



Ф.В. Тоукач

Компьютерные инструменты гликобиологии



Актуальные темы гликобиологии

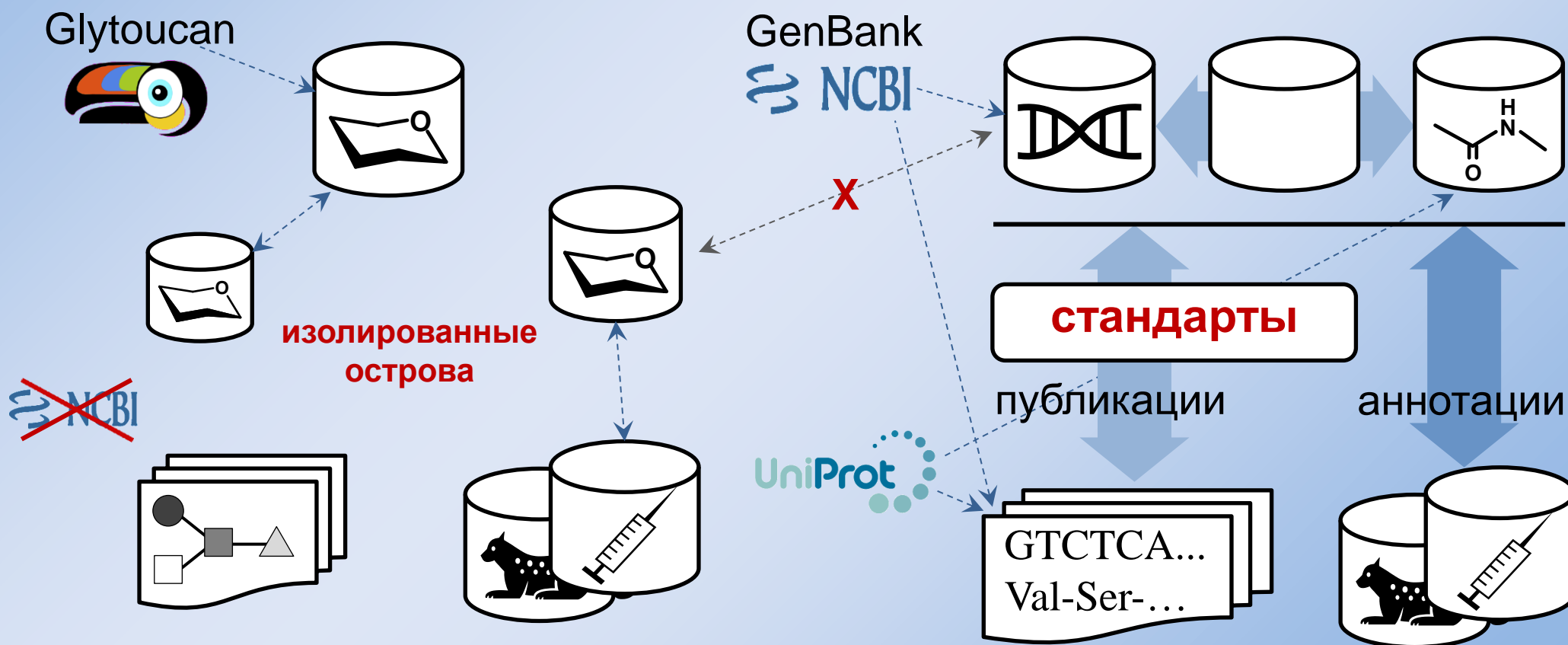
- Структура, изомерия, конформация углеводов в клетках
- Таксономия и классификация микроорганизмов
- Гликоэпитопы и иммуноспецифичность организмов
- Объяснение взаимодействия антиген-антитело
- Гликосодержащие вакцины и лекарства
- Корреляция свойств организма с его углеводами
- Биосинтез и круговорот углеводов

Гликомика vs. геномика, протеомика

3

по сравнению с другими -omics:

- сходный объем информации (>100 000 известных структур)
- бóльшая химическая вариативность
- меньшее использование IT (базы данных, сервисы)
- меньшая стандартизация

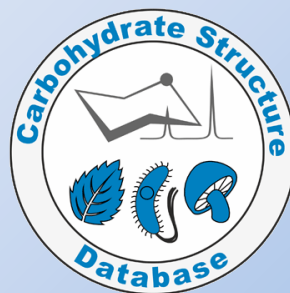


Проблемы в гликоинформатике

- Вариативность и гетерогенность объектов
- Неоднозначное описание структуры
- Сложности с вводом и визуализацией больших структур
- Отсутствие стандартов
- Изолированность проектов
- Неполнота и низкое качество данных в базах
- Ресурсоемкие алгоритмы
- Нехватка системного видения у разработчиков и пользователей
(нет общепризнанных сервисов, инициативы несовместимы друг с другом)

Углеводные базы данных

- Что это такое и какие они бывают?
- Что мы от них ожидаем и почему этого не получаем?
- Как вводить, кодировать и отображать структуры?
- Надстройки, сервисы и программы
- CSDB и другие проекты



Зачем нужна база данных?

- **Легкий доступ к знаниям и автоматизация исследований**

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, ...

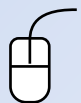
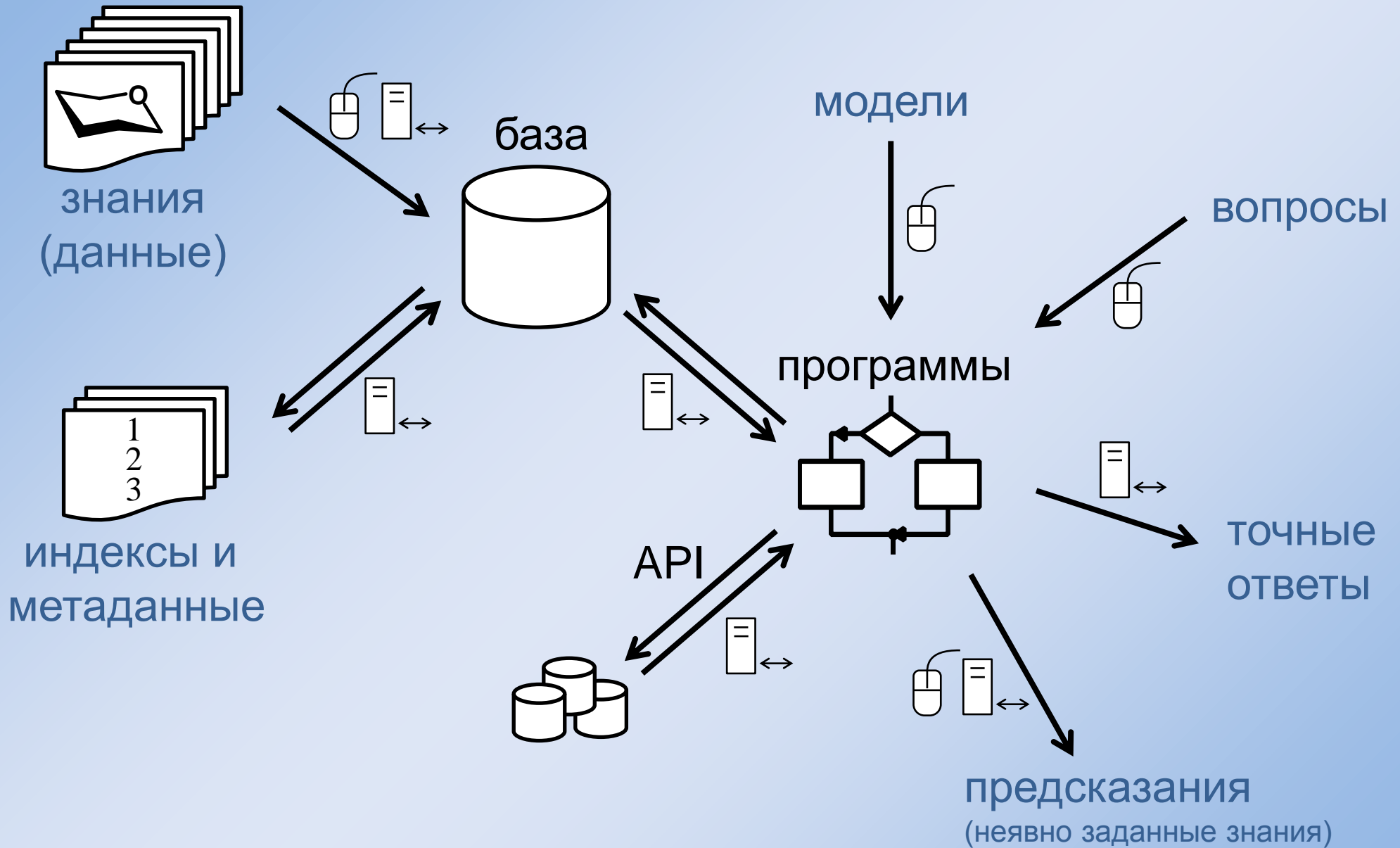
- **Предсказание структуры по наблюдаемым свойствам**

- **Предсказание свойств таксонов**

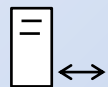
Кластеризация на основании гликомов, поиск схожести и различий таксонов, хемотаксономическая классификация

- **Идентификация и визуализация молекул** (в т. ч. в публикациях)

Что такое база данных?

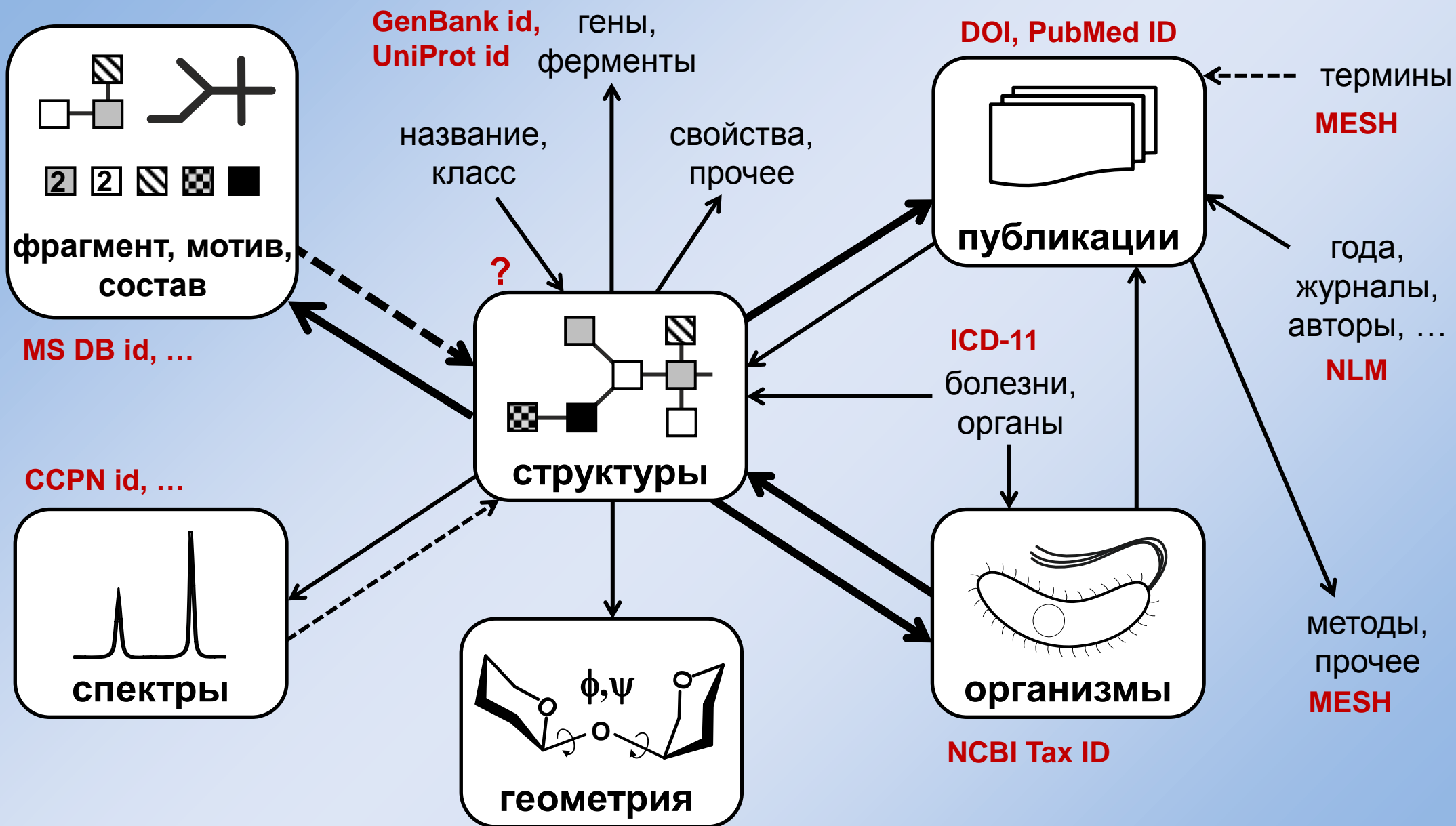


с участием человека



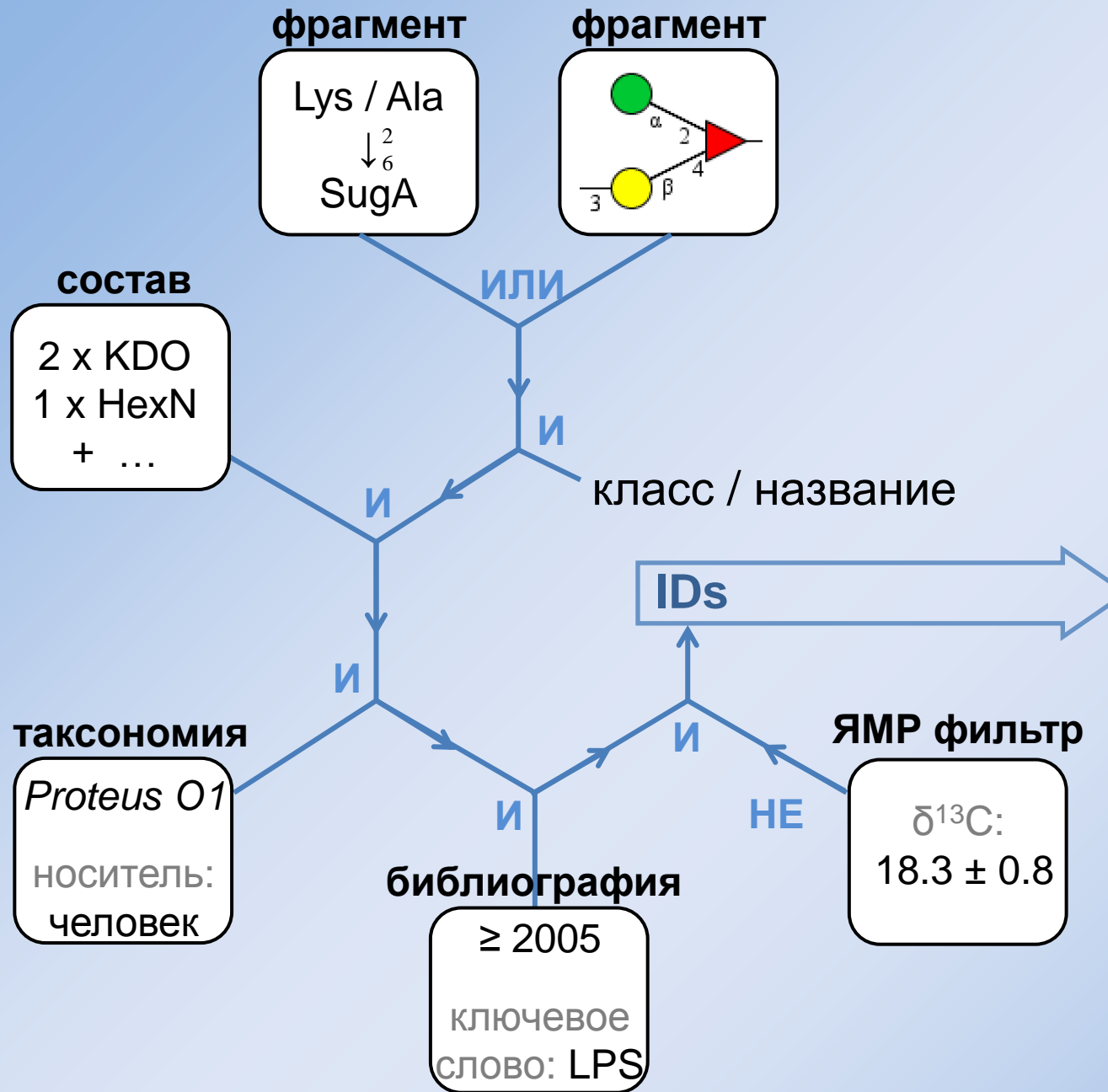
автоматически

Типичные запросы



→ однозначные переходы - - - - -> нечеткие переходы комбинации: **И, ИЛИ, НЕ**

CSDb: составной запрос



Данные сгруппированы по соединениям, публикациям, организмам и т.д.

Found 5 structures. Displayed structures from

Expand all compounds Show all as text (Sweet)

1. Compound ID: 10502

Structure type: polymer chemical repeating unit
Compound class: O-polysaccharide, O-antigen

Structural formula & atomic coordinates
Sweet-II 3D model

The structure is contained in the following publications:

- Article ID: 4266
Boyko AS, Dmitrenko AS, Fedonenko YF
O-polysaccharide of the lipopolysaccharide of *Azospirillum brasilense* strain Jm6B2 (D-Acofriose) - *Carbohydrate Research*

Two types of neutral O-polysaccharides were isolated from the aqueous phenol-water extraction from the ashy residue of the major O-polysaccharide was identified by 1D and 2D (1)H and (13)C NMR spectroscopy. D-rhamnose is indicated by italics.

Lipopolysaccharide, structural, O-antigen, physiology, Azospirillum brasilense, D-Acofriose

NCBI PubMed ID: 22575749
Publication DOI: 10.1016/j.carres.2012.04.006
Journal NLM ID: 0043535
Publisher: Elsevier
Correspondence: room308@ibppm.sgu.ru
Institutions: Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences
Methods: 1H NMR, 13C NMR, NMR-2D, methylation, chemical analysis, GLC, Smith degradation

Azospirillum brasilense Jm6B2
CSDb ID 2831

Article ID: 4833
Fedonenko YF, Boyko AS, Dmitrenko AS
The review of the polysaccharide structure of the genus *Azospirillum*

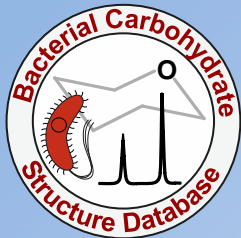
The review of the polysaccharide structure of the genus *Azospirillum*

Get MOL Spin rotate zoom move

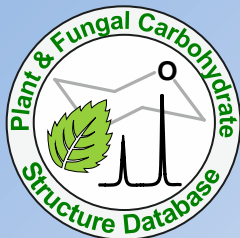
The spectrum also has 2 signals at unknown positions (not plotted).

Carbohydrate Structure Database

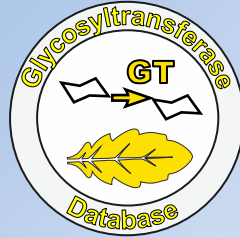
10



c 2005

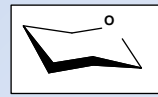


c 2012



c 2017

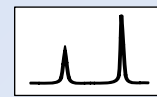
Комплекс баз данных +
Платформа для сервисов



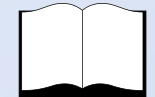
26K



14K



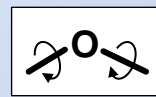
15K



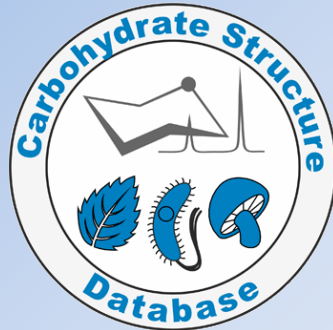
10K



2K



3K



CSDB

- регулярное пополнение
- расширяемая архитектура
- множество инструментов обработки данных
- проверка контента (15% = CarbBank, 85% = литература)
- полное покрытие (бактерии - сейчас, грибы - скоро)
- интеграция с другими базами

Структурные базы

CarbBank 23 полная до 1996 ORIG архитектура, % ошибок

GlycomeDB 120 мета-репозиторий неполные аннотации нет агликонов

GLYTOUCAN

CFG glycan млекопитающие, > 6

SweetDB, SugaBase **27 / 21** PDB

SCIENTES.DE

GLYCAN 11

Eurocarb DB архитектура только модель

BCSDB PFCSDB **15 / 6** (бактерии, археи) **11 / 5** (грибы, растения)

Carbohydrate Structure Database ORIG полная по прокариотам и грибам курируемая

Glycosyltrans Database

GlycoSuite ORIG млекопитающие+... **10 / 1** полная до 2005

JCGGDB **> 70** коллекция баз слабо аннотированы

UniCarbKB **4 / 1** курируемая

nibrT **0.7 O- & N-** ORIG

Glycoconjugate Data Bank **44** PDB

EcoDAB **0.2 E. coli** ORIG

GlyGen **34** (человек, мышь, крыса, вирусы)

GlycoBase **0.3 животные** ORIG

Специальные базы



углеводные гены человека



~0.2

MS^{2,3,4} N- и O-гликанов



~0.2

химические реакции



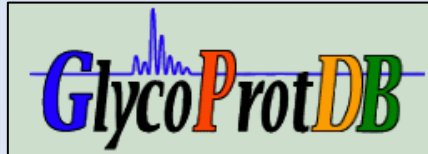
~3.0 (4.4 str)

конъюгаты & агликаны



>70

N-гликопротеины
C. elegans + мышь



~2.5

методики синтеза
и анализа



~0.2 (0.5 sub)

глико-эпитопы
и антитела



~0.2 (0.6 ABs)

GlyTOUcan,
репозиторий идентификаторов



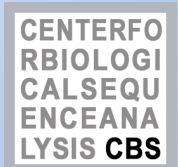
~120

адгезия к патогенам



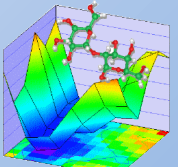
~0.9

O-glycBase,
O- и C-гликопротеины



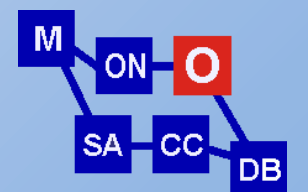
~0.2

GlycoMaps,
расчетные конформационные карты



~2.6

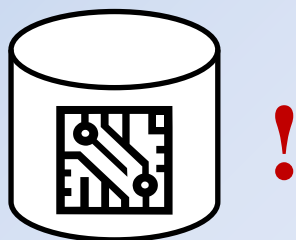
MSDB
моносахариды и номенклатура



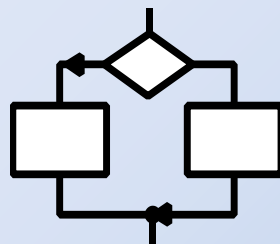
~0.8

Критерии оценки

- **Функциональность** (типы данных и индексов, обработка запросов)
- **Полнота покрытия** (+ выбранный класс)
- **Качество данных** (% ошибок, прозрачность)
- **Интеграция** (поддержка форматов, импорт-экспорт, API, RDF)
- **Интерфейс** (простота, стабильность, производительность)



внутренняя
архитектура

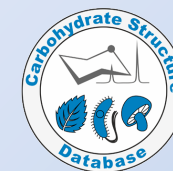


управляющие
программы



наполнение
данными

Архитектура



- Реляционная база данных
- Индексация + стандартные индексы (DOI, TaxID, ICD-11, PMID, ...)
- Структуры, таксономия, библиография – разные типы записей
- Человекочитаемый дамп (организация процесса наполнения)
- Контролируемые словари терминов (мономеров; MSDB)
- ~~Free text~~ ↗
- Таблица связности

минимум

структура,
таксономия,
библиография,
ссылки на другие БД

дополнительно

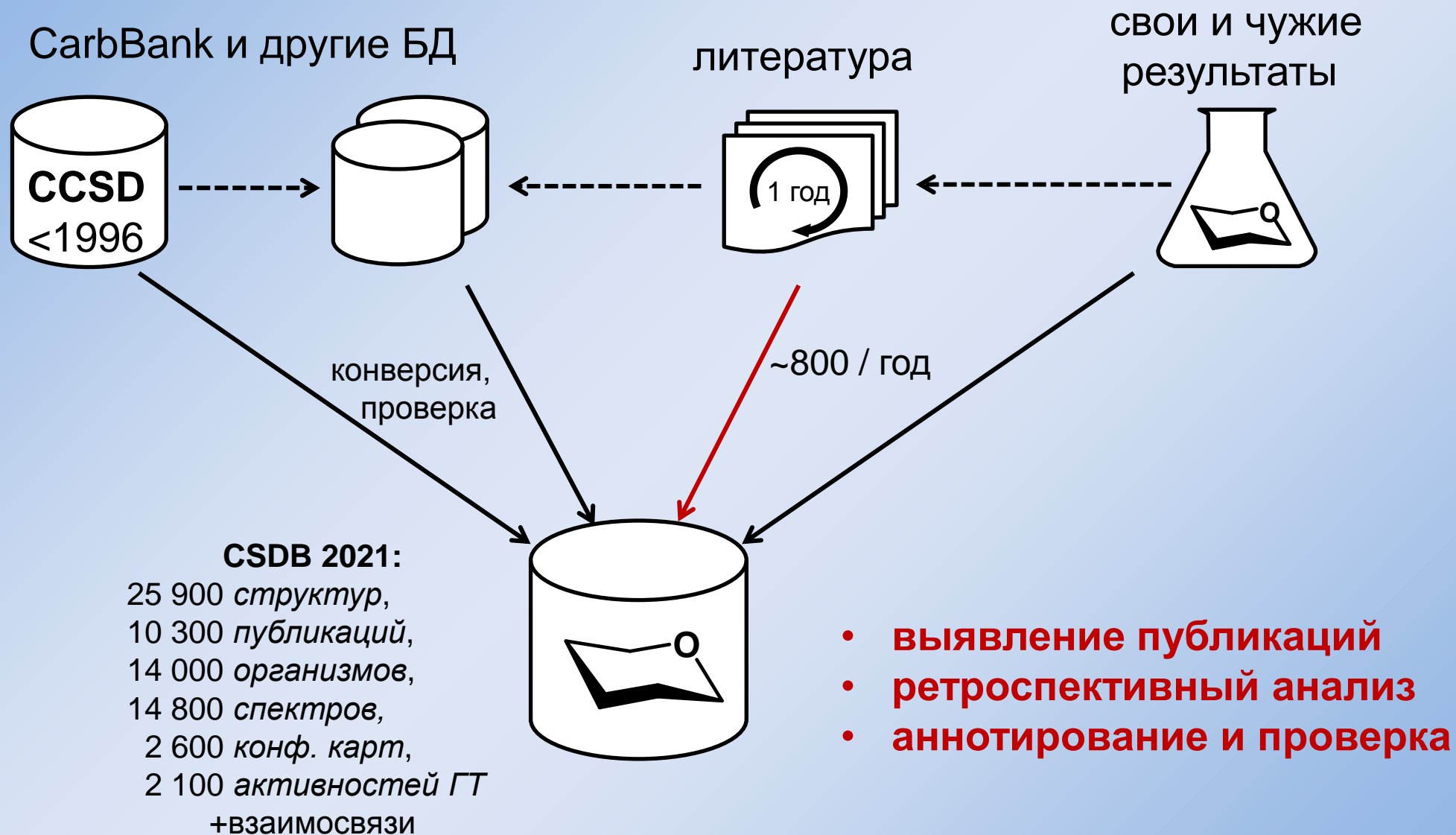
тривиальные названия,
спектры ЯМР, МС
условия съемки спектров,
конформация

гликопротеины,
биоактивность,
гены, ферменты биосинтеза

болезни
органы, ткани
генотип, стадия

ключевые слова
рефераты
институты
методы

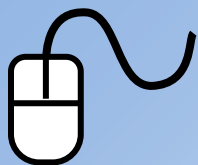
Источники данных



полное покрытие по прокариотам и грибам:

отрицательный результат поиска = значимая научная информация

Качество данных



операторские

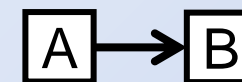


в базах

найдено &
исправлено



в статьях



в программах

ошибки, противоречия

исправляемые

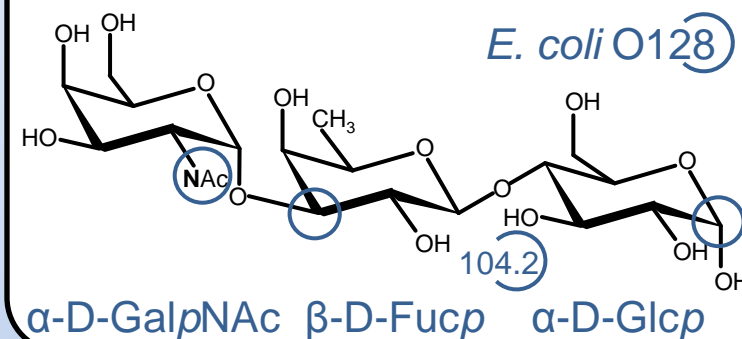
2dGlc → araHex,
α-Rib-ol → Rib-ol,
D-Kdo → Kdo,
1-methyl → 1-Me,
n.m.r. → NMR,
taxid 583 → Proteus,
...

выявляемые

Glc(1-2)GlcN,
anhydro-Kdo,
D-manHep,
Galp5N,
Ac(1-2)[Glc(1-2)]Gal,
Escherichia sapiens,
Dev Food Sci 2012,
#Ac : 23 м.д., 65 м.д.
D-Gcl, ...

невывявляемые

E. coli O127:
aDGalpN(1-4)bDFucp(1-4)bDGlc
Glc C1 103.2 ppm



CarbBank

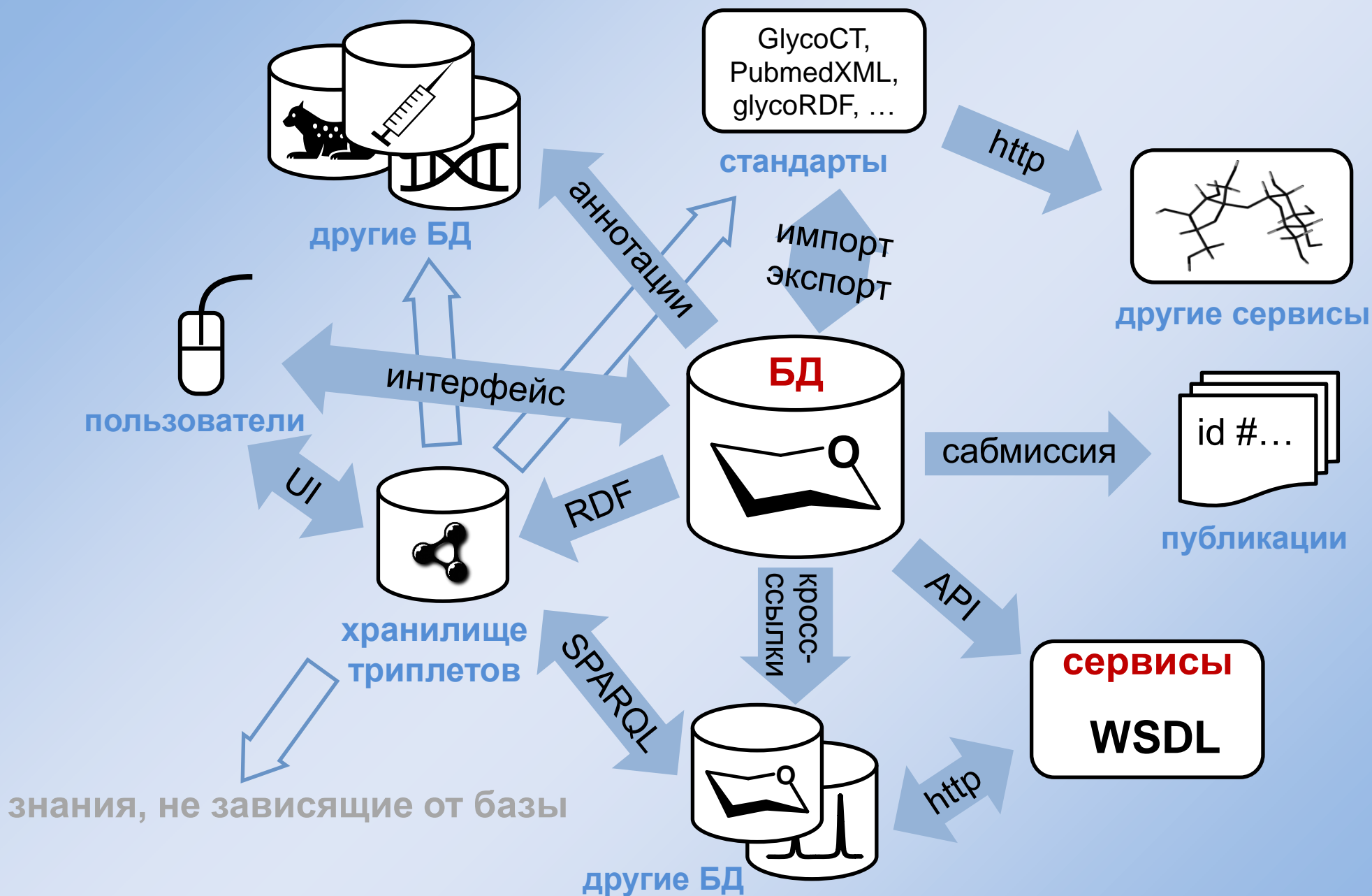
☹ >50% (неправильные, отсутствующие, ложно присутствующие структуры, штаммы, аннотации)

↳ другие БД

☺ <10%



Идеальная интеграция



Интерфейс

Конверсия данных ↔ другие форматы

Автоматические web-сервисы (WSDL)

Импорт, экспорт

Документация, HELP

Дружественность, быстрое действие

Ссылки на записи в других проектах (запросы, индексы, данные)

Ввод и вывод структур

SNFG,
WURCS,
GlycoCT,
SMILES,
MOL, PDB
Glyde II,
LinUCS,
Sweet-DB
GLYCAM,
GlycoRDF,
DCI XML, ...

DOI,
NCBI Pubmed,
NCBI Taxonomy,
Uniprot,
Genbank,
PubChem,
MonosaccharideDB,
Glytoucan,
ICD-11

Topology: 4 residues (branched: ([A->B->],[C->]D)) (A)→(B)→(D)

Structure: bDGalpNAc(1-3)aDGlc(1-4)[bDManp(1-3)]?LFucp

Residue (A): bDGalpNAc(1-

b D galactosamine (pyranose)

add substituent Acetylated at 2

add substituent

add substituent

add substituent

is terminal

Residue (B): aDGlc(1-

a D glucose (pyranose)

add substituent

add substituent

add substituent

wizard

Popular Small sugars Hexoses Higher sugars Alditols Aliphatic a

Glc GlnAc GlcA QuiNAc Gal GalNAc GalA Fuc FucNAc Man Rha LDmanH Ara Ara4N Xyl Fru

Novice Expert Insert Replace Oligo Poly Ac Am Cm Cho Fo M

Oligosaccharide

editor

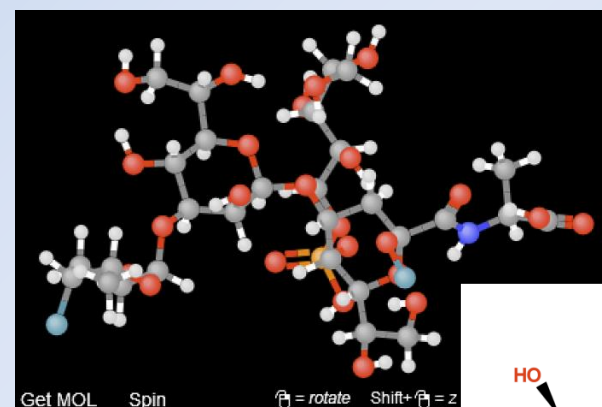
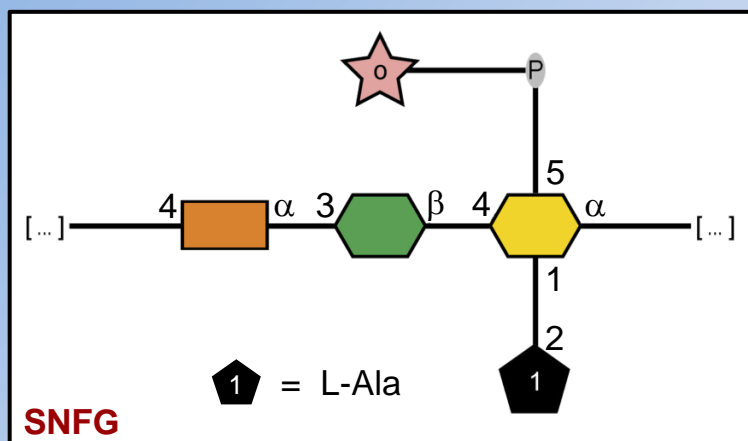
веб-помощник,
граф. редактор,
библиотека
структур,
CSDB Linear,
GlycoCT

bDManp(1-3)[Ac(1-2)bDGalpN(1-?)aDGlc(1-4)]?LFucp

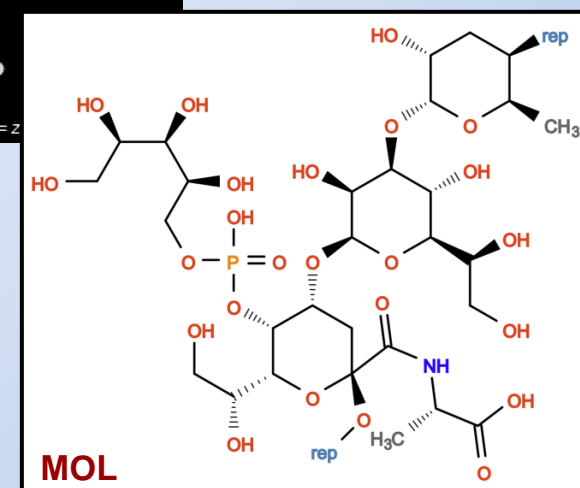
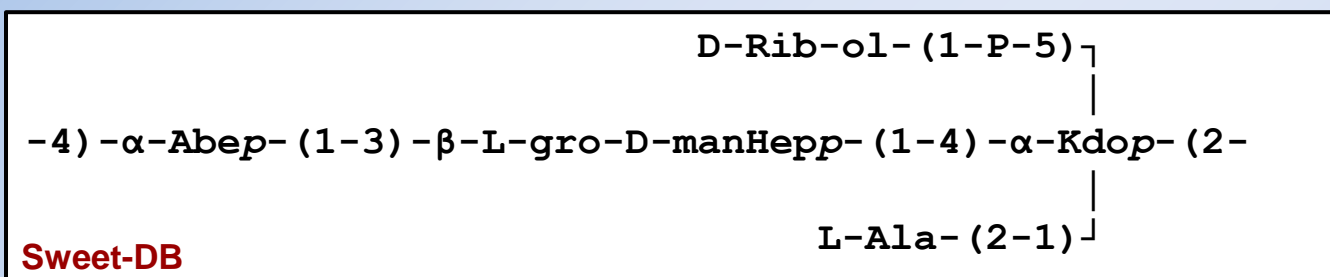
нотация CSDB Linear

Вывод структур

Визуализация в человекочитаемых форматах:



MOL (3D)



↔ Экспорт в машиночитаемых форматах:

[*]O[C@]1(C(=O)N[C@@H](C)C(=O)O)[C@@H](O[C@@H]2O[C@H]([C@@H](O)CO)[C@@H](O)[C@H](O[C@H]3O[C@H](C)[C@H](*)C[C@H]3O)[C@@H]2O)[C@@H](OP(=O)(O)OC[C@H](O)[C@H](O)[C@H](O)CO)[C@@H]([C@H](O)CO)O1

SMILES

2.0/5,5,5/[Aad1122h-2a_2-6][h222h][a11221h-1b_1-5][a2d12m-1a_1-5][A1m_2*N]/1-2-3-4-5/a1-e2_a2-d4~_a4-c1_a5-b1*OPO*/3O/3=O_c3-d1

WURCS

-4) aXAbep (1-3) bXLDmanHepp (1-4) [xDRib-ol (1-P-5) , xLAla? (2-1)] aXKdop (2-

CSDB Linear

Редакторы структур

редактор	мономеры	неопределенности	повторы	импорт, экспорт	хим. модель
GlycanBuilder™	70+30	конф., связи, классы, топология	-	++	-
Sugar Sketcher	70+20	конф., связи, классы	+	+	-
Polys builder	100+0	конф., классы	-	+	-
GlycoGlyph	80+10+user	конф., связи, классы	++	+	-
DrawRings	70+0	конф., классы	-	+	-
CSDB editor	300+300+user	конф., связи, классы, альтернативы	++	++	+

Углеводные,
SNFG

Химические

программа	плагин	мономеры	результат	оптимизация
Hyperchem	Sugar Builder	30+26+user	3D, можно задать φ, ψ, ω	MM+, Amber, OPLS: MM, MD, ...
PCModel	Template: sugars	24+28	3D	MM3, MMFF, OPLS: MM, MD, ...
ACDLabs Chemsketch	Template: sugars	36+35+...	2D: stereo, chair, Haworth, Fisher	MM2: MM
ChemDraw / Chem3D	Template: hexoses	16+24+...	2D: chair, Fisher	MM2, MMFF: MM, MD

Инструмент CSDB/SNFG editor

21

онлайн-редактор

Chemical repeating unit; n=10

Subst

Rib-ol

P

-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 Эксперт / Новичок
- любые поли- и олигомерные топологии
- 600+ моносахаридов и неуглеводных остатков
- SMILES для нестандартных фрагментов
- любые связи (в т. ч. хелатные и C-C)
- поддержка неопределенностей и суперклассов
- внутренние повторы и альтернативные ветви

Previews Refresh

Hi-res image

Subst = chrysin = SMILES O=c2cc(c1cccc1)oc3c{7}c(0)c{5}c(0)c23

«ЖИВОЙ» экспорт

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- //

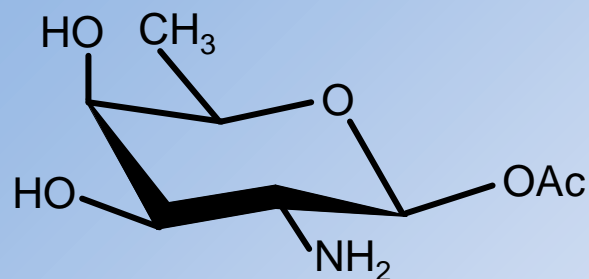
Files: MOL, PDB, Glycam

SMILES code:

```
[*][C@H]1O[C@H]([C@H](O)COC2O[C@H](CO[C@H]3O[C@H](C)[C@H](O)[C@H]([*])[C@H]3O)[C@H](OP(=O)(O)OC[C@H](O)[C@H](OC3cc(O)c4c(=O)cc(-c5cccc5)oc4c3)[C@H](O)CO)[C@H](O)[C@H]2O)[C@H](O)[C@H]1NC(C)=O
```

- экспорт в углеводные нотации
- экспорт в химические кодировки
- списки изомеров
- оптимизация и экспорт координат атомов
- визуализация, рендер, SNFG-окраска
- работа с мол. моделью в браузере
- импорт CSDB Linear и GlycoCT

Неоднозначность номенклатуры



bDFucpN(1-1)Ac (CSDB) ←

однозначно соотнесено со структурой, но при этом понятно людям

D-FucpN-β1OAc

beta-fucosamine acetate

1-acetoxy-beta-D-fucopyranosamine

2-deoxy-2-amino-β-D-fucopyranosyl acetate (IUPAC)

β-D-fucosamine acetic ester

β-6-deoxy-D-galactosamine acetate

b-dgal-HEX|1:5|2-amino|1-acetate (GlycoCT)


β-D-фукозамин-1-О-ацетат (на других языках)

(2S,3R,4R,5R,6R)-3-amino-4,5-dihydroxy-6-methyltetrahydro-2H-pyran-2-yl acetate (IUPAC)

N[C@H]([C@H]([C@H]([C@@H](C)O1)O)O)[C@@H]1OC(C)=O (SMILES)

1S/C8H15NO5/c1-3-6(11)7(12)5(9)8(13-3)14-4(2)10/h3,5-8,11-12H,9H2,1-2H3/t3-,5-,6+,7-,8+/m1/s1 (InChI)

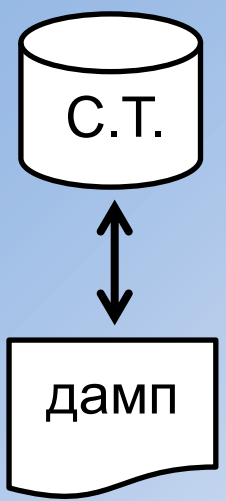
Почему не аналог MOL?

- Очень сложный перевод поатомного описания в семантическое (в структуру типа α -D-Galp-(1-3)- β -D-Glcp)
- Сложный перевод семантического описания в поатомное => трудоемкость аннотирования статей
- Нельзя описать структуры с неопределенностями
- Данные визуально не сопоставлены со знаниями

- Нечеловекочитаемый → трудно курировать → ошибки в данных
- Координаты атомов не являются первичными данными но неполный MOL (без 3D) может быть воспринят как 3D MOL
- Громоздко для хранения и передачи в сети (и не передается как параметр в URL)

52.0606	6.3910	-0.1606	O	0	0	0
51.8591	8.6986	-0.1875	N	0	0	0
52.9844	9.0584	0.7259	C	0	0	1
53.8550	8.9929	0.0662	H	0	0	0
52.9684	10.5530	1.3121	C	0	0	2
52.2705	11.0903	0.6993	H	0	0	0
1117	1	0	0	0	0	
1118	1	0	0	0	0	
1119	1	0	0	0	0	

атомы, координаты, связность

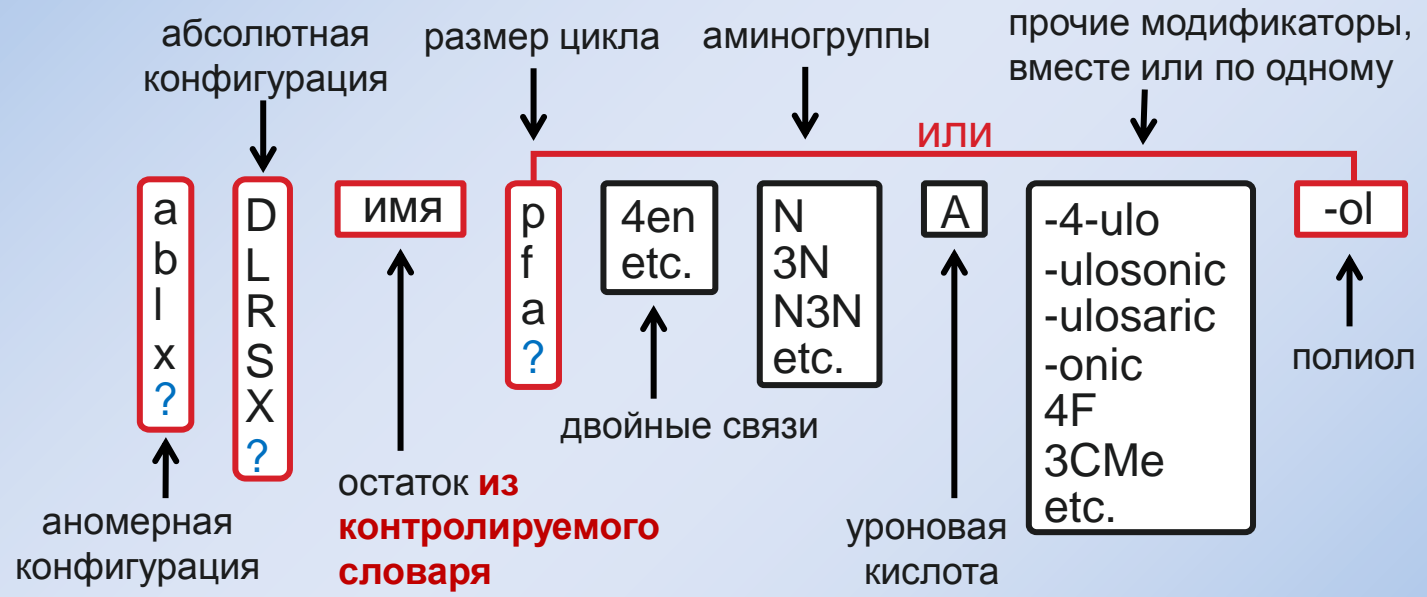
Особенности структур и язык



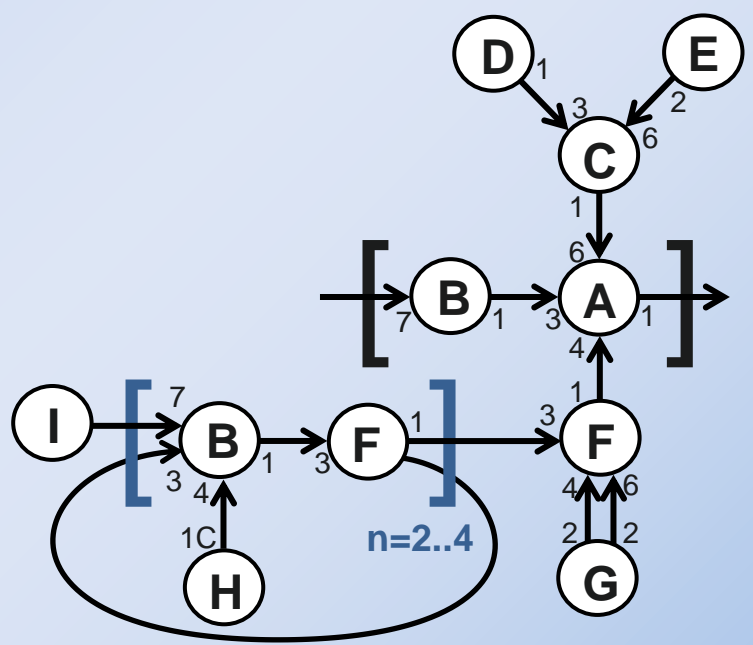
углеводный язык
CSDB Linear

аннотаторы,
операторы,
другие языки
(IUPAC, GlycoCT, WURCS, ...)

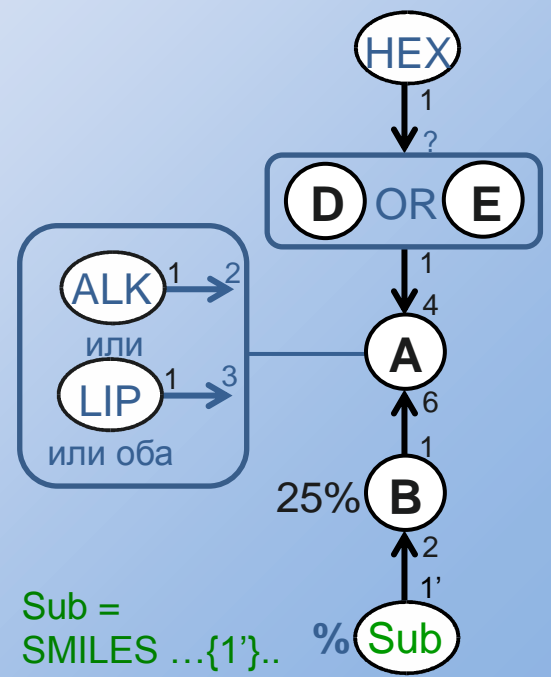
МОНОМЕРЫ
















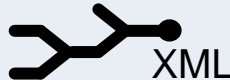






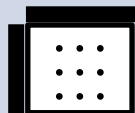











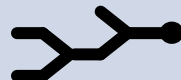























ТОПОЛОГИЯ



НЕТОЧНОСТИ

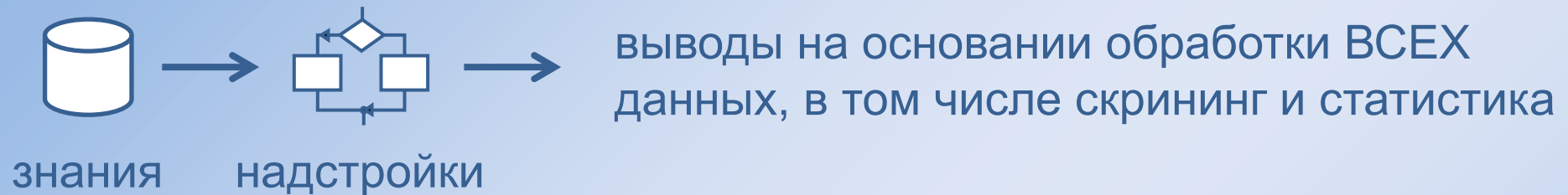


Сравнение языков

	<i>подход</i>	полнота	однозначность	контроль	парсинг	неточные структуры
IUPAC	 					
IUPAC extended (SweetDB, Carbbank)	pseudo-graphics 					
Glyde I	 XML 				 URL	
WURCS (JCGGDB, ChEBI, PDB)					 URL	
GlycoCT (Glycome-DB)						
LinearCode (CFG)					 URL	
LinUCS (GlycoSCIENCES)					 URL	
KCF (KEGG)						
CSDB Linear (CSDB)					 URL	

 СОВМЕСТИМОСТЬ

хуже     лучше



- Анализ путей биосинтеза (база гликозилтрансфераз)
- Конформационные карты олигосахаридов
- Предсказание и отнесение спектров ЯМР ^{13}C , ^1H , 2D
- Предсказание структуры по спектрам и другим данным
- Кластеризация таксонов на основании их гликомов
- Распределение фрагментов по таксонам и положению в структурах
- Классификация мономеров и агликонов

Гликозилтрансферазы

Критерии:
(в любом сочетании)

- идентификаторы в базах
- название фермента / группа
- название гена / кластер
- семейство CAZy
- организм (вид, штамм)
- синтезируемая связь
- донор (или его фрагмент)
- акцептор (или его фрагмент)
- роль объекта в клетке
- уровень достоверности

71B1
Uniprot
Q9LSY9.1
Genbank
821729

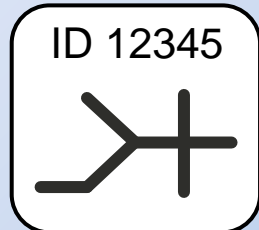
идентификаторы



Объект:

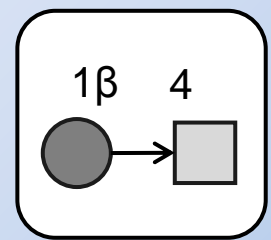


организм,
орган, ткань

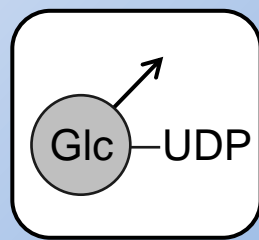


полная
структура

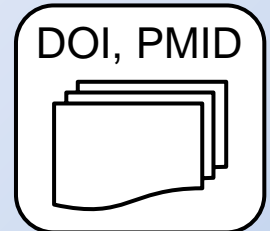
Активность:



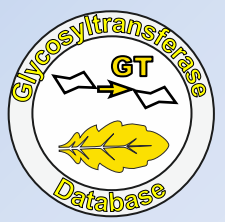
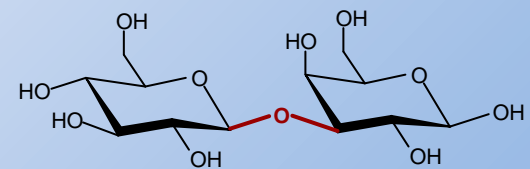
синтезируемый
фрагмент



донор и
акцептор



публикации



CSDB GT

Гликоферментные базы



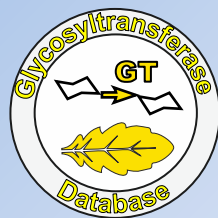
300K+ ферментов CAZy, подтверждено ~4%,
все организмы
 (полноценное использование – через соавторство)



гликозилтрансферазы, подтверждено 7%
E. coli (полная)



гликозилтрансферазы и др. CAZy
O. sativa



гликозилтрансферазы, подтверждено 80%
E. coli, A. thaliana, S. cerevisiae, (*ESKAPE* – в процессе)
 (полная по подтвержденным)



биосинтетическая информация среди прочей
все организмы



биосинтетическая информация среди прочей
A. thaliana



биосинтетическая информация, нет поиска
H. sapiens, M. musculus, R. norvegicus, ...

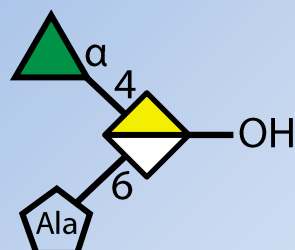


углеводные гены,
H. sapiens, C. elegans

Анализ конформаций

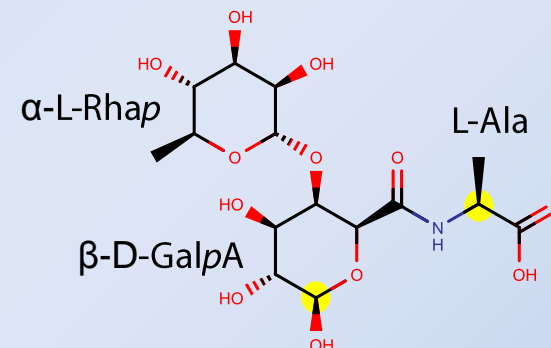
aLRhap(1-4)[x?Ala?(2-6)]?DGalpA

структуры,
в т.ч. неполные

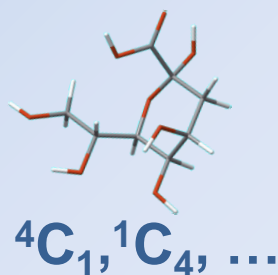


другие варианты
(α -GalA, D-Ala, и т. д.)

SMILES

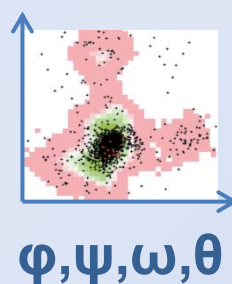


выгодные
конформеры
~1000 остатков

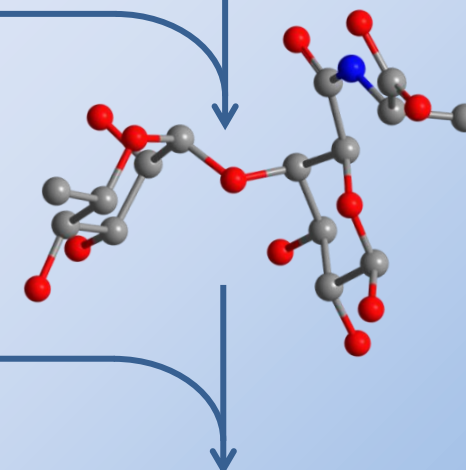


«креслификация»

заселенные
состояния
мостиков

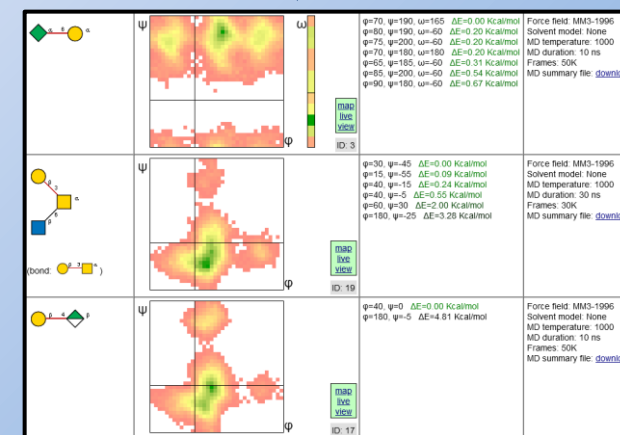


выбор минимумов
ММ-релаксация



мол. динамика
300К, 100нс, H₂O

конформеры
+ их энергии

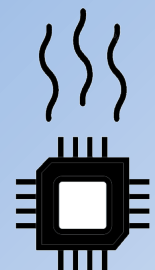


Конформационный модуль CSDB

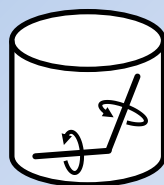
30



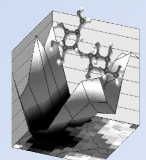
анализ
встречаемости
фрагментов



пайплайн
100-нс 300K MM3
MD с явной водой



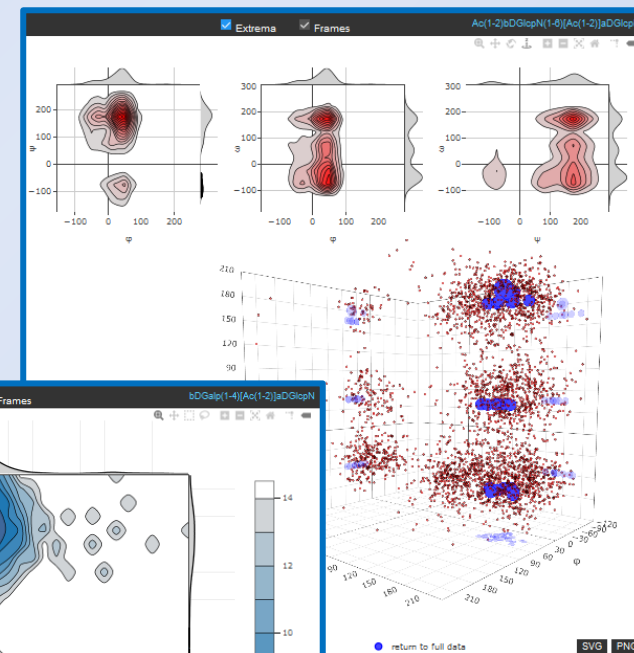
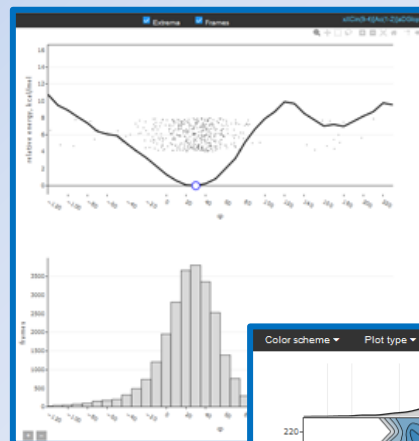
связи с 1-4 степенями
свободы в ди- и три-
мерных фрагментах
(заполнение базы)



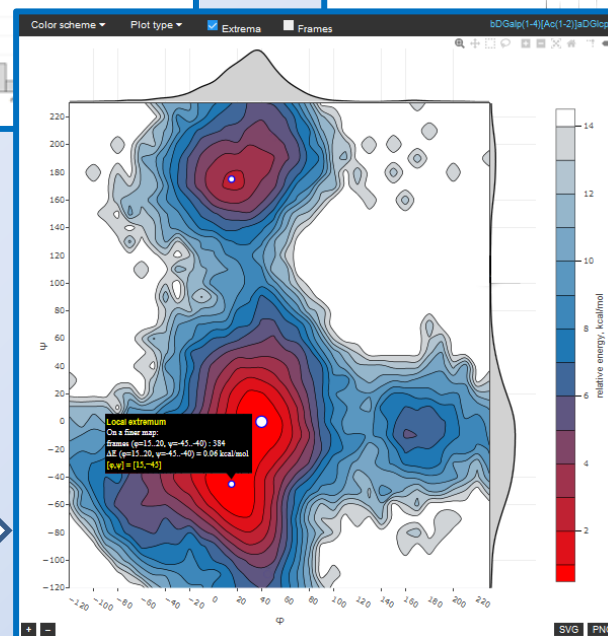
импорт GlycoMaps



1D



3D



2D

- поиск ID, структур, параметров расчета
- работа с картами энергий и заселенности
- проекции, экстремумы, экспорт

ЯМР – основной метод анализа первичной структуры в гликобиологии

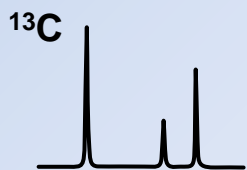


- Моделирование спектров (^{13}C ~ 0.7 ppm, ^1H ~ 0.06 ppm)
- Помощь эксперту в установлении структуры
- Отнесение сигналов и проверка гипотез
- Перебор структур и сравнение симуляции с экспериментом
- Верификация расчетной геометрии молекул

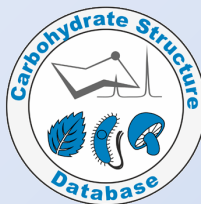
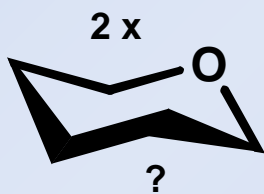
Итератор структур



неотнесенный
спектр ЯМР

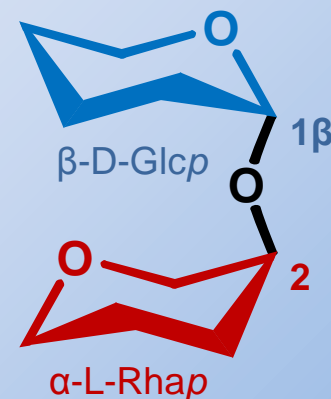


известные
данные
о структуре



алгоритм GRASS

Generation, Ranking
and Assignment of
Saccharide Structures



полная структура
(оставшиеся
неизвестные)

мономеры, конфигурации,
модификации, позиции
замещения,
последовательность

Как воспользоваться?

S. Scherbinina, Ph. Toukach **3D structures of carbohydrates and where to find them** (2020) *Int J Mol Sci* **21**, ID 7702. doi: [10.3390/ijms21207702](https://doi.org/10.3390/ijms21207702)

Abrahams J. et al. **Recent advances in glycoinformatic platforms for glycomics and glycoproteomics** (2020) *Curr Opin Struct Biol* **62**, 59-69. doi: [10.1016/j.sbi.2019.11.009](https://doi.org/10.1016/j.sbi.2019.11.009)

K.F. Aoki-Kinoshita **A practical guide to using glycomic databases** (2017) Springer. doi: [10.1007/978-4-431-56454-6](https://doi.org/10.1007/978-4-431-56454-6)

Ph. Toukach, K. Egorova **Carbohydrate Structure Database merged from bacterial, plant and fungal parts** (2016) *Nucl Acid Res* **44**, D1229–36. doi: [10.1093/nar/gkv840](https://doi.org/10.1093/nar/gkv840)



<http://glytoucan.org>



http://jcggdb.jp/index_en.html

T. Lütke **The use of glyco-informatics in glycochemistry** (2012) *Beilstein J Org Chem* **8**, 915-929. doi: [10.3762/bjoc.8.104](https://doi.org/10.3762/bjoc.8.104)

[GLYCO-SCIENCES.DE](http://glycosciences.de)

<http://glycosciences.de>



<http://www.genome.jp/kegg/glycan/>



<http://www.unicarbkb.org/>



<http://csdb.glycoscience.ru>

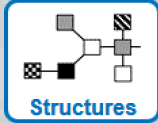
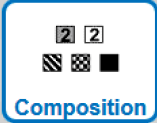


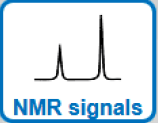


<http://toukach.ru/rus/glyco-db.htm>



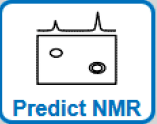
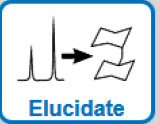
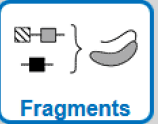
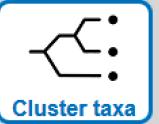


Сайт в Интернете

Database search

 Structures
  Composition
  Organisms
  Publications
  NMR signals

Additional operations are available from the [left menu](#). If you don't see it [click here](#)

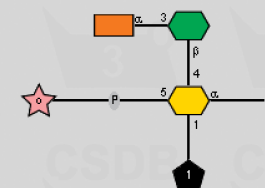
Useful tools

 Predict NMR
  Elucidate
  Fragments
  Cluster taxa
  GT activities
  Examples

NMR spectrum simulation

Please, select how to input a structure:

- [Input using Structure Wizard](#)
- [Select from library](#)
- [Draw in Glycan Builder](#)
- [Convert from GlycoCT](#)
- [Use expert form \(field below\)](#)



Structure in CSDB encoding:

`aXAbep(1-3)bXLDmanHepp(1-4)[xDRib-ol(1-P-5),xLAla?(2-1)]aXKdop`
(this field is editable) [Help on structure encoding](#)

Nucleus: [?](#)

More parameters... [?](#)

Solvent: [?](#) [Coverage](#) [?](#)

Carbohydrate Structure Database

Prokaryotes » Plants » Fungi

7006 publications (1941-2017):
18923 compounds from
8859 organisms
last update: 2017 Jun 2

Search

- CSDB IDs
- (Sub)structure
- Composition
- Taxonomy
- Bibliography
- NMR signals

Help

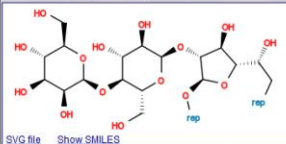
Extras

- NMR simulation
- Elucidation from NMR
- Monomer namespace
- Fragment abundance
- Coverage stats
- Taxon clustering
- Submit record
- Translate structure
- Feedback

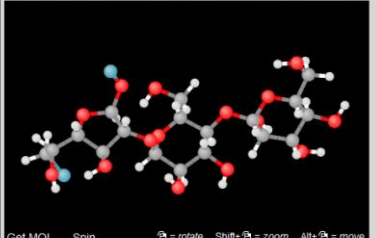
Maintenance

Related record ID(s): [101](#)
NCBI Taxonomy refs (TaxIDs): [64489](#)

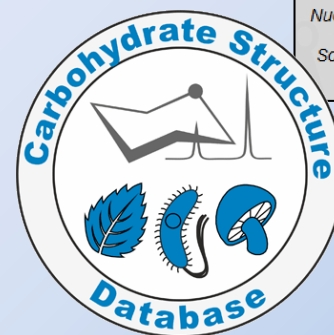
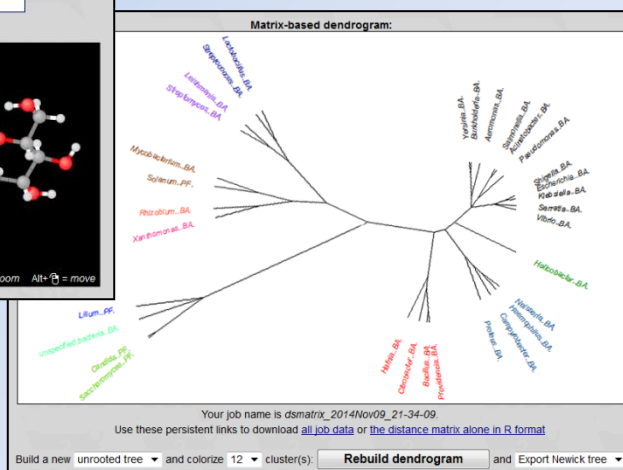
There is only one chemically distinct structure:



[SVG file](#) [Show SMILES](#)



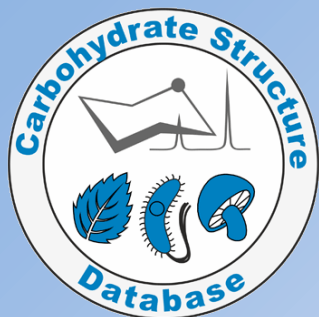
[Get MOL](#) [Spin](#) [rotate](#) [Shift+ = zoom](#) [Alt+ = move](#)



<http://csdb.glycoscience.ru>

- свободный доступ
- подробная документация
- примеры решения задач

Участники и спонсоры











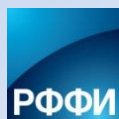
Carbohydrate Structure Database

проверяемый контент, полное покрытие, сопутствующие сервисы



программирование
аннотирование и проверка данных
общая поддержка и сбор данных
интеграция, онтология
конформационный анализ
идеи, язык, архитектура, интерфейс,
программирование, координация
партнеры

-   Роман Капаев, Андрей Бочков, Иван Чернышов, ...
-  Ксения Егорова, Надежда Калинчук, Кирилл Казанцев, ...
-  Юрий Книрель
-   Рене Ранцингер, Кийоко Аоки-Киношита, Томас Люттеке, ...
-  Виктор Стройлов, Софья Щербинина, ...
-  Филипп Тоукач



Российский Фонд
Фундаментальных
Исследований

2005-2007,
2012-2020 (x3)



Международный
Научно-Технический
Центр

2004-2005



Комиссия по
грантам при
президенте РФ

2005-2007



Немецкий Центр
Исследования
Рака

2007-2010 (x4)



Фонд Содействия
Отечественной
Науке

2008-2009



Российский
Научный
Фонд

2018-2020
2021-2022

Дополнения

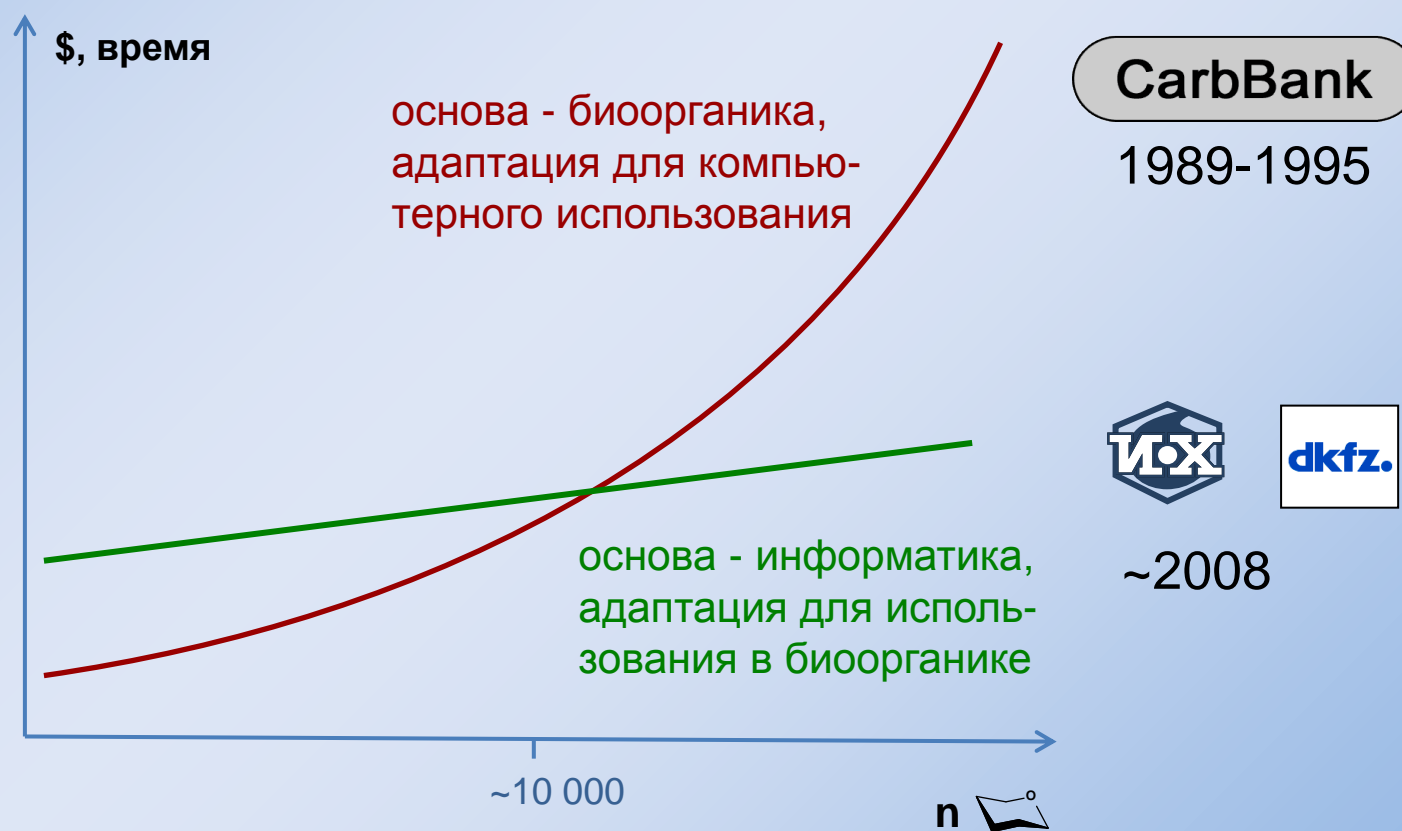
(демонстрируются в рамках ответов на вопросы)

Примеры вопросов к CSDB

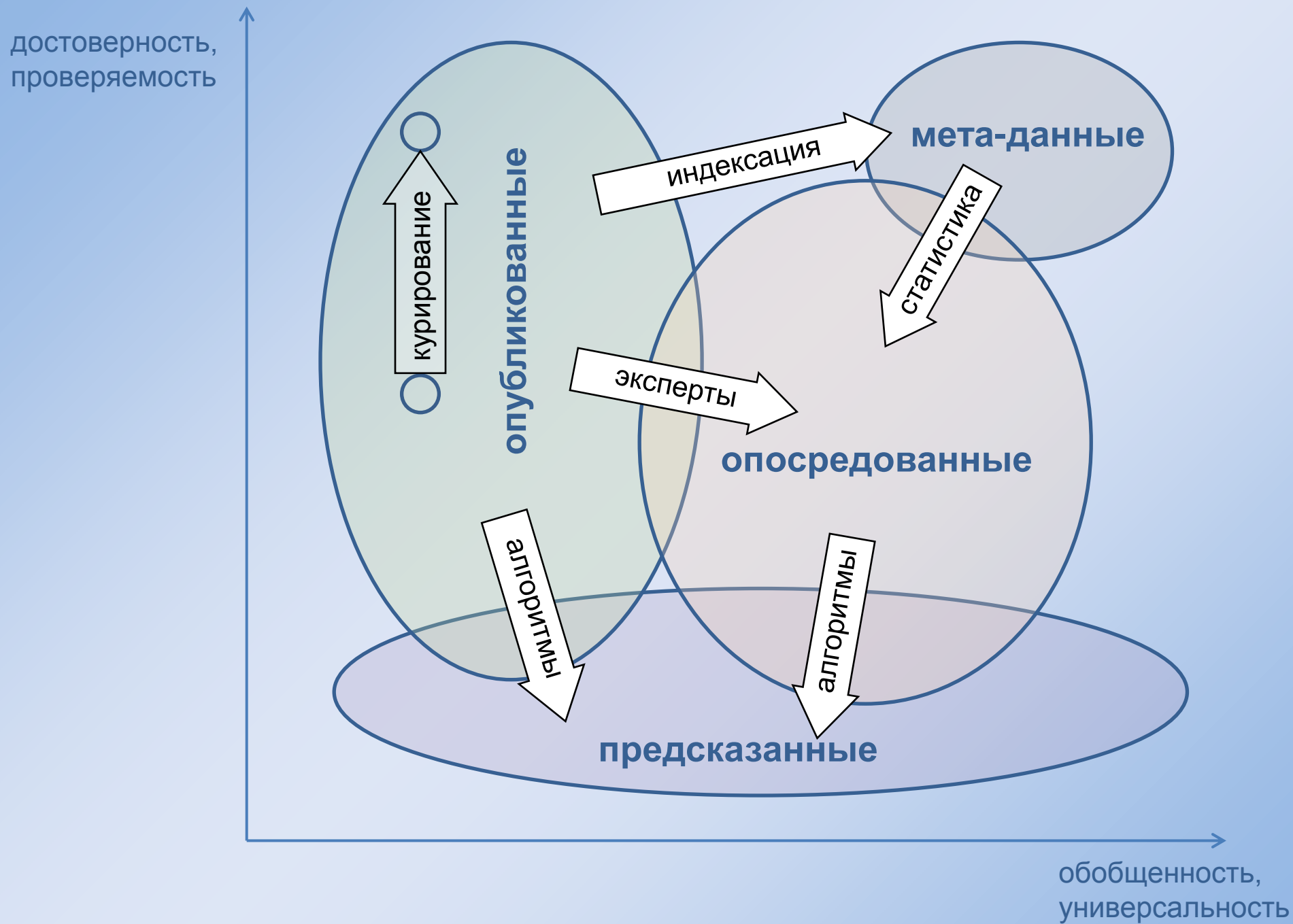
- Как введение аминогруппы влияет на химические сдвиги ЯМР в лактозном фрагменте?
- Какие структуры О-антигенов, содержащие галактуроновую кислоту и еще как минимум одну гексозу, были опубликованы после 2005-го года?
- Какие гликозиды, выделенные из растений рода паслёна, содержат агликон соланидин?
- Какие углеводы, кроме октозосодержащих, имеют сигнал ЯМР ^{13}C около 34 м.д. ?
- Какие бактериальные структуры, опубликованные А.С. Шашковым или Ю.А. Книрелем, содержат хиновоз-4-амин, амидированный любой N-ацетилированной аминокислотой?
- Гомополимеры каких нонапираноз встречаются в бактериях?
- Каков ожидаемый спектр ЯМР ^{13}C 3-О- α -абеквозил-6-деокси- β -D-манногептопиранозил-(D-рибитол-1)-фосфата в воде и на основании каких источников предсказаны химические сдвиги, для которых указана наименьшая достоверность?
- Какова наиболее вероятная последовательность остатков бациллозамина, глюкуроновой кислоты и лизина в олигомере с указанным экспериментальным спектром ЯМР ^{13}C ?
- Какие моносахариды склонны занимать концевые позиции в гликанах Аспергилла дымящего и Аспергилла кодзи?
- Какие димерные фрагменты (включая дисахариды) гликанов высших растений специфичны для рода люпинов?
- Для скольких гликанов протеобактерий опубликованы спектры ЯМР?

Организация разработки

- База данных и платформа должны быть построены по правилам информатики.
- Правила были конкретизированы и адаптированы для углеводов.



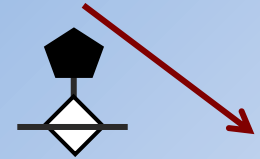
Уровни данных



Уровни абстракции

-3) [xLAla (2-6) ,Ac (1-2)]bDGalpNA (1-

точный фрагмент и его связи

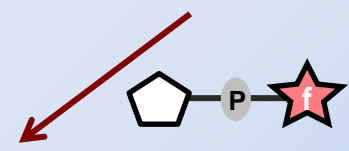


L-Ala (2-6) +

-5) α-D-Ribf(1-3) β-D-GalpNAcA (1-2) D-Gro (1-P-

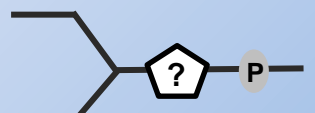
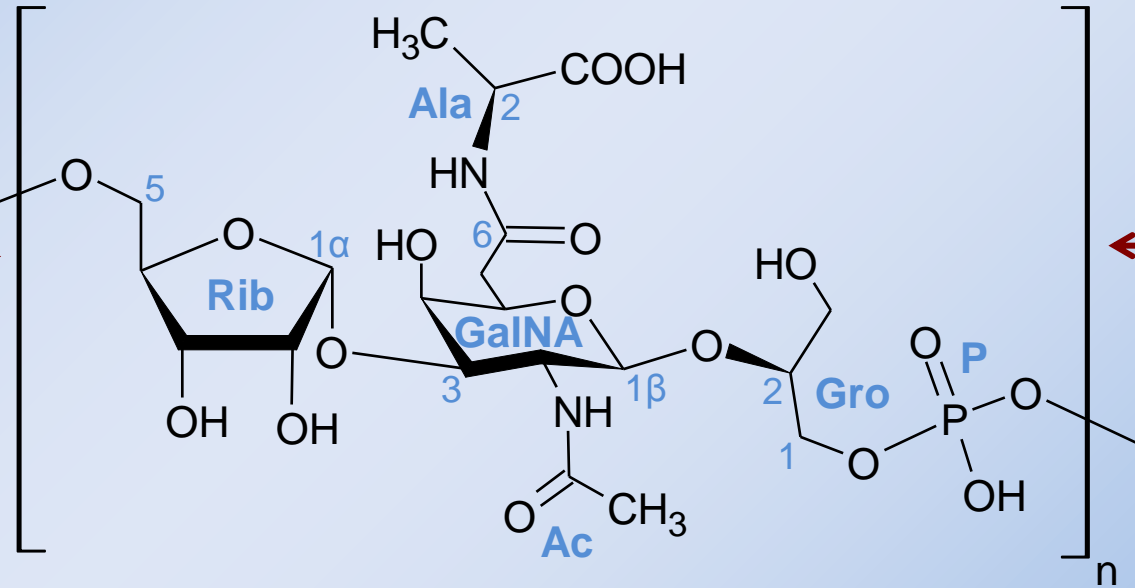
xDGro (1-P-5) aDRibf

не указана топология и места присоединения



-?) ?Dhex (1-

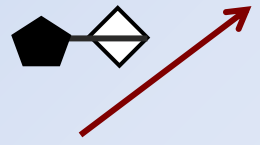
указаны только класс и 1-связь



приблизительный мотив

x?Ala (2-?) ?DGal?NA

не указаны конфигурации, позиции замещения, размер цикла, N-ацетилирование



HEX, xDRib?, PER

частичный состав



Resource Description Framework

RDF – модель данных в виде триплетов *объект-предикат-субъект*.

- 😊 допускает распределенные запросы с минимальным знанием форматов баз
- ☹️ требует репозитория триплетов и согласованной онтологии



Задача: найти белок-носитель для произвольного гликана из JCGGDB.

Проблема: JCGGDB не связана ссылками с белковыми базами.

Преамбула:

Записи в JCGGDB имеют ссылки на идентификаторы в GlycomeDB.

Как GlycomeDB, так и UniCarbKB могут экспортировать структуры в формате GlycoCT.

Записи в UniCarbKB имеют ссылки в белковую базу UniProt.

Решение *(9-строчный скрипт на SPARQL)*:

Сопоставить идентификаторы JCGGDB и UniCarbKB, используя GlycomeDB, и получить идентификаторы UniProt из UniCarbKB для каждого идентификатора JCGGDB.



Требуется:

стандартная онтология → экспорт данных в RDF → репозиторий триплетов → интерфейс SPARQL


GlycoRDF – первая формальная углеводная онтология (OWL)



Дальнейшее развитие

42

● сделано в CSDB ● предстоит сделать ● близко к завершению

- **Стандартный человекочитаемый язык** (SNFG, CSDB Linear, ...)
- **Расширение онтологий** (интеграция через GlycoRDF и GlycoCoO)
- **Кросс-проектные сервисы**
(ввод-вывод структур, конформационные расчеты, предсказание свойств)
- **Стандартные индексы**
(Glytoucan ID, MSDB, PMID, DOI, TaxID, ICD-11, PDB id, Genbank, ...)
- **Стандартные протоколы** (API, WSDL, SPARQL, REST, ...)
- **Идеологическая замена CarbBank** 
- **Требование включать ID в публикации** (Glytoucan ID?)
(кто будет чистить базы от неопубликованных/ошибочных данных?)

Structure wizard

Topology: 3 residues (linear: A->B->C) (A)→(B)→(C)

Structure:

Residue (A):

()

[aLFucp](#)

substitutes of Residue B

is terminal

add substitution
 add substituent Acetylated at
 add substituent
 add substituent
 add substituent

Residue (B):

()

[DRib-ol](#)

substitutes of Residue C

add substitution
 add substituent
 add substituent
 add substituent

Residue (C):

()

[a?6dTal?](#)

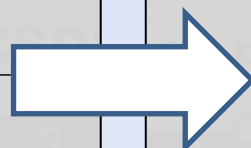
has aglycon:

add substitution
 add substituent
 add substituent
 add substituent

Structure in CSDB encoding:

[Return the structure to the search page and close this window](#)

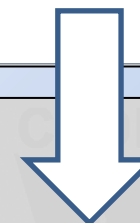
[Home](#) [Help](#)



Glycan Builder

File Edit Structure View Help

Linkage: Chirality: Ring:



Search for (sub)structure

Please, select how to input structure:

- [Input using Structure Wizard](#)
- [Select from library](#)
- [Draw in Glycan Builder](#)
- [Convert from GlycoCT](#)
- [Copy from the previous query \(aDFucp3N\)](#)
- [Use expert form \(field below\)](#)

Structural fragment in CSDB encoding:

(this field is editable) [Help on structure encoding](#)

Only those containing text: in aglycons, aliases or linear code in trivial names

Search scope:

Search the whole database
 Search in the result of the previous query (logical AND)
 Combine with the result of the previous query (logical OR)
 Negate search (find results NOT matching current query)

Treat search term as a
 Search for molecule types:
 Search for structures with published NMR data only
 Restrict compound class:
 Restrict taxonomical domain:

Previous results: 122 structures: [<ID list>](#)

& display records per page.

[Predict NMR](#) [Sweet 3D model](#) [Home](#) [Help](#) [HELP !!!](#)

Поиск по таксономии

Found **12** organisms. Displayed organisms from **1** to **12**
[Expand all organisms](#) [Show all as text \(SweetDB notation\)](#)

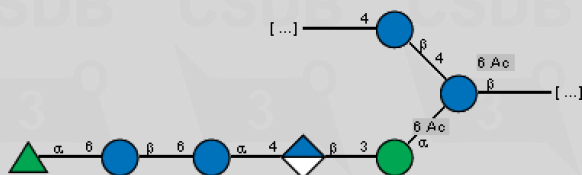
1. (Organism ID: 1005)

[Acetobacter xylinum](#)
 (Ancestor NCBI TaxID 28448, [species name lookup](#))

Later renamed to: [Komagataeibacter xylinus](#)
 Taxonomic group: bacteria
 Phylum: Proteobacteria

The following compound(s) are assigned to this organism:

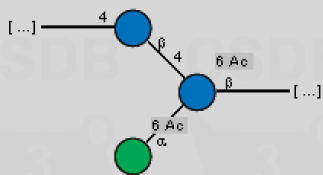
- Compound ID: 1717



[Show legend](#)
[Show as text](#)

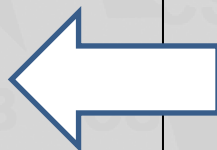
Carbohydrate Research 2004, "Synergistic interactions between the genetically modified bacterial polysaccharide P2 and carob or konjac mannan"
[CSDB ID 9262](#) (all data & tools)

- Compound ID: 1720



[Show legend](#)
[Show as text](#)

Carbohydrate Research 2004, "Synergistic interactions between the genetically modified bacterial polysaccharide P2 and carob or konjac mannan"
[CSDB ID 9414](#) (all data & tools)



Search for organism

Display domains: bacteria archaea protista algae fungi plants animals

Genus:

Species:

Strain / subspecies:

Specify:

Search scope:

Search the whole database Search among HOST organisms
 Search in the result of the previous query (logical AND) Use NCBI taxID
 Combine with the result of the previous query (logical OR) Include subtaxons
 Negate search (find results NOT matching current query)

Previous results: 6 structures: [<ID list>](#)

& display records per page.

[List of organisms](#) [Home](#) [Help](#)

Process taxonomy in NCBI Taxonomy DB (fields are editable):

Genus: Species:

Поиск по библиографии

Found 3 publications. Displayed publications from 1 to 3

[Expand all publications](#) [Show all as text \(SweetDB notation\)](#)

1. (Article ID: 1525)

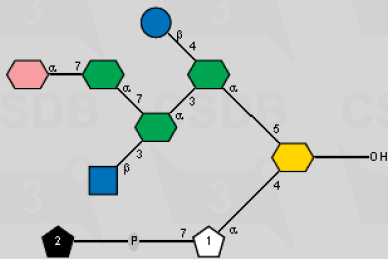
Knirel YA, Lindner B, Vinogradov EV, Shaikhutdinova RZ, Senchenkova SN, Kocha
Cold temperature-induced modifications to the composition and structure of lipopolysaccharide of Yersinia pestis
Carbohydrate Research **340(9)** (2005) 1625-1630

Following a report of variations in the lipopolysaccharide (LPS) structure of Y. pestis grown at 15 degrees C and flea (25 degrees C) temperatures, a number of changes to the structure of the bacterium was cultivated at a temperature of winter-hibernating rodents (6 degrees C). The LPS of the known Y. pestis LPS types, LPS of a new type was isolated from Y. pestis KM218. The latter differs in: (i) replacement of terminal galactose with terminal mannose; (ii) phosphorylation of terminal oct-2-ulosonic acid with phosphoethanolamine; (iii) the absence of glycine; lipid A differs in the lack of any 4-amino-4-deoxy-4-phosphoryl group; (iv) the presence of a 4-amino-4-deoxy-4-phosphoryl group (di)oxygenation of a fatty acid(s). The data obtained suggest that cold temperature-induced modifications of control of the synthesis of Y. pestis LPS.

Lipopolysaccharide, structure, core, modification, agent, composition, Yersinia pestis, Plague

The publication contains the following compound(s):

• Compound ID: 4209

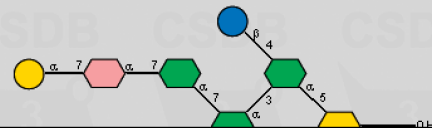


1 = a-Kop
2 = EtN

[Show legend](#)
[Show as text](#)

Yersinia pestis KM218
[CSDB ID 10076](#) (all data & tools)

• Compound ID: 4210



Search for bibliography

Authors: "Knirel YA" OR Toukach Author index start with: tou
[Help on author/keyword query syntax](#) [ă ö ü á é í ó ç š](#)

Title: pestis OR plague* search also in abstract
(content of title) [Help on title/abstract query syntax](#)

Keywords: structure? OR composition? search also in title
(content of keyword section) [Help on author/keyword query syntax](#)

Journal: Carbohydrate Letters
Carbohydrate Polymers
Carbohydrate Research
Cell
Cell Chemical Biology
Cell and Tissue Research

Year: 1983
1984
1985
1986
1987
1988
1989

Vol: *

Page: *

Search scope:

- Search the whole database
- Search in the result of the previous query (logical AND)
- Combine with the result of the previous query (logical OR)
- Negate search (find results NOT matching current query)
- Publications with structure elucidation only
- Restrict taxonomical domain: All domains

Previous results: 3 publications: 2953,201,1525

Go! & display 30 records per page.

[PubMed XML](#)

[Home](#)

[Help](#)

Author index:

[Toubetto K](#) [Toussaint A](#)
[Toukach FV](#)

The listed author names start with 'Tou'.
Click an author name to copy it to the author field in the caller form.

[Close this window](#)

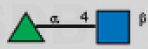
Поиск конформаций

Search for disaccharide conformation maps

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

Conformation ID:

Model:
(only those components are listed for which conformation maps are stored)

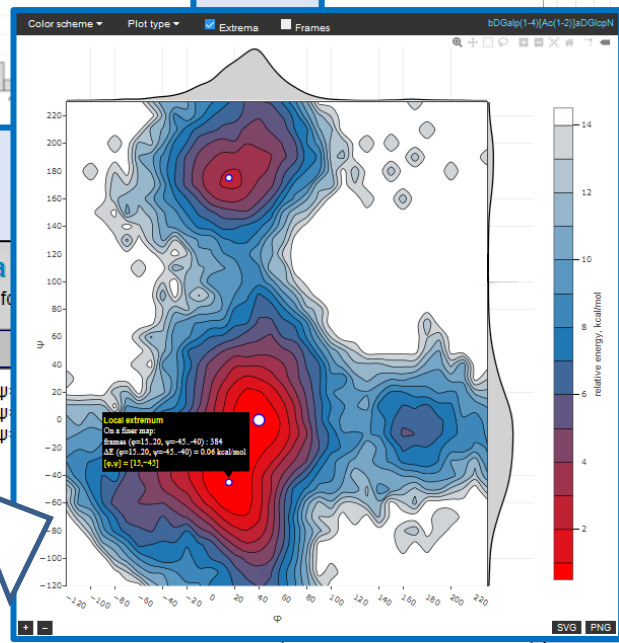
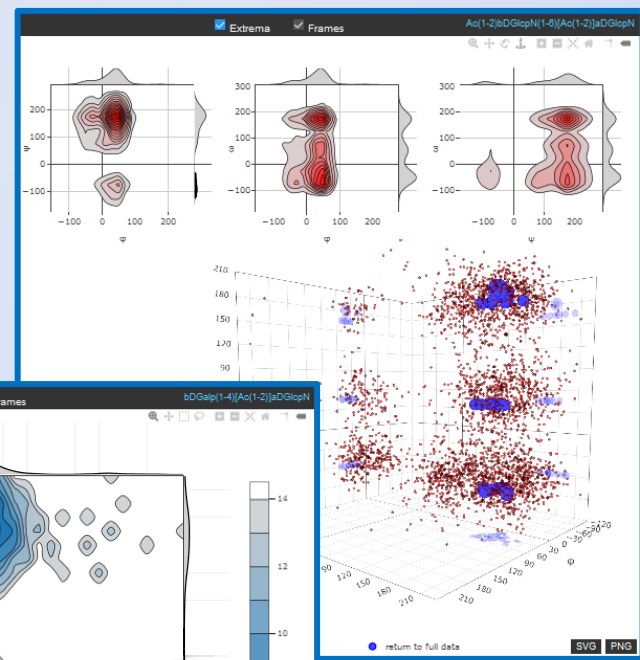
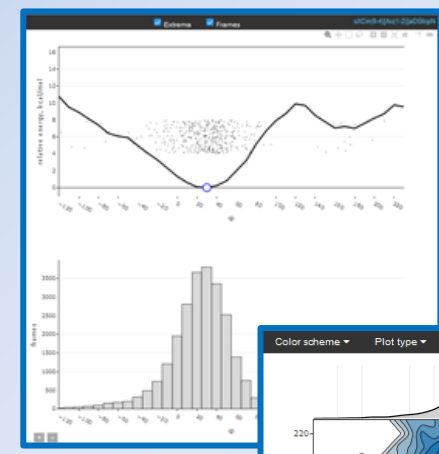
or type dimeric fragment in CSDB encoding
 


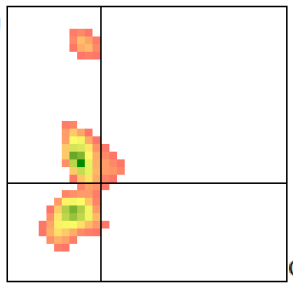
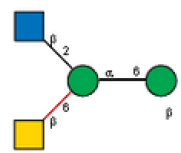
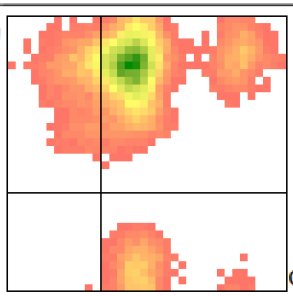
Force field:

Temperature:

Solvent model:

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



CSDB conformation data			
12 conformation maps have been found			
Model structure	Conformation map		
		ψ	$\phi = -30, \psi = -35, \omega = -25$
		ψ	$\phi = 25, \psi = 165, \omega = -75$ $\phi = 45, \psi = 170, \omega = -60$ $\phi = 45, \psi = 195, \omega = -75$ $\phi = 50, \psi = 185, \omega = -60$ $\phi = 25, \psi = 165, \omega = 180$ $\phi = 30, \psi = 155, \omega = -60$ $\phi = 35, \psi = 180, \omega = 180$ $\phi = 45, \psi = 215, \omega = -60$ $\phi = 40, \psi = 170, \omega = 165$ $\phi = 20, \psi = 185, \omega = 165$ $\phi = 55, \psi = 195, \omega = 165$ $\phi = 55, \psi = 155, \omega = -60$ $\phi = 45, \psi = 185, \omega = 165$ $\phi = 30, \omega = 145, \omega = -60$

map live view

ID: 1610

map live view

ID: 907

Force field: MM3-1996
 Solvent model: None
 MD temperature: 1000
 MD duration: 30 ns
 Frames: 30K
 MD summary file: [download](#)

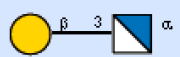
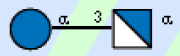
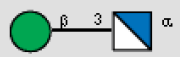
3D

2D

Поиск гликозилтрансфераз

CSDB glycosyltransferase search

42 glycosyltransferase activities have been identified in the CSDB database.
Please note that GTR database covers only two species: *Escherichia coli* and *Streptococcus pneumoniae*.

Enzyme	Gene	Activity
Name: WbbD UniProt ID: Q03084*	?	Synthesized dimer: bDGalp(1-3)aDGlcPn  Donor (ID 19342): DGalp(1-P-P-5)nucU Acceptor (ID 19715): Ph(1-11)[Ac(1-2)aDGlcPn(1-P-1)]Subst // Subst = undecan-1,11-diol Status: evidence <i>in vitro</i> ? Confirmation methods: <i>in vitro</i> (crude extract) ID: 2053
Name: WbbG UniProt ID: Q0H8C8*		Synthesized dimer: aDGlcP(1-3)aDGlcPn  Status: indirect evidence <i>in vivo</i> ? Confirmation methods: mutation (knockout) Notes: Repeating unit of the O148 antigen. ID: 2151
Name: WbaD UniProt ID: Q1L815*	Name: wbaD GenBank ID: 7156002*	Synthesized dimer: bDManp(1-3)aDGlcPn  Donor (ID 19855): DManp(1-P-P-5)nucG


CSDB glycosyltransferase search



Use the following conditions alone or in any combination to search for glycosyltransferases. Any field may be left blank for no restrictions.

GT names and IDs: Type enzyme name, e.g. "Orf10". Wildcards (* and ?) are supported.
 Enzyme name:

Organism: Select species Type strain/serogroup

Molecule role: Filter by target structure

Synthesized bond: Type dimeric fragment in CSDB encoding or use tools
 [Use Wizard](#) 

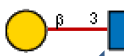
Donor & acceptor: Type donor CSDB encoding or use tools
 [Use Wizard](#) 
Type acceptor CSDB encoding or use tools
 [Use Wizard](#) 


Treat donor/acceptor as fragments

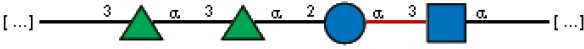
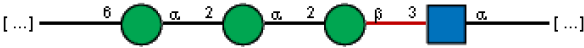
Confirmation status: Filter results to those

[Search!](#)

[Home](#) [Help](#) [HELP !!!](#) ?

Organism (ID 8...)
 Full structure (ID...)

 1 = u...
 Analog in...



 CSDB ID(s): 11572 , 21578 , 23062 , 26257 , 27289 Molecule role: O-antigen	
Organism (ID 1863): <i>Escherichia coli</i> O77 Full structure (ID 4600): 	Zhou et al. 2016 DOI: 10.1016/j.carres.2016.02.007 Wang et al. 2007 DOI: 10.1099/mic.0.2007

Кластеризация таксонов

Scope settings

Limit taxonomical scope to: **phylum**

Display groups: bacteria archaea protista algae fungi plants animals

Phylum: (select multiple with CTRL key)

- (unspecified bacteria)
- (unspecified protista)
- Actinobacteria
- Bacteroidetes
- Chlamydiae
- Chloroflexi**
- Crenarchaeota
- Cyanobacteria

General settings

species Rank of taxons to compare (should be lower than selected scope). [Specify exact species \(all\)](#)

50 **Taxon population threshold.** Minimal number of structures* assigned to a taxon or its subtaxons, to include this taxon in calculation (affects selection of taxons). Check to use this filter.

15 % **Normalized taxon population threshold.** Minimal part of structures* assigned to a taxon or its subtaxons, to include this taxon in calculation (affects selection of taxons). Normalized by the total number of structures* in the database. Check to use this filter.

50 **Structure abundance threshold.** Minimal number of structures* in which a fragment should be contained to be qualified as 'present in biota' (affects selection of fragments)

60 **Fragment abundance threshold.** Minimal number of instances* in which a fragment should be present to be qualified as 'present in biota' (affects selection of fragments)

2 **Fragment presence threshold.** Minimal number of instances* in which a fragment should be present in organisms of a taxon to be qualified as present in this taxon (affects occurrence codes and thus, taxon dissimilarity)

two residues **Type of fragments to analyze (dimeric or monomeric)**

only polymers **Type of structures to analyze.** Only structures of this type are considered in fragment analysis and where marked (*). 'Optimized' = only polymers from bacteria, archaea and fungi, and only mono/oligomers from plants.

R-project **Format of the dissimilarity matrix**

Fragment pool generation settings

- Combine anomeric forms.** All sugar residues will be treated as 'any anomer'
- Exclude underdetermined residues.** Residues with unknown anomeric, absolute or ringsize configuration will be omitted from analysis.
- Exclude monovalent residues.** Residues like Me, Ac, etc. will be omitted from analysis. Please note, that Ac in N-acetylated amnosugars is a separate residue.
- Exclude superclasses.** Fragments with residues represented by aliases and superclasses will be omitted from analysis.
- Differentiate aliases.** Residue aliases (used for atypical residues) will be differentiated by actual residue names, otherwise they are combined under an alias name.
- Sugars only.** Fragments with non-sugar residues (including monovalent residues, like N-acetyls) will be omitted from analysis.
- Exclude aglycons.** Fragments with atypical residues at non-reducing ends will be omitted from analysis.
- Differentiate location.** The same fragments at different locations (inline, terminal, reducing) will be treated as different.
- Strict comparizon** of fragments. Unknown configurations and ringsizes are always unequal to those known (otherwise a fuzzy comparizon is performed).

Distance matrix based on fragment presence

The analysis was performed over all cellular organisms

Prepared 20 monomers
Prepared 32 genera
Generated occurrence bit-codes. [Show](#)
Generated dissimilarity matrix. [Show](#)

Calculation parameters:

Hamming mode: YES
Fragment size: monomer
Fragment abundance filter: instance threshold: 550
Fragment abundance filter: structures threshold: 500
Fragment presence threshold: 2
Differentiate structures of this type: any
Filter: differentiate monomer positions (inline/terminal/reducing) in structures: NO
Matrix data format: R

Coverage data on used taxons:

(taxons, number of organisms in a taxon, number of structures assigned to these organisms)

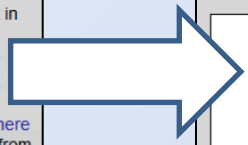
Acinetobacter (BA)	68	140
Aeromonas (BA)	64	122
Bacillus (BA)	95	234
Burkholderia (BA)	36	219
Solanum (PF)	46	127

Matrix-based dendrogram:

Your job name is *dsmatrix_2014Nov09_21-34-09*

Use these persistent links to download [all job data](#) or [the distance matrix alone in R format](#)

Build a new **unrooted tree** and colorize **12** cluster(s): **Rebuild dendrogram** and **Export Newick tree**



Предсказание структуры

Structure generation constraints:

The structure contains 6 residue(s): [Add residue](#)

α/β	D/L	Residue	Ring form
1. ?	D	galact-2N-uronic acid	pyranose
2.		acetic acid	
3.	D	show all residues	
4.		phosphoric acid	
5. α	?	any octose	pyranose
6.	L	alanine	

Allowed linkages:

	C1	C2	C3	C4	C5	C6	C7+
D-GalpNA	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	None
Ac	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D-Rib-ol	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None
P	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
α -Octp	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	None
L-Ala	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None

Advanced options: [Hide](#)

Min in	Max in	Location	Ac at N	Acceptors	Remove
1	2	any	demanded	any	<input checked="" type="checkbox"/>
?	?	any		any	<input checked="" type="checkbox"/>
?	?	terminal		3,5	<input checked="" type="checkbox"/>
?	?	reducing			<input checked="" type="checkbox"/>
?	?	any	forbidden	1	<input checked="" type="checkbox"/>

Search depth: Widespread structures only

Scope: oligomers polymers

Advanced scope: β -anomers: = 1 CH₂ carbons: ? no furanoses

Top 15 matches:

#Rank	Structure	Experimental spectrum	Simulated spectrum	Comments
#1.				$\Delta \sim 0.94$ ppm Corr = 1.000 RMS dev = 1.46 ppm Trust = 46%
#2.				$\Delta \sim 0.95$ ppm Corr = 1.000 RMS dev = 1.46 ppm Trust = 46%
#15.				$\Delta \sim 1.42$ ppm Corr = 0.999 RMS dev = 1.99 ppm Trust = 49%

Find best matching structures:

Experimental ¹³C NMR spectrum in water (24 signals of 24 expected):

17.4 22.9 34.7 50.5 52.4 63.9 64.9 66.2 68.3 70.6 72.3 72.4 72.7 73.3 73.6 76.5 78.6 78.8 99.2 102.6 171.2 175.2 176.0 176.5

\pm 2 signals

Find 15 best-fitting structures

Save generated structures

[Go!](#)

E-mail for results: [why?](#)
user@gmail.com

