 Kcal/mol φ=40, ψ=195, ω=180 ΔE=0.71 Kcal/mol φ=55, ψ=195, ω=-60 ΔE=0.86 Kcal/mol φ=40, ψ=185, ω=45 ΔE=0.86 Kcal/mol φ=40, ψ=150, ω=180 ΔE=0.86 Kcal/mol φ=30, ψ=180, ω=45 ΔE=0.86 Kcal/mol φ=15, ψ=170, ω=165 ΔE=0.86 Kcal/mol φ=50, ψ=175, ω=165 ΔE=0.86 Kcal/mol φ=25, ψ=155, ω=-60 ΔE=1.01 Kcal/mol φ=40, ψ=-210, ω=-60 ΔE=-1.01 Kcal/mol	Force field: MM3-1996 Solvent model: None MD temperature: 1000 MD duration: 30 ns Frames: 30K MD summary file: <a href="#">download</a>

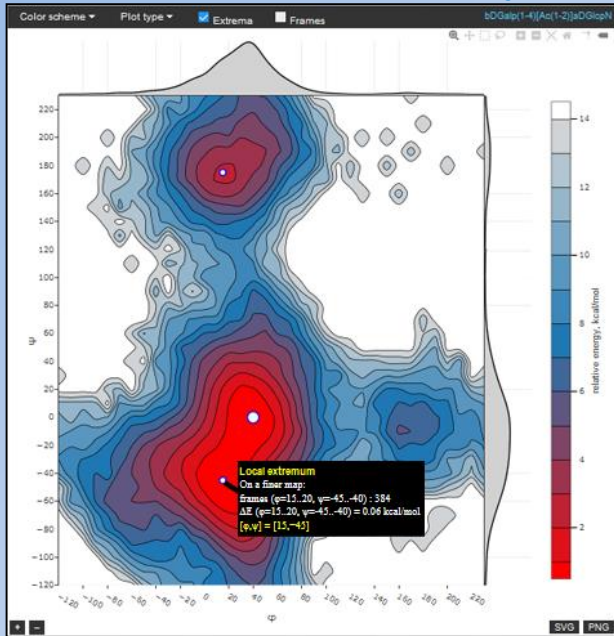


ПОИСК ПО:  
 ID,  
 структурам  
 и их фрагментам,  
 параметрам расчета

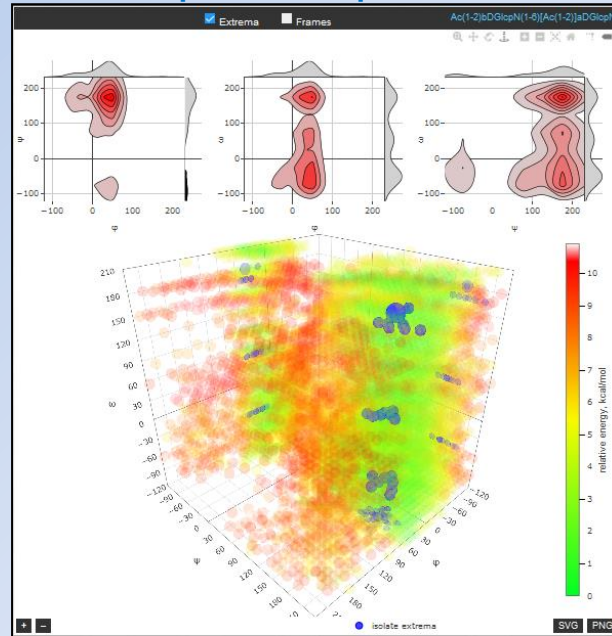
результаты  
 кратко

# Работа с конформационными картами<sup>8</sup>

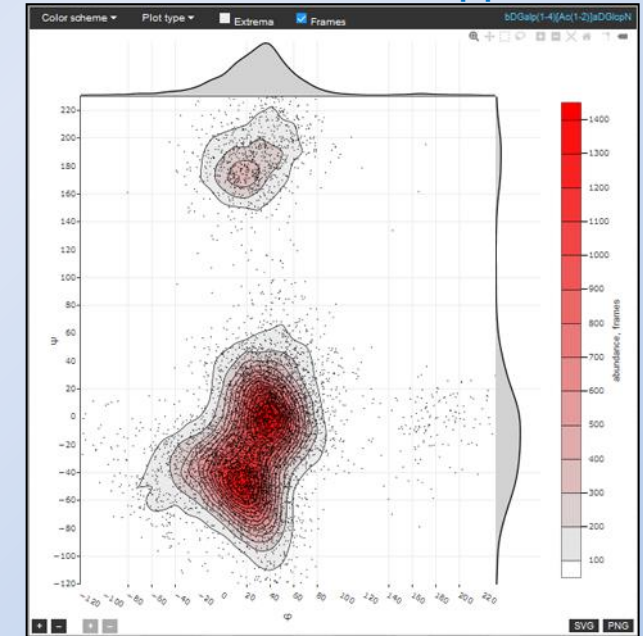
## 2D, энергия + экстремумы



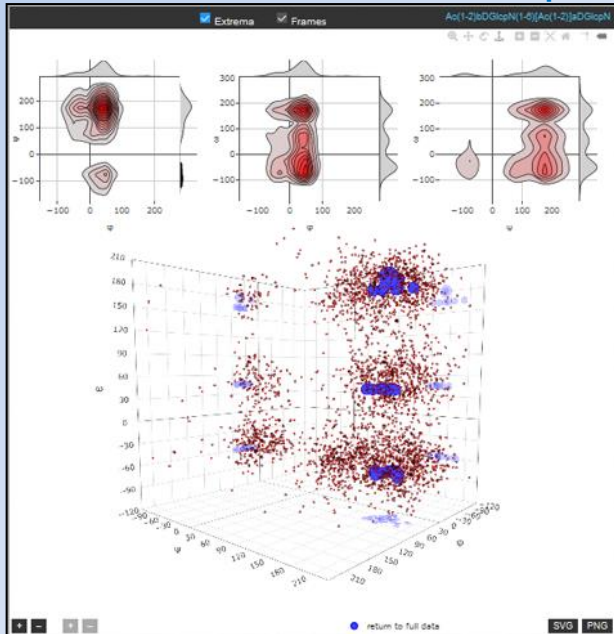
## 3D, энергия + проекции



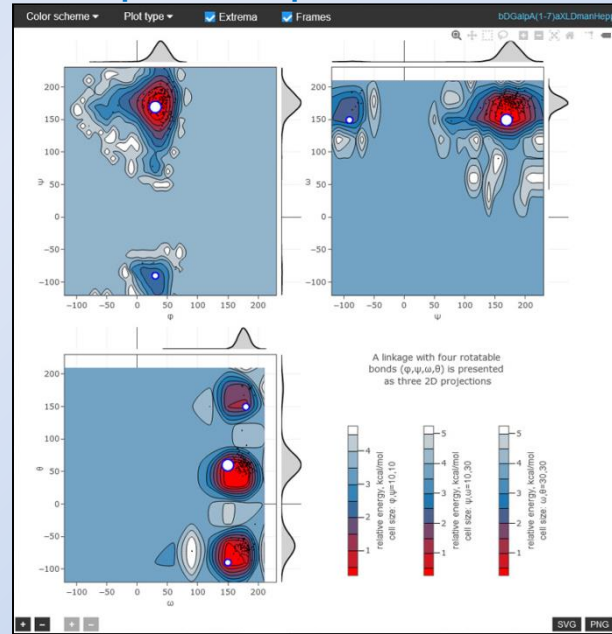
## 2D, заселенность + фреймы



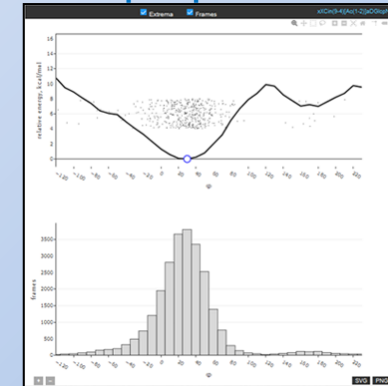
## 3D, заселенность + экстрем.



## 4D: три 2D-проекции



## 1D-профиль

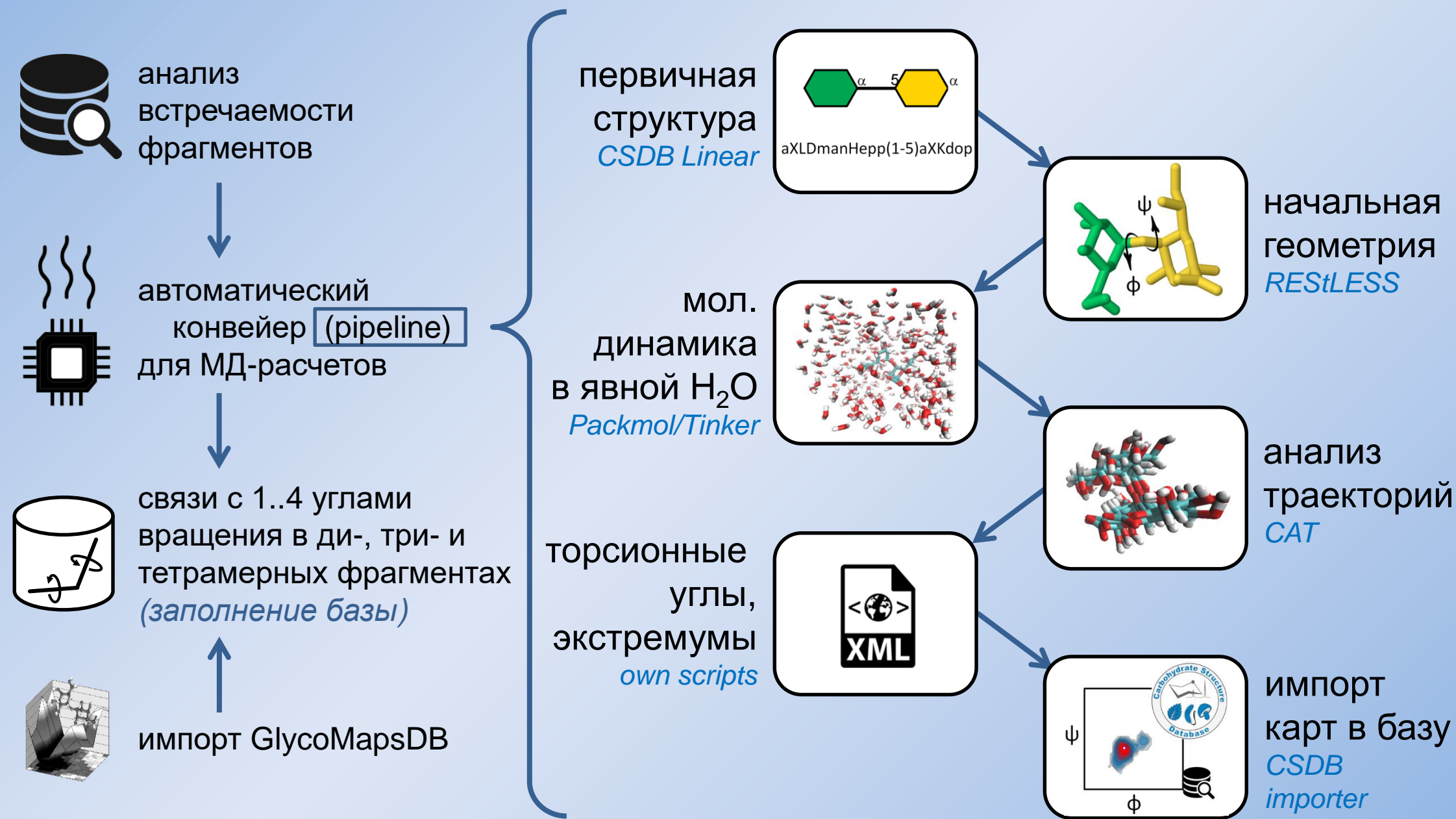


- до 4 углов / связь
- интерактивность (прокрутка, масштаб, вращение, контроль плотности, цвет, слои)



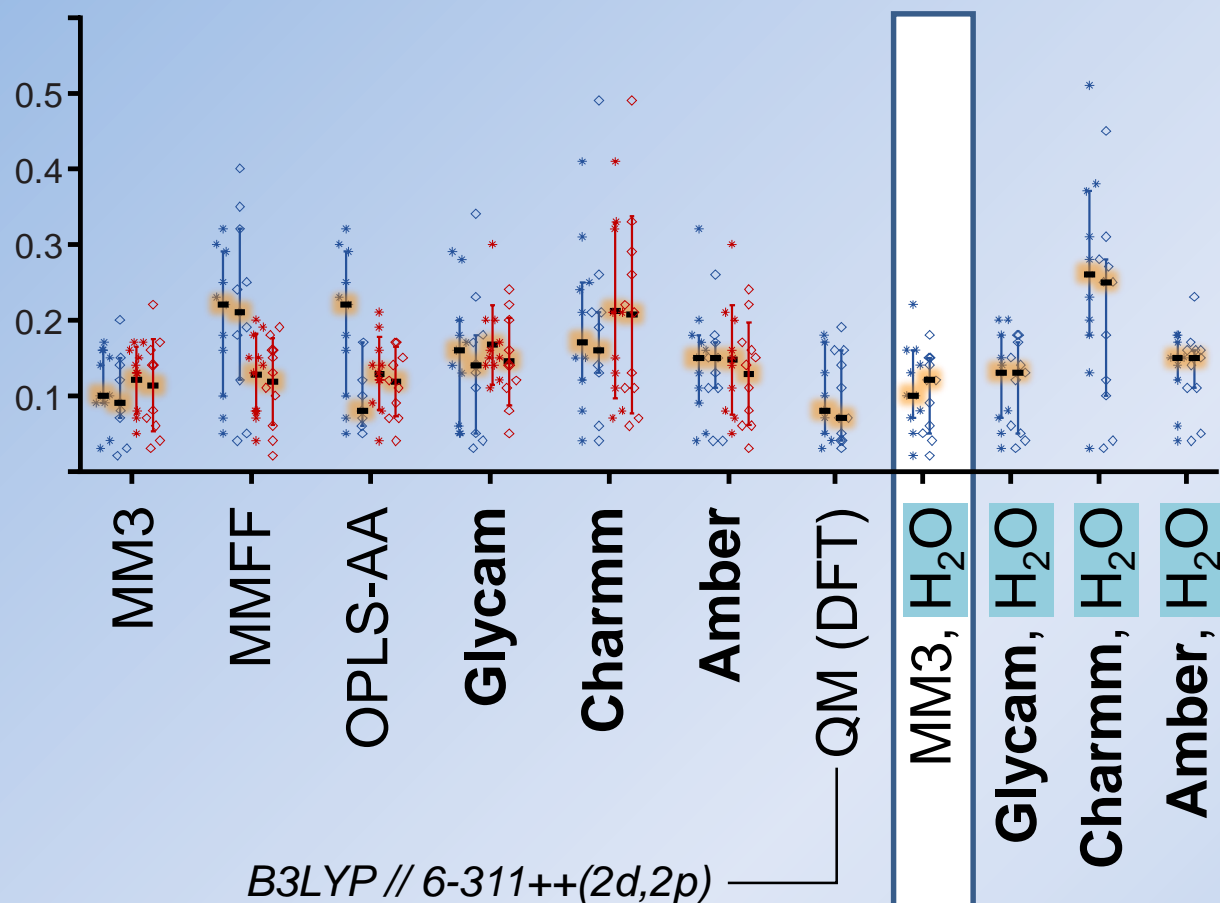
# Заполнение базы

9



# Метод расчета

NOE, RMSD



выборка:

11 дисахаридов разных классов

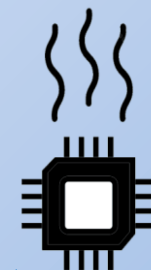
\* все H-H NOEs, 300K

◇ трансгликозидные NOEs, 300K

\* все H-H NOEs, 1000K

◇ трансгликозидные NOEs, 1000K

B3LYP // 6-311++(2d,2p)



непрерывно

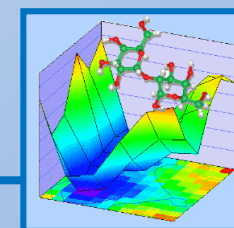
~300 моделей / год

~200 моделей: MM3, 100+ ns, 300K, явная H<sub>2</sub>O

~2400 моделей: MM3, 10-30 ns, 1000K, вакуум

GlycomapsDB

концепция



- Добавление NOESY в симулятор 2D-ЯМР  
(расчет из межатомных расстояний)
- Заполнение базы конформаций:
  - в порядке распространенности (200-300 расчетов / год)
  - олигомеры с неуглеводными остатками
  - замена устаревших данных GlycomapsDB
- Автоматизация вставки углов в модели  
(в генераторе начальных геометрий)
- Ансамбли конформаций: web-explorer / сравнение / экспорт

# Основные участники



программирование

аннотирование и курирование

сбор данных, поддержка

интеграция, онтология

конформационный анализ

R&D (разработка),

интерфейс, менеджмент

команда и партнеры CSDDB  
участники 3D-темы

 Роман Капаев, Андрей Бочков, Иван Чернышов, ...

 Ксения Егорова, Надежда Калинчук, Кирилл Казанцев, ...


 Юрий Книрель

 Рене Ранцингер, Кийоко Аоки-Киношита, Томас Люттеке, ...

 Виктор Стройлов, Софья Щербина, ...

 Филипп Тоукач

хостер

 Институт органической химии им Н.Д. Зелинского РАН

партнеры



деньги



[http://toukach.ru/rus/CSDDB\\_3D.htm](http://toukach.ru/rus/CSDDB_3D.htm)