Zelinsky Institute
Moscow, Russia

Philip Toukach

3D structures in Carbohydrate Structure Database

http://toukach.ru/CSDB_3D.htm

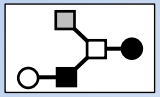
Carbohydrate Structure Database

CSDB

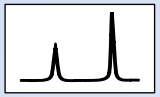
prokaryotes, fungi, plants, protista

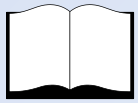
Database of natural carbohydrates

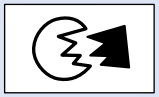
Platform for glycoinformatic services

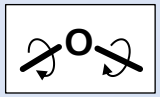

29K glycan structures


15K taxa


18K NMR spectra


11K articles


2K glycosyl-transferases


3K geometries


3K building blocks

- annual updates
- curated content
- complete coverage

(*microorganisms* – to 2021, *plants* – to 2001)

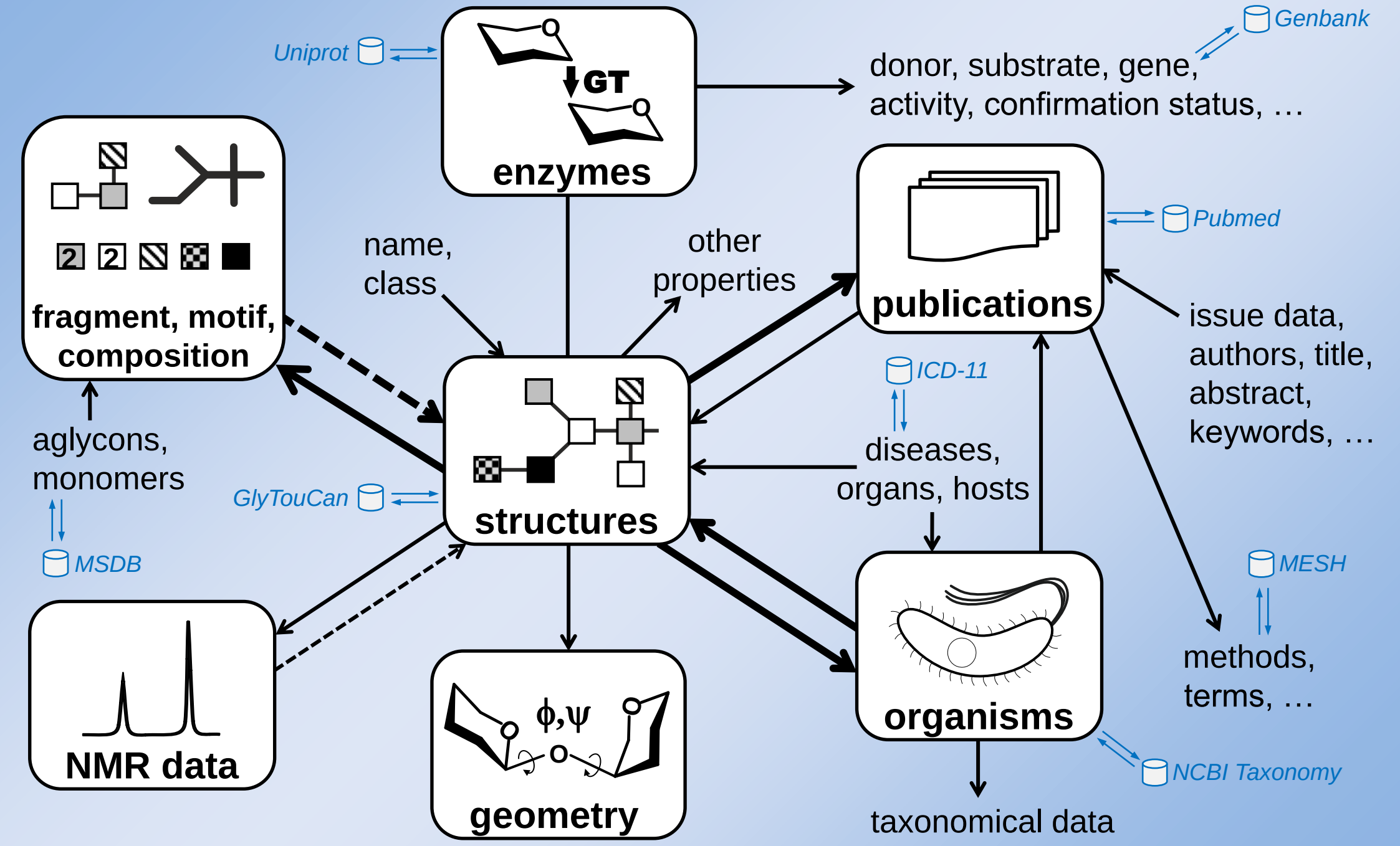
- data analysis tools
- NMR simulator / elucidator
- integration with other DBs

http://csdb.glycoscience.ru

- free access
- detailed manuals
- problem solution examples

Egorova K, Toukach Ph **CSDB_GT: a new curated database on glycosyltransferases** *Glycobiology 2017, 27:285-290*
Toukach Ph, Egorova K **Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts** *Nucl Acid Res 2016, 44:D1229-1236*
Egorova K, Toukach Ph **Expansion of coverage of Carbohydrate Structure Database (CSDB)** *Carbohydr Res 2014, 389:112-114*

Main data in CSDB



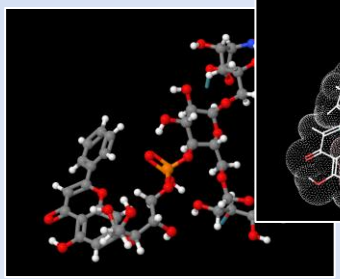
Access to 3D structures

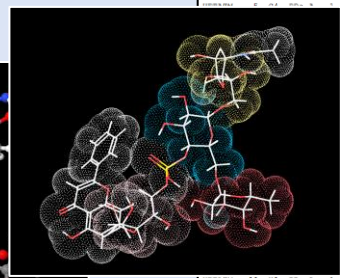
- <5% papers report 3D data
- Most DBs cover experimental / simulated 3D structures of mammalian glycans (as a part of glycoproteins)
- Stored simulations are obtained under different conditions
- Glycopolymers are totally uncovered
- Every simulation becomes a separate research
- We need an out-of-the-box tool for non-IT specialists
- → fully automated, standardized models of solution structures
- → massively pre-calculated, even if imperfect
- → exportable to atomic models

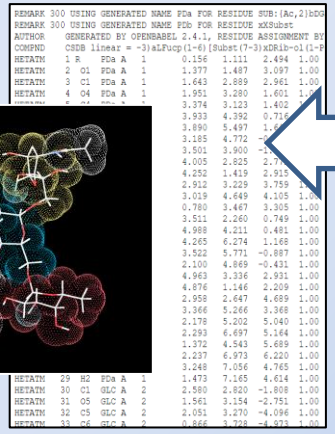
Initial geometry generator

- expert and novice modes
- all poly- & oligomeric topologies
- 600+ monosaccharides and other residues
- SMILES for atypical entities
- any bond types (incl. chelate and C-C)
- uncertainties, alternatives, superclasses, repeats

- online model explorer
- export of glyco- and all-chemical notations
- export of atomic coordinates
- render, SNFG-coloring



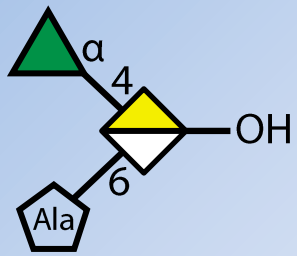




Conformation analysis

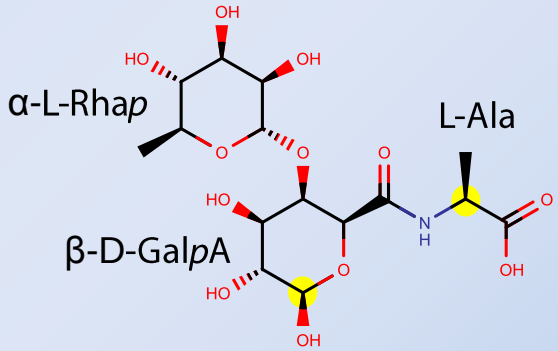
aLRhap(1-4)[x?Ala?(2-6)]?DGalpA

structure
(can be incomplete)

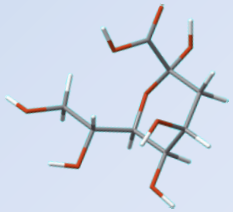


other variants
(α-GalA, D-Ala, etc.)

SMILES

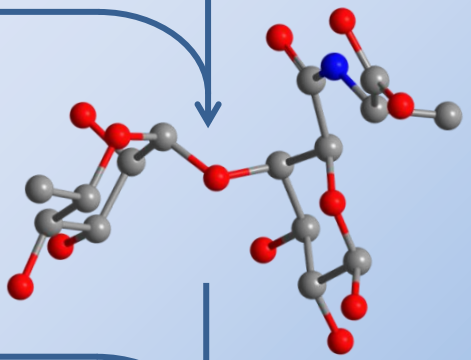


populated ring
conformers
~1000 residues

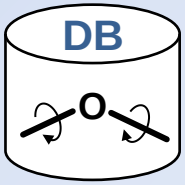


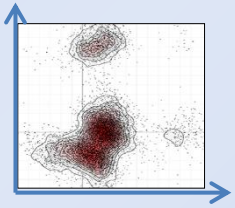
$^4C_1$,$^1C_4$, …

«chairification»



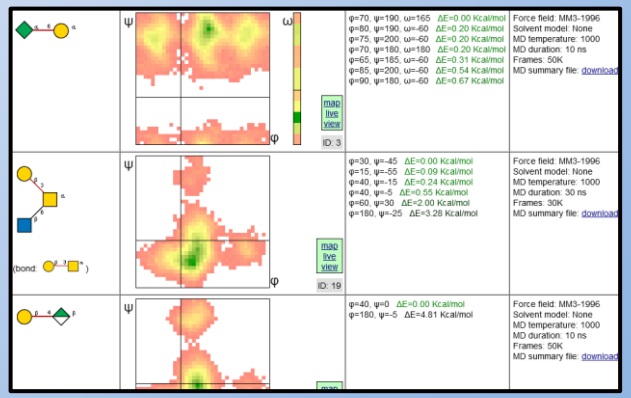
populated
bond torsions





φ,ψ,ω,θ

picking of minima
MM-relaxation



conformers
+ energies

mol. dynamics
300K, 100ns,
explicit H_2O

Conformation search

Search for conformation maps

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

Conformation ID: *Type CSDB conformation ID or range, e.g. 1-5,10,12* — ANY — Search!

Model bond: *Use selectors* β D-GlcpNAc →? α D-Manp
(only those components are listed for which conformation maps are stored)

or type dimeric fragment in CSDB encoding
Ac(1-2)bDGlcpN(1-?)aDManp

☐ Strict modification search

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

Force field: *Filter by MD method* — any

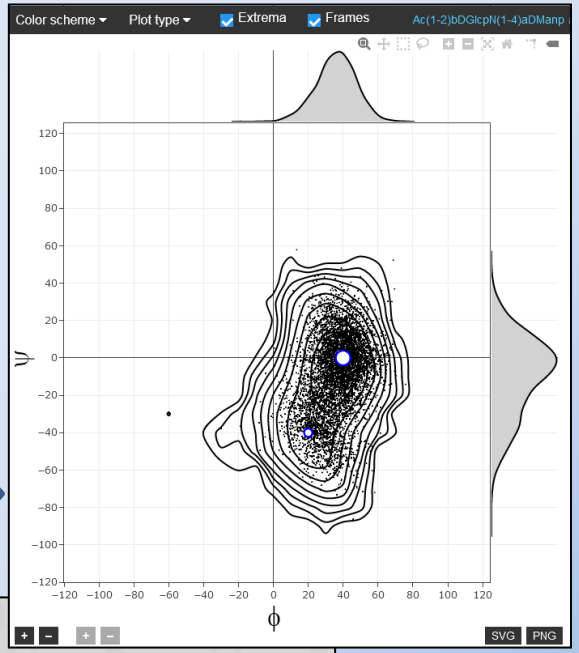
Temperature: *Filter by MD temperature* — any

Solvent model: *Filter by solvent model* — any

Home Help

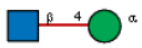
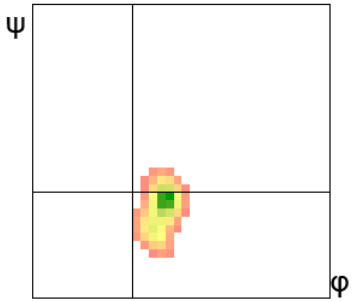
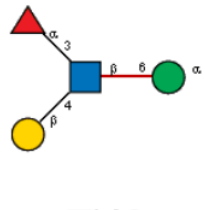
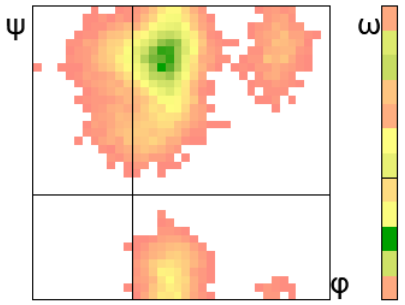
search for:
IDs,
(sub)structures,
sim parameters

explorer:
live view and export of
energy / abundance maps,
extremum detection



CSDB conformation data search

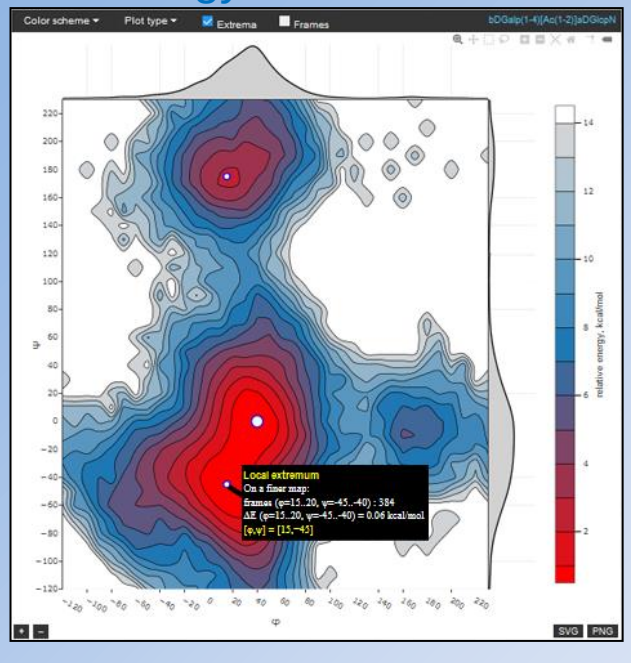
73 conformation maps have been found.

Model structure	Conformation map	Energy minima	Details
	 map live view ID: 953	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: download
 (bond: )	 map live view ID: 1857	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: download

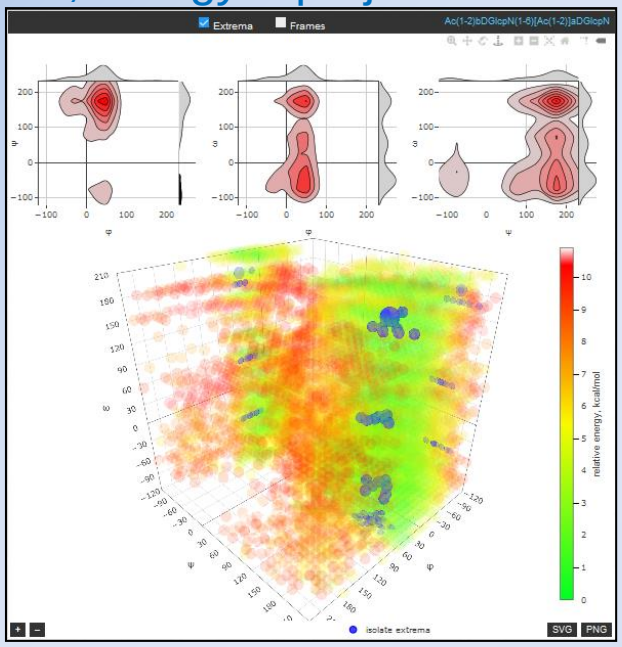
result
summary

Conformation explorer

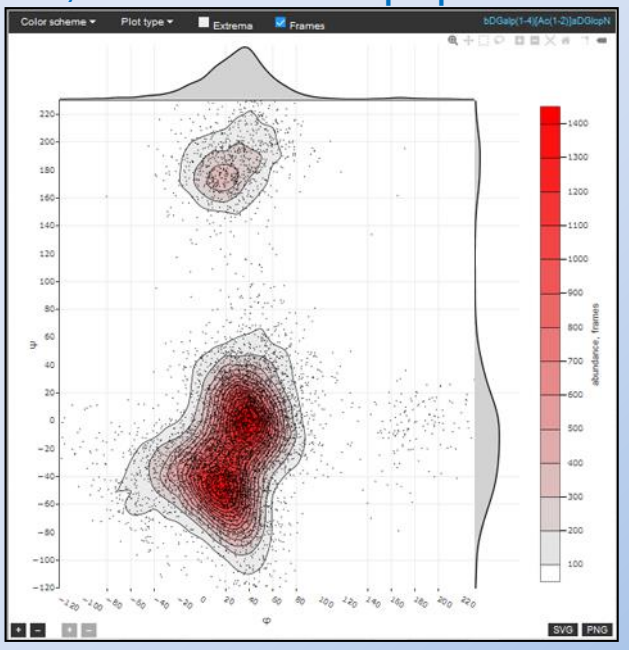
2D, energy + extrema



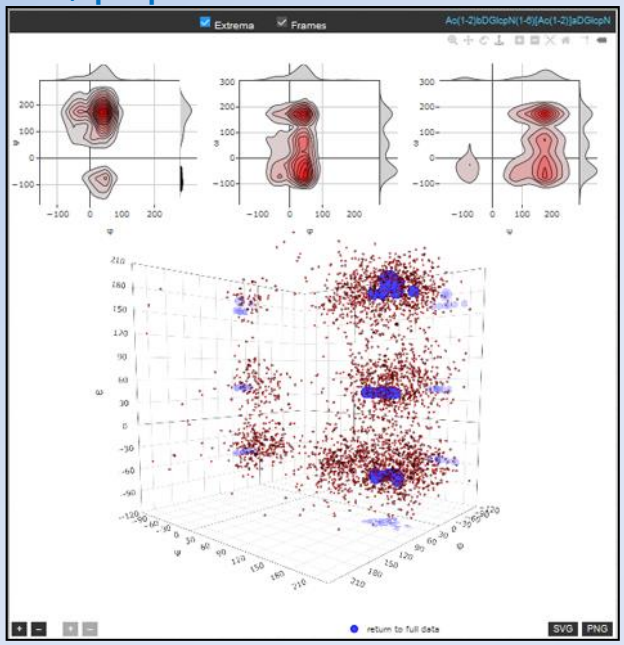
3D, energy + projections



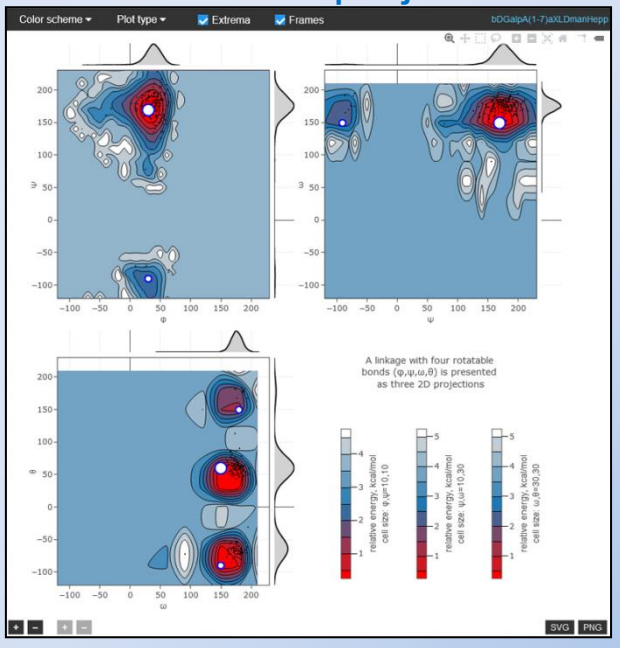
2D, abundance + population



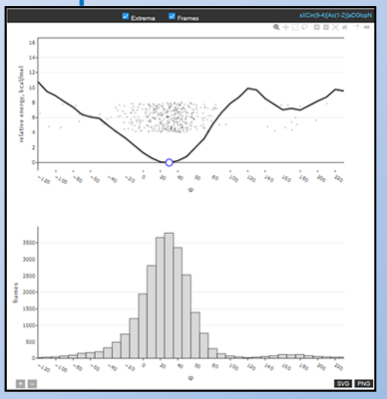
3D, population + extrema



4D: several 2D projections



1D profile



- 1..3 dihedrals per linkage
- interactive view
  (zoom, pan, orbit, density control, color schemes, switch layers)

Database filling

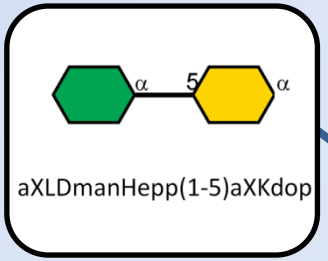
fragment abundance analysis

automated pipeline for molecular dynamics simulations

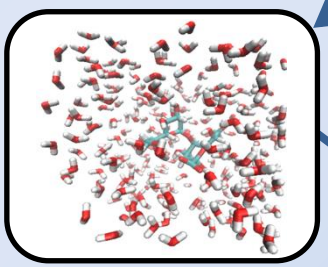
bonds with 1..4 torsions in di-, tri- and tetrameric fragments *(database filling)*

GlycoMapsDB import

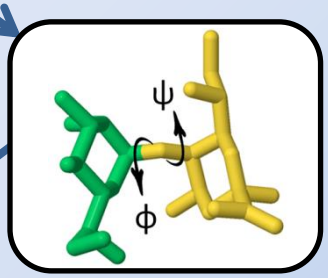
primary structure *CSDB Linear*



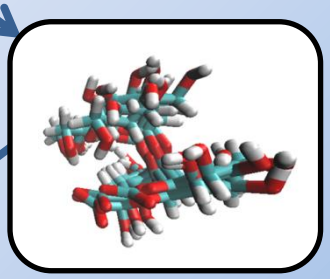
molecular dynamics, explicit H_2O *Packmol/Tinker*



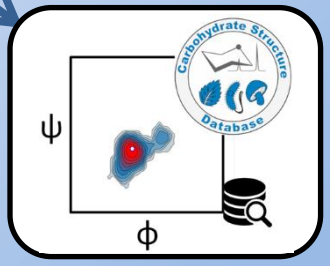
torsions, extrema *own scripts*



initial geometry *REStLESS*

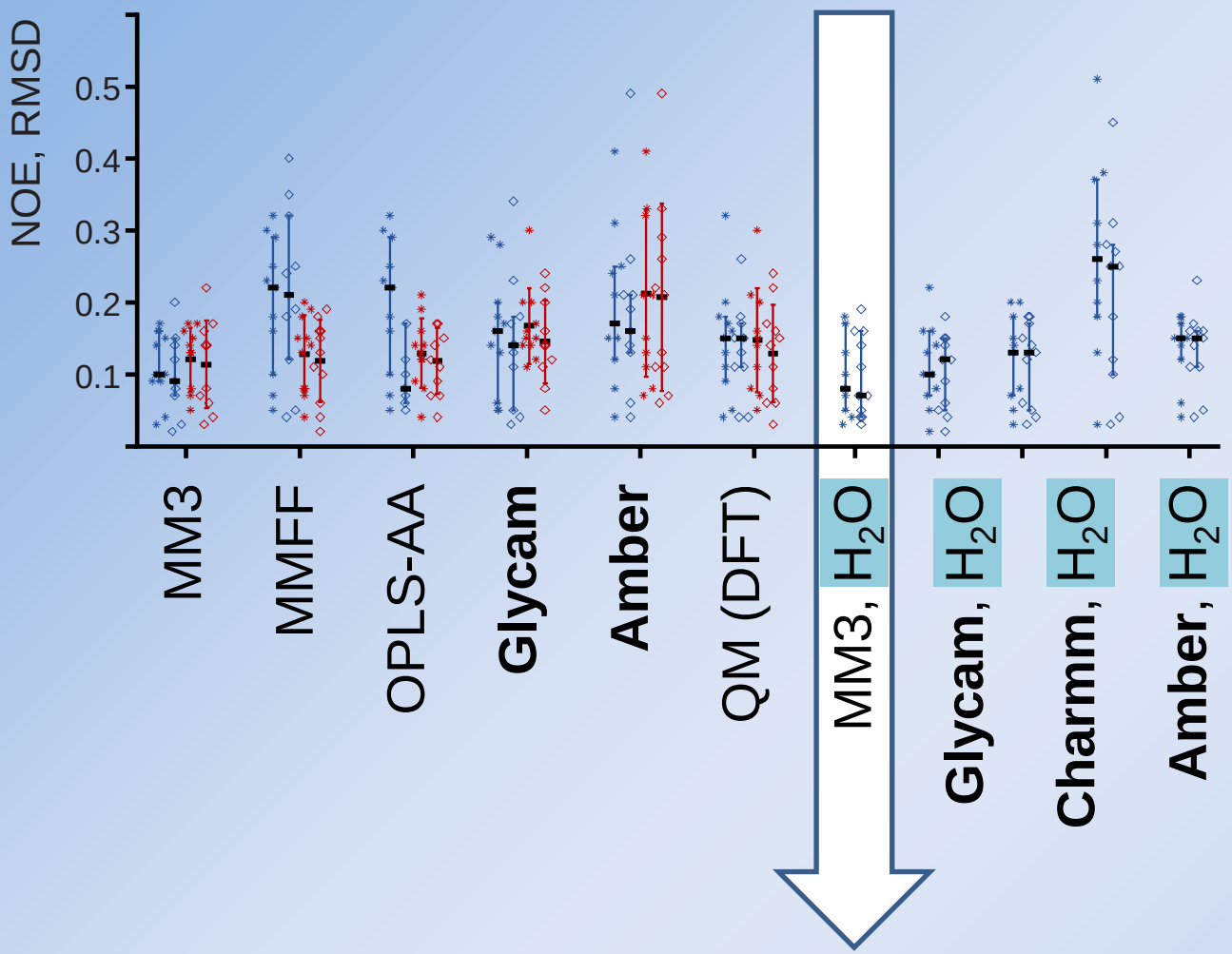


trajectory analysis *CAT*



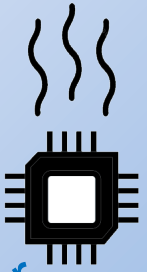
map import *CSDB importer*

Source of data



Sampling:
11 typical disaccharides

- all H-H NOEs, 300K
- transglycosidic NOEs, 300K
- all H-H NOEs, 1000K
- transglycosidic NOEs, 1000K

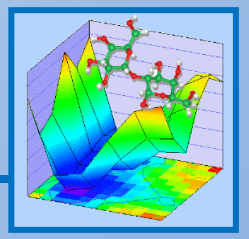


continuously
~300 models / year

~200 models: MM3, 100+ ns, 300K, explicit H_2O

~2400 models: MM3, 10-30 ns, 1000K, *in vacuo*

GlycomapsDB
proof of concept



Future perspective

- Add distance-based NOESY to 2D NMR simulator
- Fill conformation subdatabase
    - occurrence-ordered (200-300 MD simulations / year)
    - oligomers with non-carbohydrate residues
    - replacement of GlycomapsDB data
- Automatize insertion of linkage conformations into atomic models
- Conformation ensemble web-explorer / comparer / exporter

Credits

CSDB members and collaborators
involved in 3D topic

programming	Roman Kapaev, Andrei Bochkov, Ivan Chernyshov, …
literature processing & verification	Ksenia Egorova, Nadezhda Kalinchuk, Kirill Kazantsev, …
general support, data collection	Yuriy Knirel
integration, ontology	René Ranzinger, Kiyoko Aoki-Kinoshita, Thomas Lütteke, …
conformation analysis	Victor Stroylov, Sofya Scherbinina, …
ideas, R&D, notation, programming, interface, supervision	Philip Toukach

host	Zelinsky Institute
partners	GlycomeDB, JCGGDB, UniCarbKB, GLYCOSCIENCES.DE
funding	RSF, RFBR, dkfz., ISTC