



# RosettaCarbohydrates: Glycan Modeling in Rosetta

In Collaboration with Jason  
Labonte, Ph.D (Gray Lab)

Jared Adolf-Bryfogle,  
PhD  
Schief Lab

# Glycoproteins and glycans are important!

## Immune evasion

- Pathogens “hide” from the immune system using glycan shields

## Antibody binding

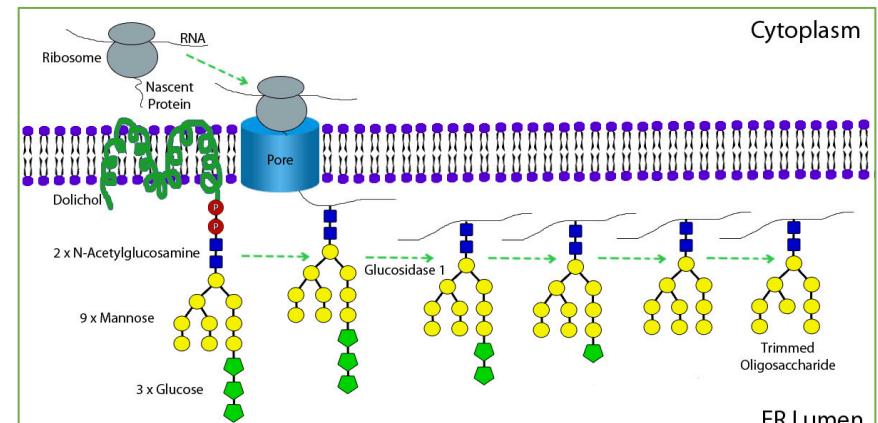
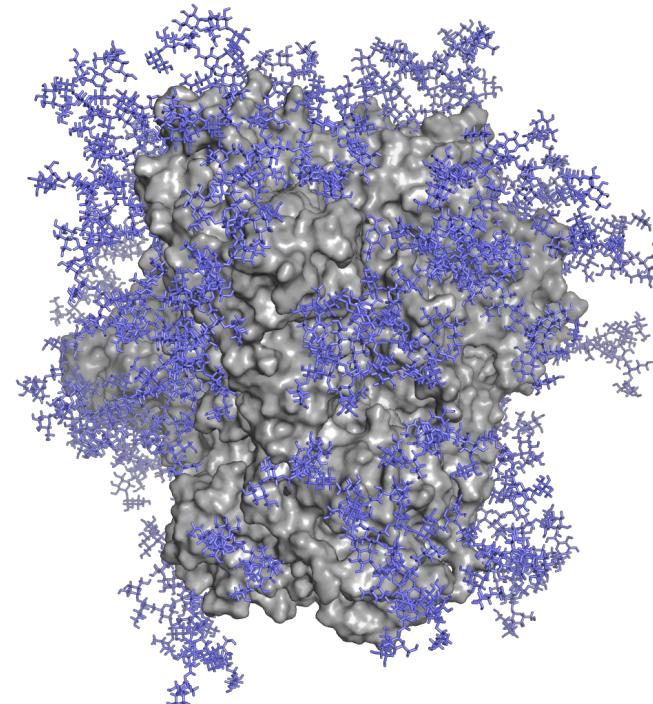
- Antibodies engage glycans for target binding
- Carbohydrate vaccines

## Protein-carbohydrate interactions

- Gene regulation
- Metabolism
- Cell-cell communication

## Solubility/Folding

- Glycosylation often improves solubility
- Can increase stability



<http://www.writeopinions.com/n-glycosylation>

## Challenges in Modeling Sugars

---

- **The nomenclature problem**
- **The scoring problem**
  - Force field parameters for sugars are not as well-developed as for proteins.
  - Sugars have several “odd” electronic effects, (e.g., the anomeric effect).
- **The sampling problem**
  - Sugars have far more **degrees of freedom (DoFs)** than proteins.

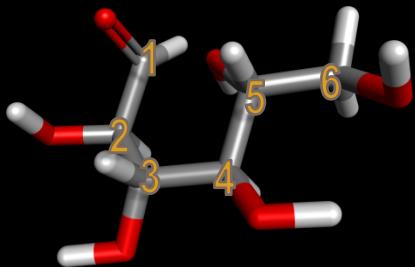
# Carbohydrate Nomenclature

---

Slides from Jason Labonte: Carbohydrates 101

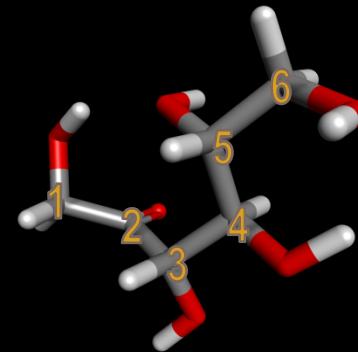
# NOMENCLATURE/PROPERTIES

Aldose: linear form is an aldehyde



glucose

Ketose: linear form is a ketone



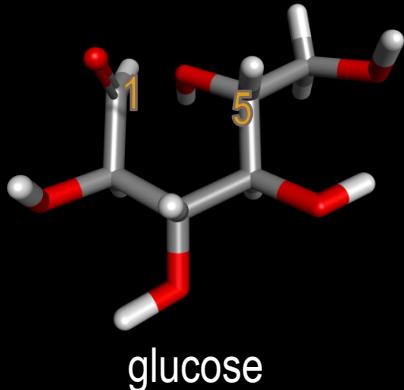
fructose

All monosaccharides contain at least 3 carbons (trioses).

Most monosaccharides contain 6 carbons (hexoses) or fewer.

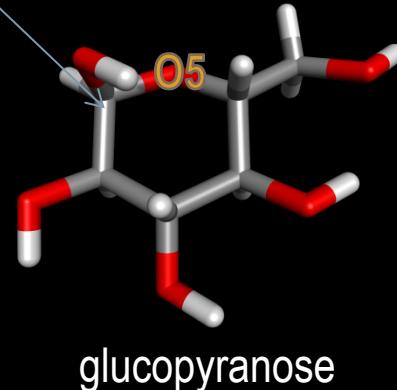
# NOMENCLATURE/PROPERTIES

Acyclic



Cyclic

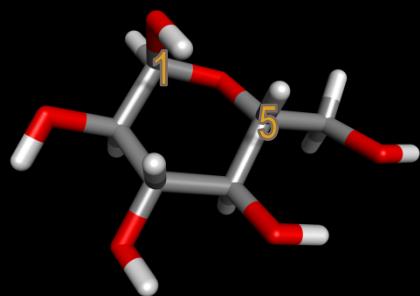
anomeric carbon



Most sugars can form rings, (trioses can't,) and most exist primarily in cyclic forms.

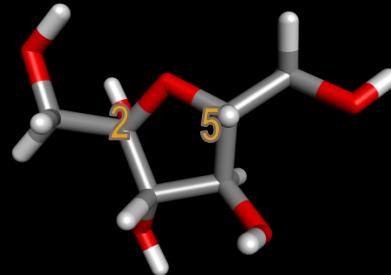
# NOMENCLATURE/PROPERTIES

Pyranose: 6-membered ring (like pyran)



glucopyranose

Furanose: 5-membered ring (like furan)



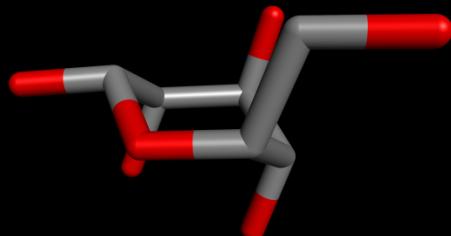
fructofuranose

Pentoses and hexoses can exist as both pyranoses and furanoses.

Long sugars can also form septuloses, but these are less stable.

# NOMENCLATURE/PROPERTIES

Chair (2 distinct forms)



Boat (6 distinct forms)

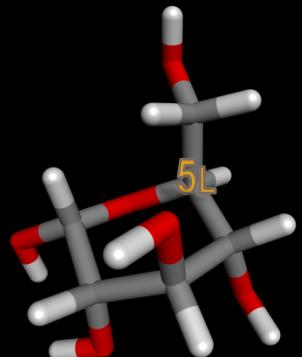


Chairs are energy minima.

Boats are energy maxima; however, the presence of oxygen in the ring helps reduce steric clash.

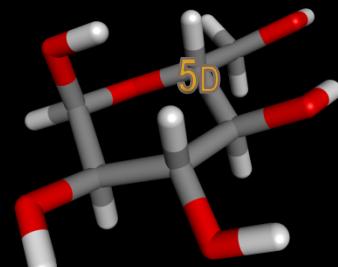
# NOMENCLATURE/PROPERTIES

L-Sugar: C<sub>n-1</sub> shares the same relative stereochemistry as L-glyceraldehyde



L-glucopyranose

D-Sugar: C<sub>n-1</sub> shares the same relative stereochemistry as D-glyceraldehyde



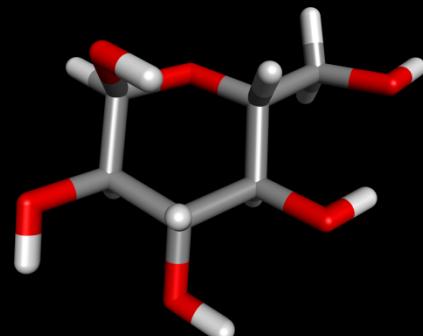
D-glucopyranose

Note that every stereocenter has flipped.

Most natural monosaccharides are D-sugars.

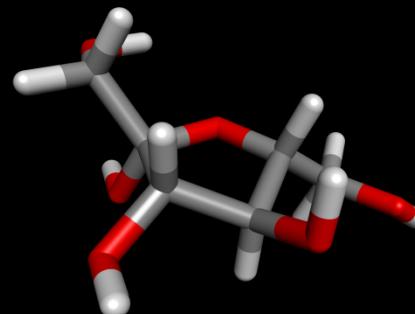
# NOMENCLATURE/PROPERTIES

$\alpha$ -Sugar: anomeric sidechain trans to sidechain at  $C_{n-1}$



$\alpha$ -D-glucopyranose

$\beta$ -Sugar: anomeric sidechain cis to sidechain at  $C_{n-1}$

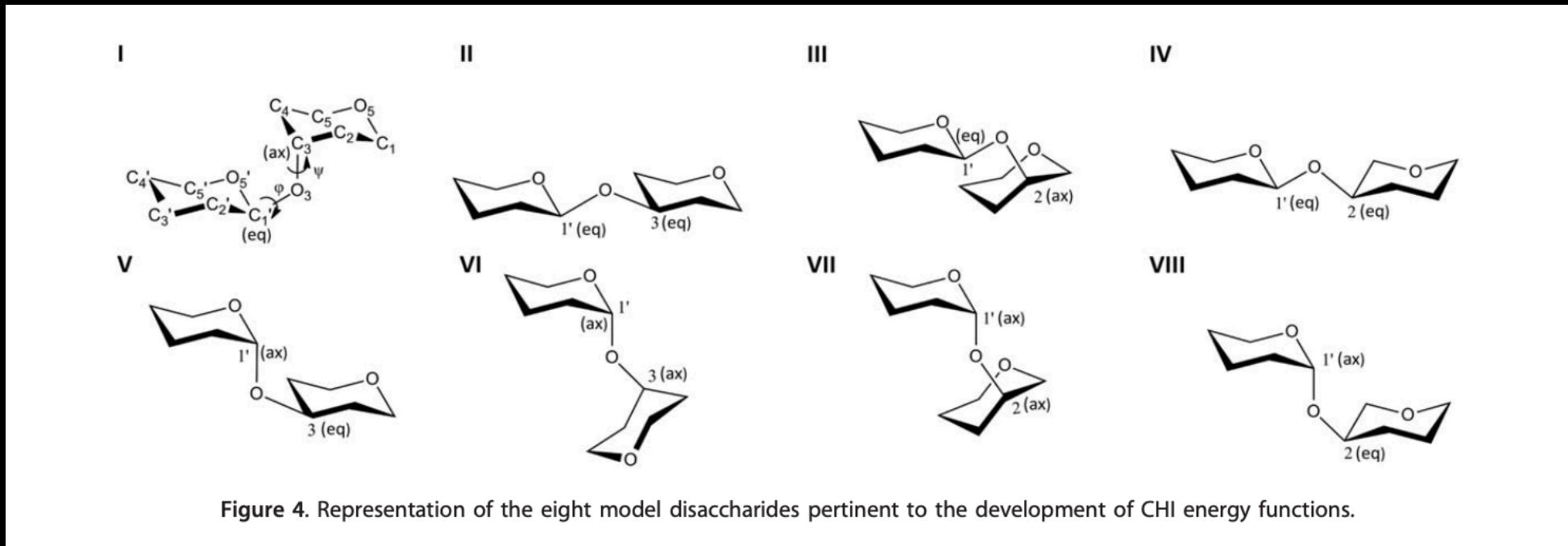


$\beta$ -D-fructofuranose

Note that changing the stereochemistry of any single non-anomeric carbon yields another sugar, an epimer.

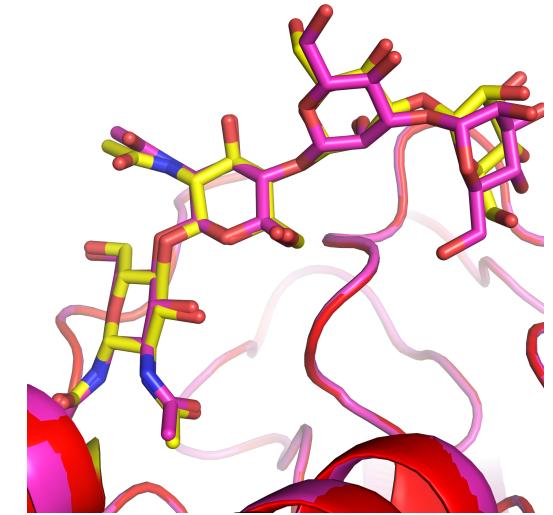
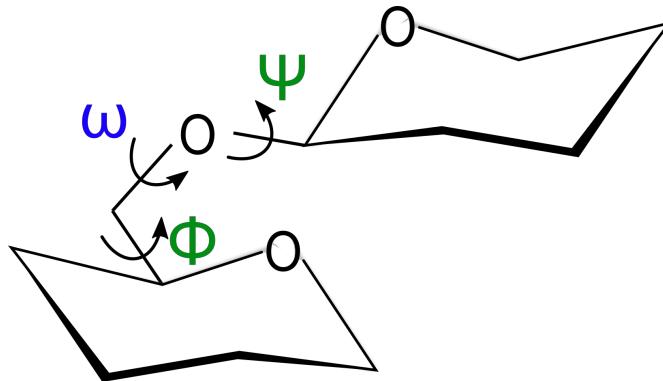
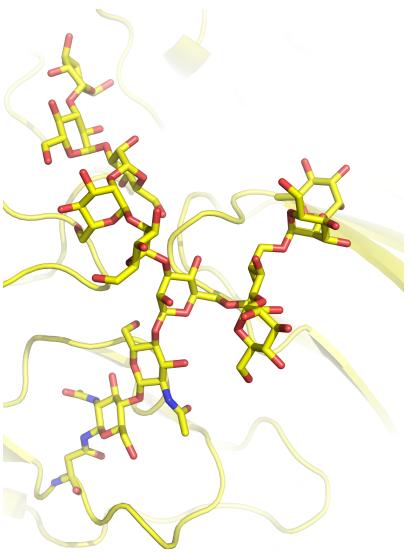
The anomeric effect causes  $\alpha$ -D-glucopyranose to be more stable than  $\beta$ -D-glucopyranose, even though the anomeric hydroxyl group is axial.

# AXIAL VS EQUITORIAL GLYCAN LINKAGE



# A framework to represent glycans in Rosetta

- Numerous types of Sugar Residues and modifications
- Typically 2-3 dihedrals between residues ('backbone')
- Structure of tree depends on C-C connection (1-4, 1-6, etc)
  - Can be branching



Journal of  
**COMPUTATIONAL  
CHEMISTRY**

WWW.C-CHEM.ORG

FULL PAPER

## Residue-Centric Modeling and Design of Saccharide and Glycoconjugate Structures

Jason W. Labonte,<sup>[a]</sup> Jared Adolf-Bryfogle,<sup>[b]</sup> William R. Schief,<sup>[b,c]</sup> and Jeffrey J. Gray\*,<sup>[a]</sup>

## Challenges in Modeling Sugars

---

- The nomenclature problem
- **The scoring problem**
  - Force field parameters for sugars are not as well-developed as for proteins.
  - Sugars have several “odd” electronic effects, (e.g., the anomeric effect).
- The sampling problem
  - Sugars have far more degrees of freedom (DoFs) than proteins.

# Scoring carbohydrates

- CHI (CarboHydrate Intrinsic) - energy function
  - Derived from QM
  - sugar\_bb ScoreTerm
- Specific for types of Linkage

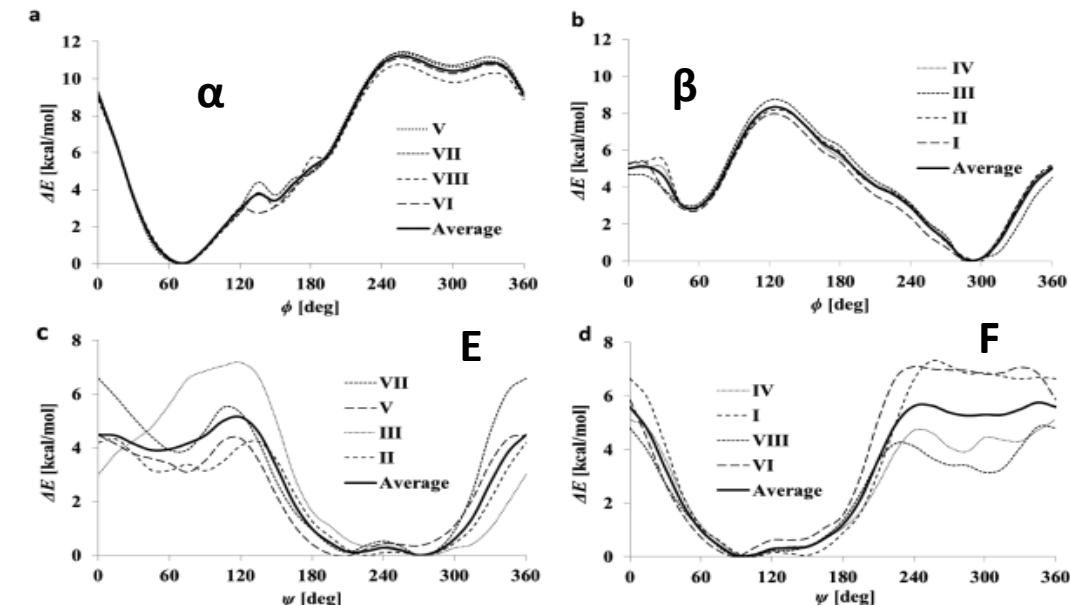
Phi	$\alpha/\beta$ linkages
Psi	1-2ax, 1-4ax and 1-3eq linkages (E) and 1-2eq, 1-4eq and 1-3ax linkages (F)
Omega	Axial/Equitorial (statistically derived)

A.K. Nivedha et al. J. Comput. Chem. 2014, 35, 526-39

A.K. Nivedha et al. JCTC 2016, 12, 892-901.

## Residue-Centric Modeling and Design of Saccharide and Glycoconjugate Structures

Jason W. Labonte,<sup>[a]</sup> Jared Adolf-Bryfogle,<sup>[b]</sup> William R. Schief,<sup>[b,c]</sup> and Jeffrey J. Gray\*<sup>[a]</sup>



# Challenges in Modeling Sugars

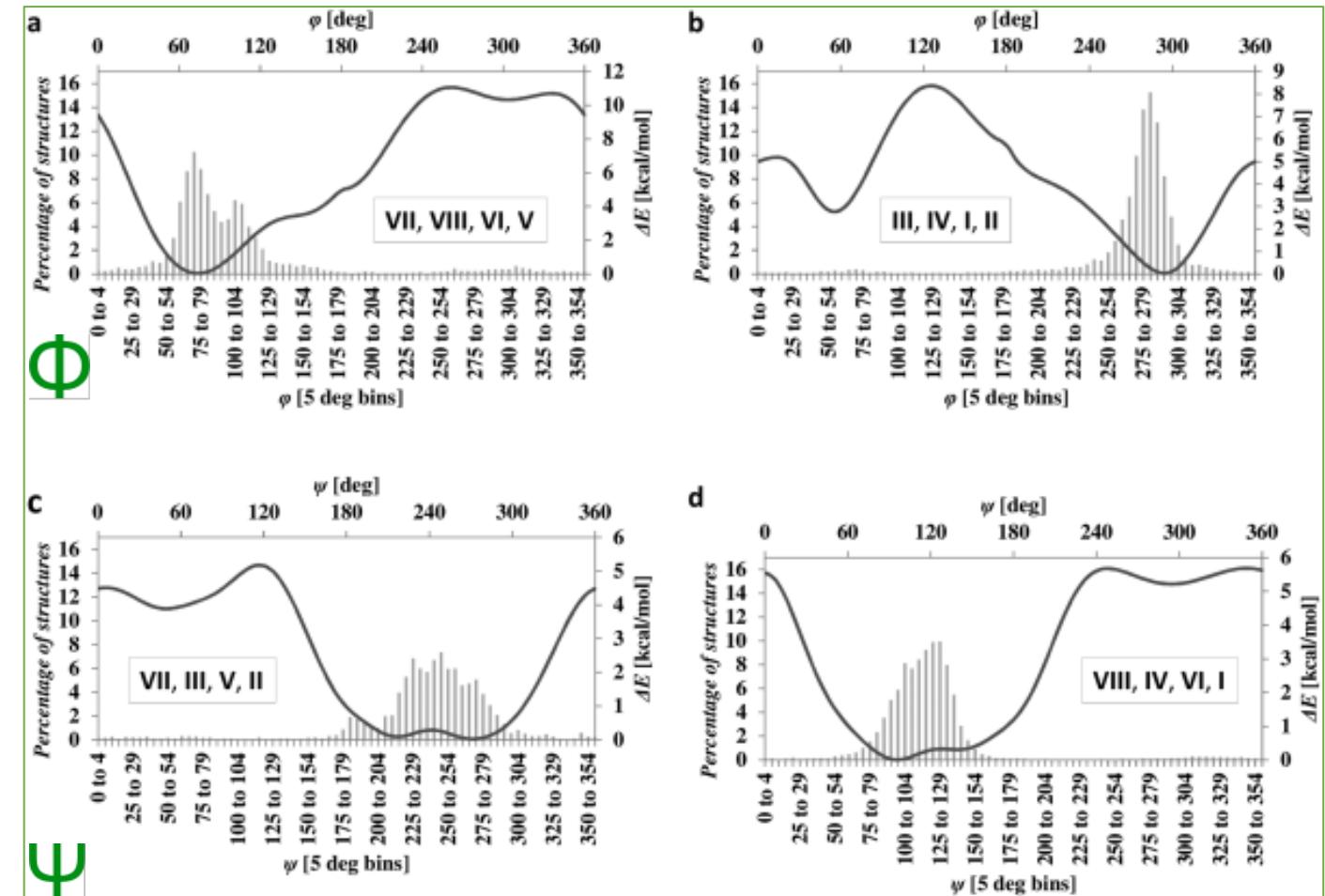
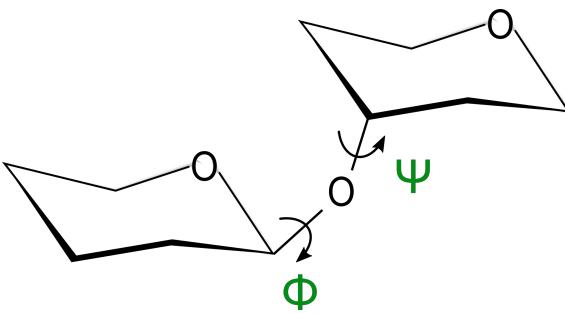
---

- The nomenclature problem
- The scoring problem
  - Force field parameters for sugars are not as well-developed as for proteins.
  - Sugars have several “odd” electronic effects, (e.g., the anomeric effect).
- **The sampling problem**
  - Sugars have far more degrees of freedom (DoFs) than proteins.

# Conformational Sampling: Phi/Psi

- CHI energy function converted into phi/psi/omega probabilities
- Implemented with new  
BB Sampling framework

## SugarBBSampler

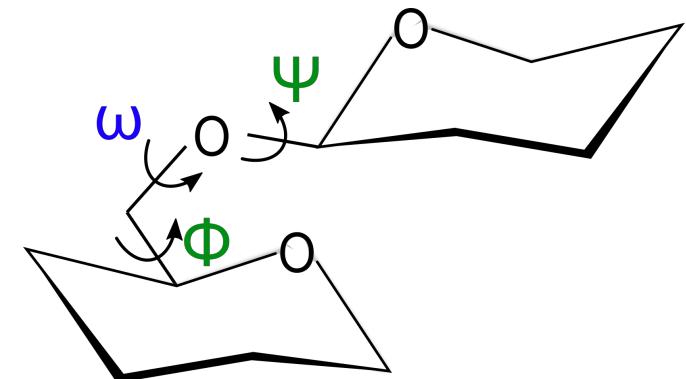


## LinkageConformerMover

### Residue-Centric Modeling and Design of Saccharide and Glycoconjugate Structures

Jason W. Labonte,<sup>[a]</sup> Jared Adolf-Bryfogle,<sup>[b]</sup> William R. Schief,<sup>[b,c]</sup> and Jeffrey J. Gray\*<sup>[a]</sup>

- Sets all torsions of a linkage at once
- Specific for types of sugar-sugar linkage
- **Sampling:** Phi/psi/omega mean (+/- gaussian of SD) of glycosidic torsions at same time



Petrescu, AJ; Petrescu, SM; Dwek, RA; & Wormald, MR. (1999), *Glycobiology*

Petrescu, AJ; Milac, A-L; Petrescu, SM; Dwek, RA; & Wormald, MR. (2004), *Glycobiology*

# Improved sampling with much more conformers

Updated 1999/2004 data using better methodology and more structures.

- Collaborate with *Maxim Shapovalov* and *Roland Dunbrack*
- Data provided by *Thomas Lutteke* ([glycosciences.de](http://glycosciences.de))

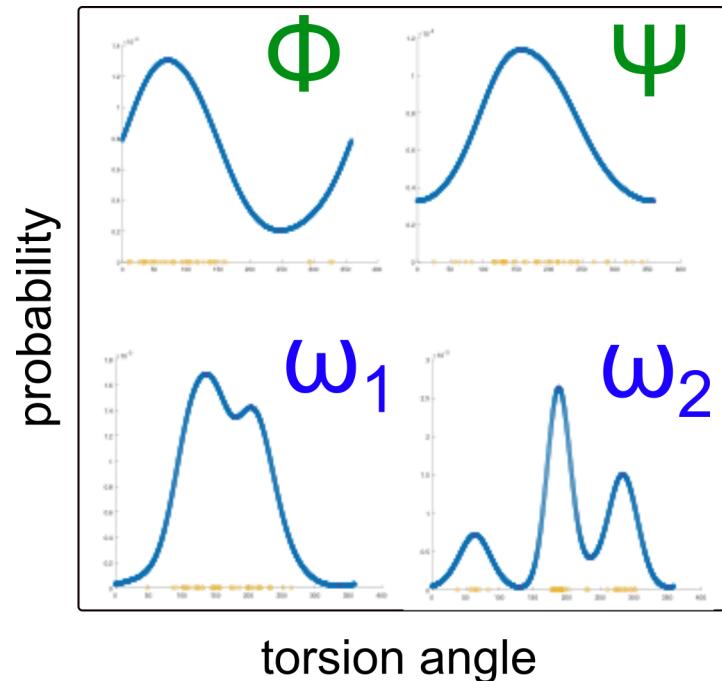
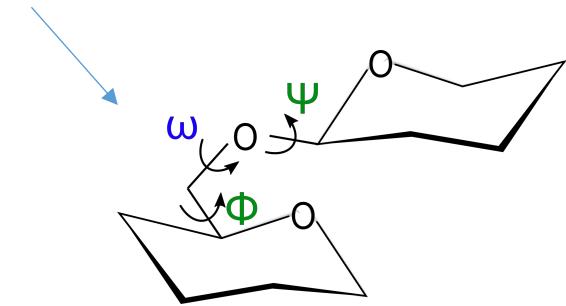
**Challenge:** Unknown Torsional bins for each torsion type (g+/trans/etc)

- 1) Generate **Adaptive Kernel Densities** using a **Von Misses Kernel** and *lowN smoothing* on filtered data
- 2) Generate **cubic splines** on the density
- 3) Calculate interdependent conformers by assigning bins to each torsion using derivatives

## Conformers

14,000 High-quality, filtered data points at  $\leq 2.0\text{\AA}$  resolution:

- 65 unique torsion types (previously 13)
- $\sim 150$  conformer (previously 27)



# Glycan Trees

pose.glycan\_tree\_set()

1 new member:

GlycanTreeSetOP\_

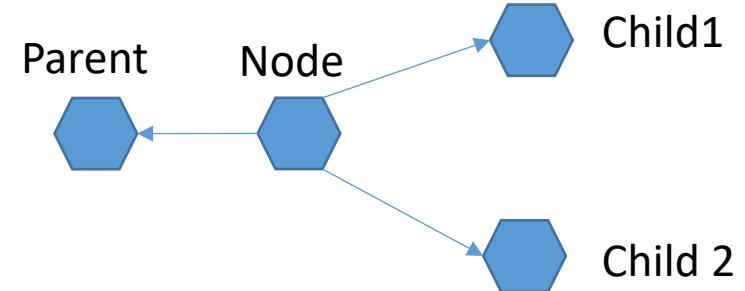
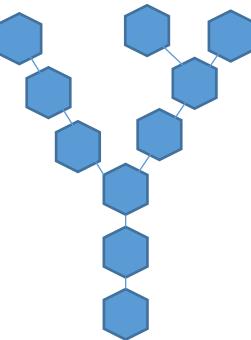
pose/carbohydrates

3 new classes:

GlycanTreeSet

GlycanTree

GlycanNode



## GlycanTree container

Map

1<sup>st</sup> glycan residue    GlycanTree object

## Node container

Map

residue number    GlycanNode object

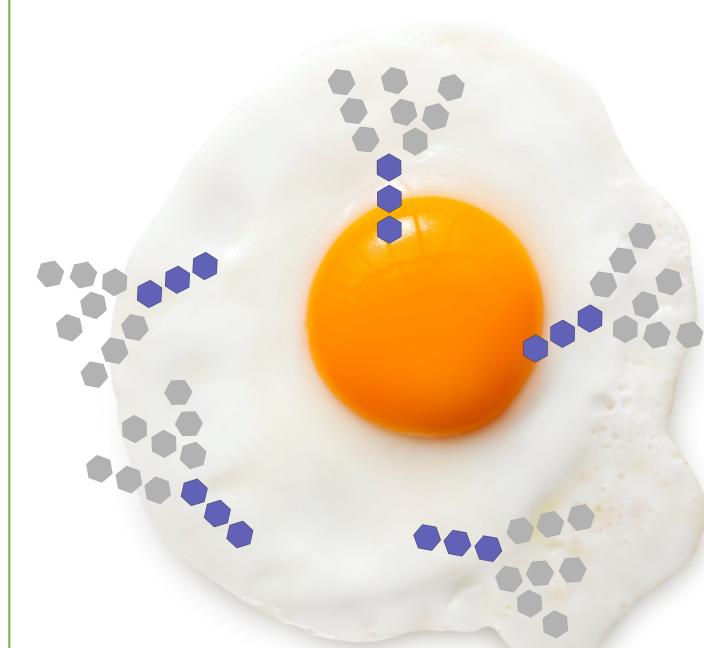
## Node

- Residue number
- Parent
- Children
- Connection numbers to all children
- Nr. of torsions for a connection

## Graph structure enabled easy implementation of ResidueSelectors

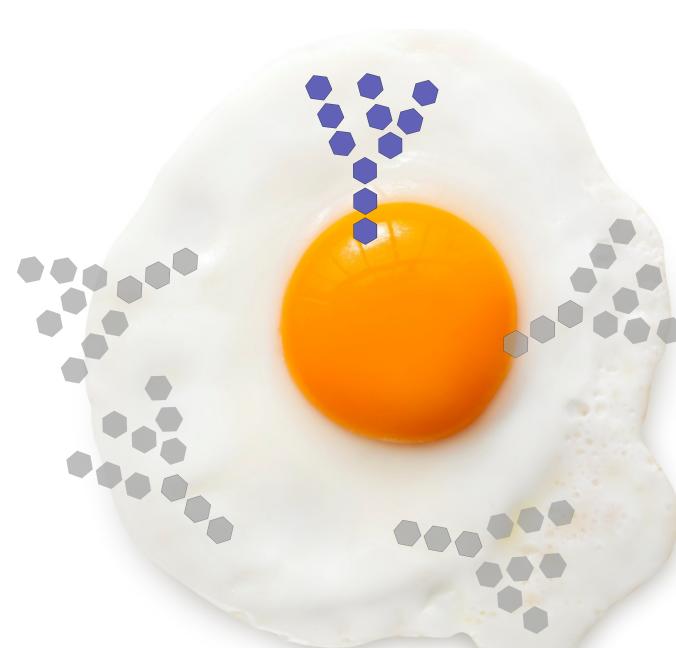
Glycan specific **ResidueSelectors** use the graph to select groups of glycan residues

GlycanLayerSelector



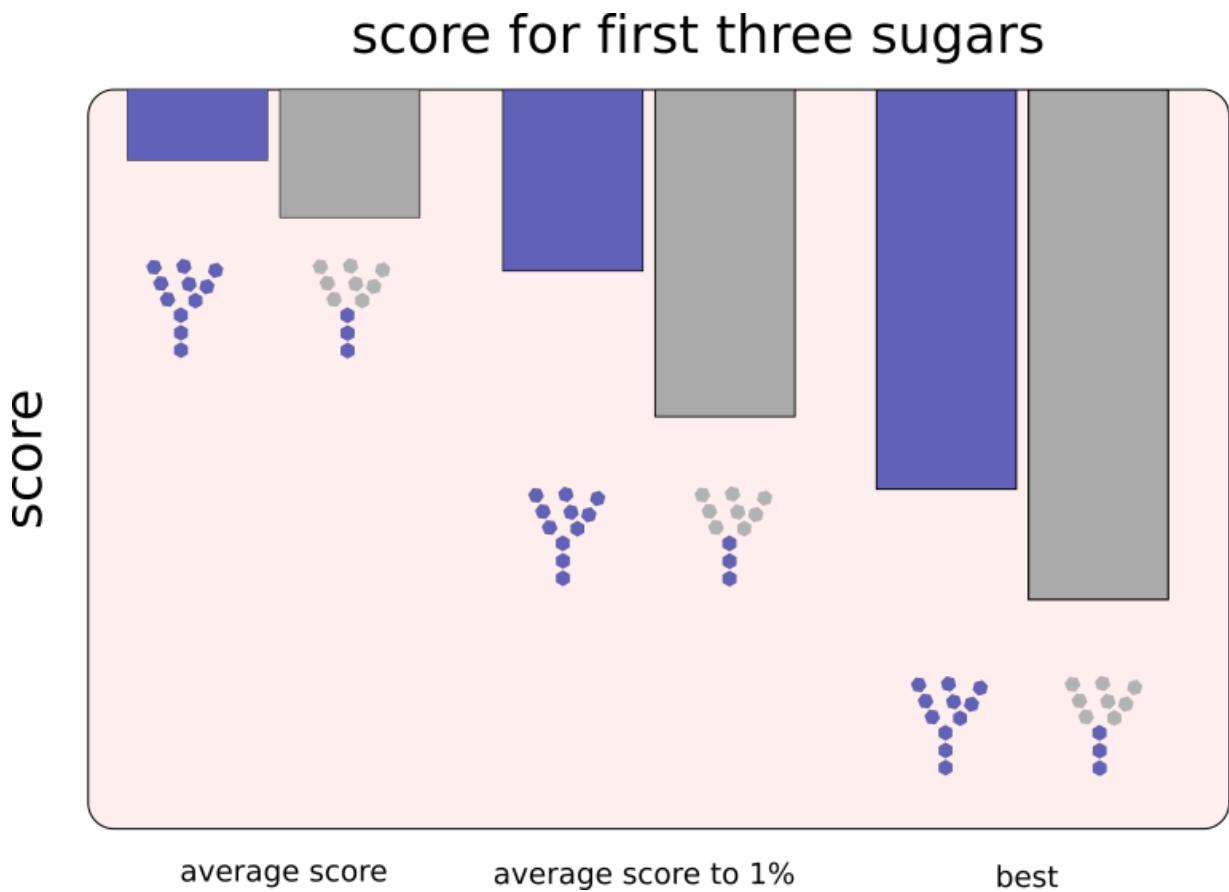
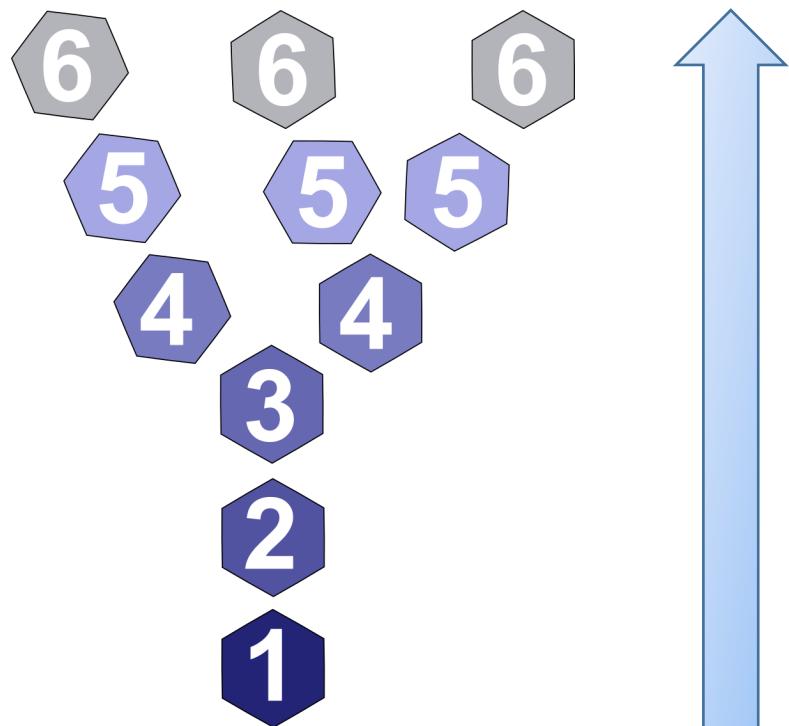
Glyco-protein

GlycanResidueSelector



Glyco-protein

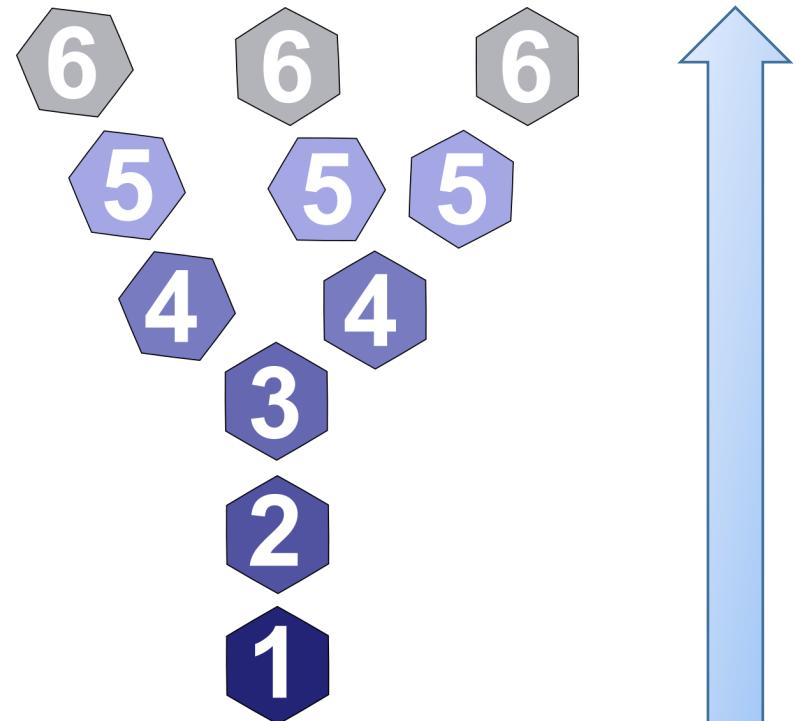
# Layer-based glycan modeling improves energies



## GlycanTreeModeler Details

---

- Turn all residues Virtual
- Build glycan(s) out in defined layers
- Use MC **GlycanSampler** to sample DOFs
  - LinkageConformerMover
  - SugarBBSampler
  - GlycanTreeMinMover
  - PackRotamersMover
    - Layer + neighbor protein residues
  - SmallMover
    - +/- 15, 30, 45 degrees at decreasing probabilities



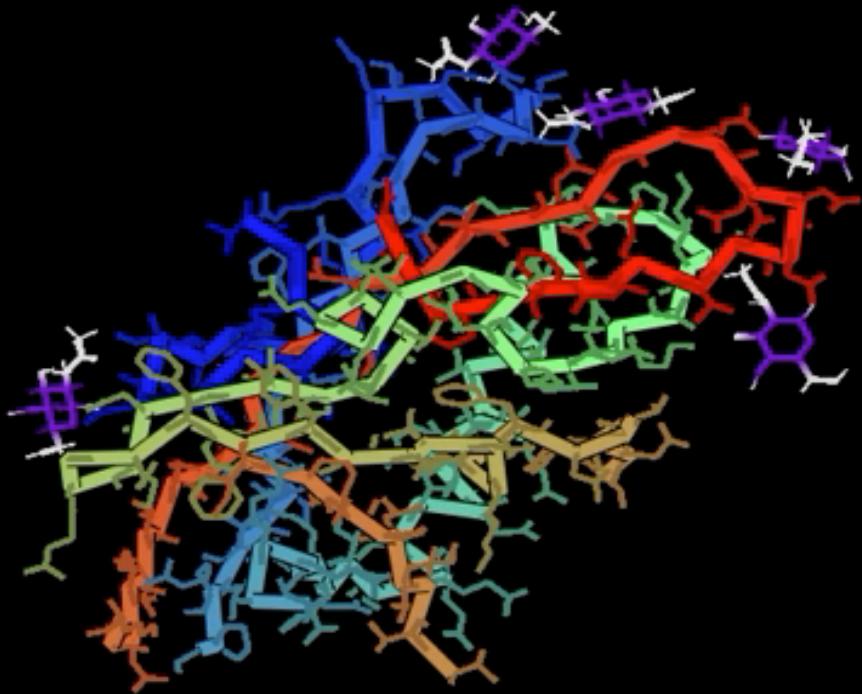
## *Layer-based glycan modeling: Example*

```
<MOVERS>
<SimpleGlycosylateMover
    name="glycosylate"
    positions="133G,137G"
    glycosylation="man9"
/>

<GlycanTreeModeler
    name="model_glycans"
    window="0"
    layer="1"
/>
</MOVERS>

<PROTOCOLS>
    <Add mover_name=glycosylate />
    <Add mover_name=model_glycans />
</PROTOCOLS>
```

Models from root out, making rest of glycan residues *virtual* until all are *real*.



## *Layer-based glycan modeling: Typical*

---

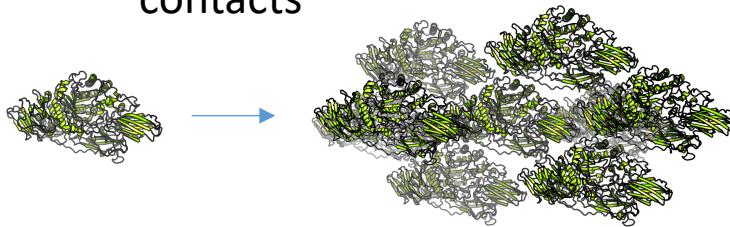
```
<MOVERS>
<SimpleGlycosylateMover
    name="glycosylate"
    positions="133G,137G"
    glycosylation="man9"
/>

<GlycanTreeModeler
    name="model_glycans"
    window="1"
    layer="2"
/>
</MOVERS>

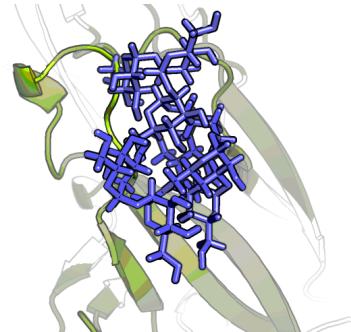
<PROTOCOLS>
    <Add mover_name=glycosylate />
    <Add mover_name=model_glycans />
</PROTOCOLS>
```

## Preliminary benchmarking and parameter optimization

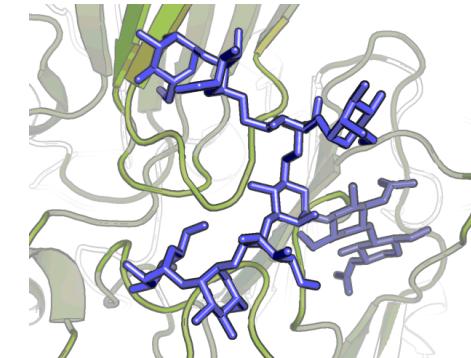
1. Re-generate crystal contacts



2. Randomize torsion angles

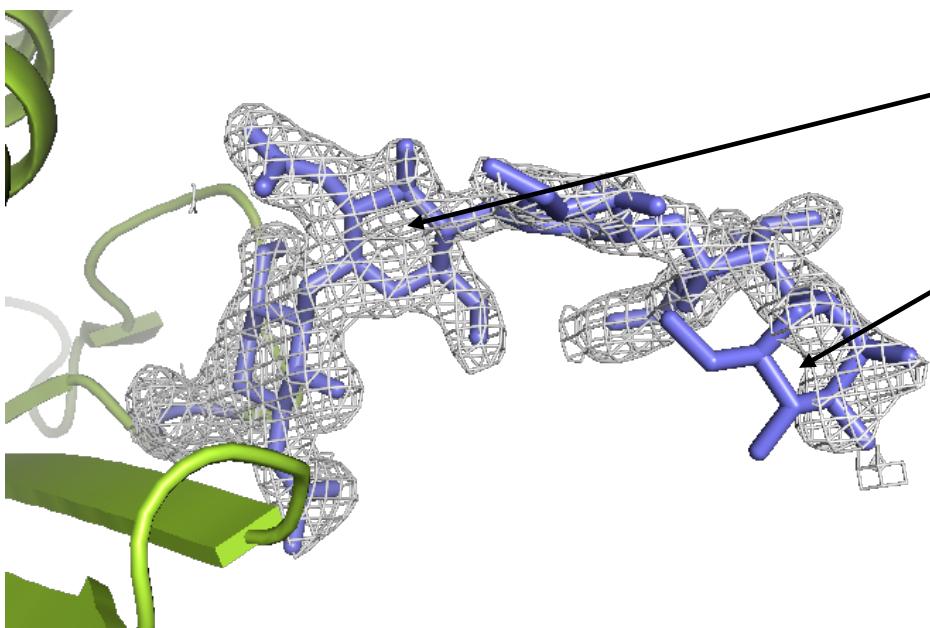


3. Model glycans



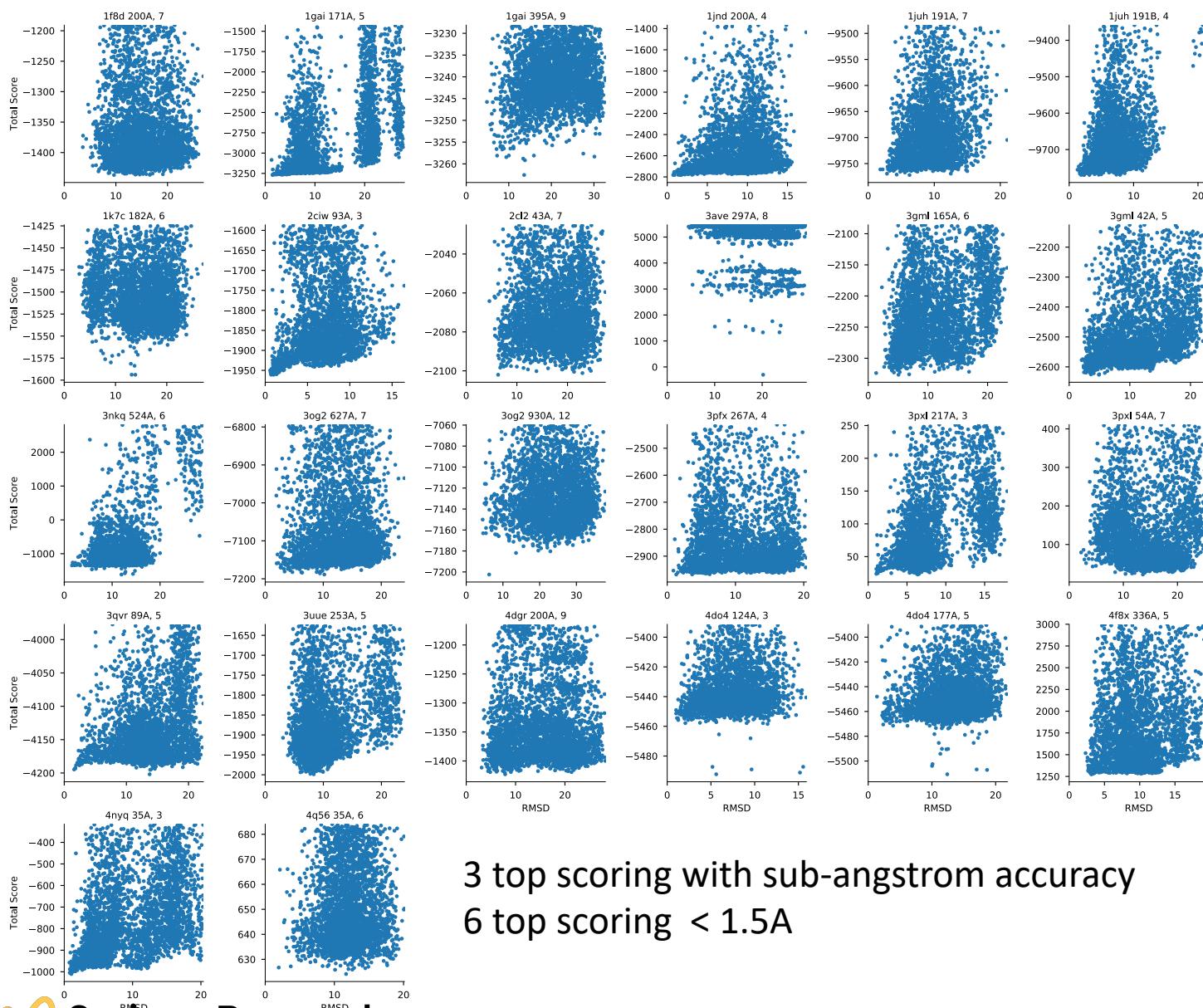
Density fit > 0.8

Density fit < 0.8

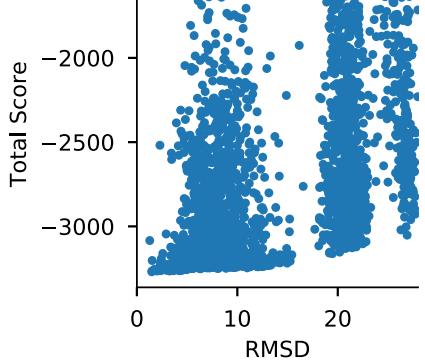
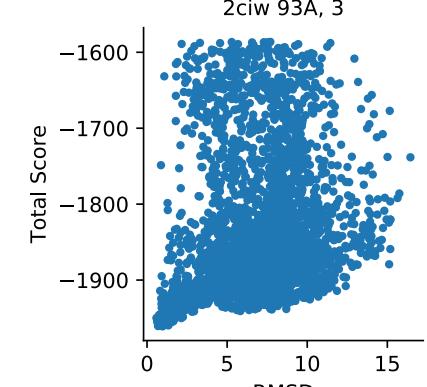


- nstruct = 4000
- rounds = 60
- 5-10 min / decoy
- Resolution: < 2A
- 26 Glycan “Trees”
- 3-12 residues

# Some glycan trees show “Folding” funnel

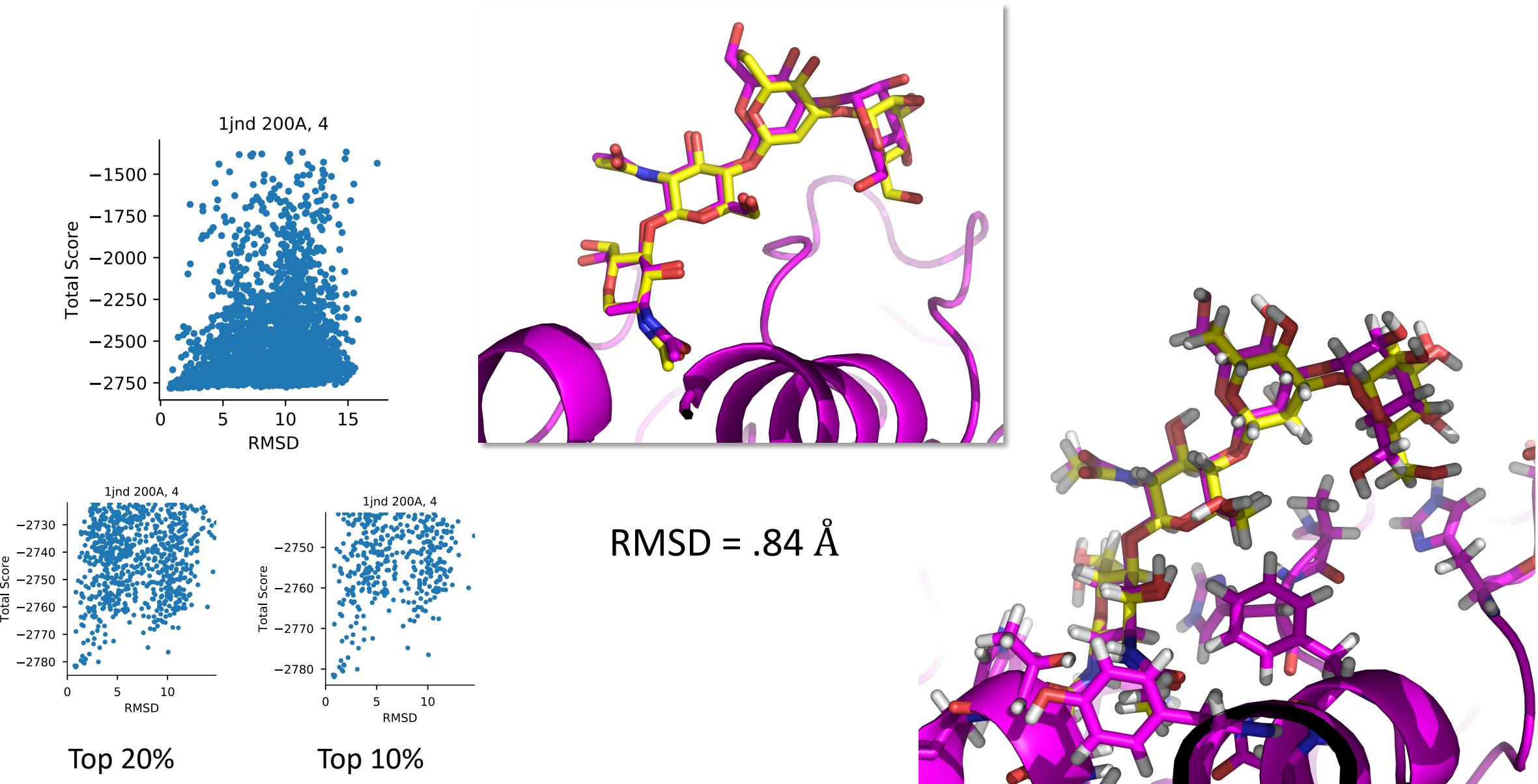


Notable examples

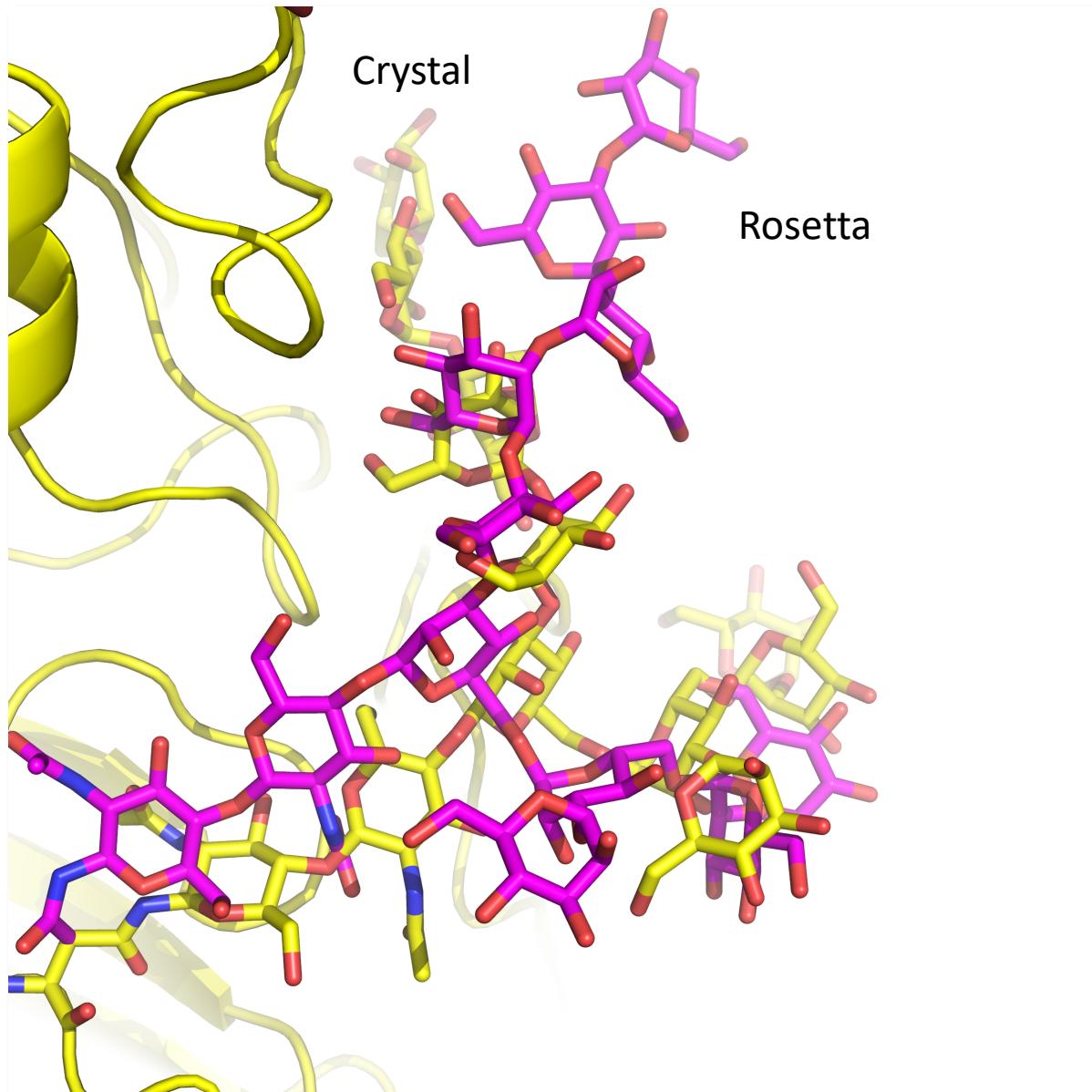
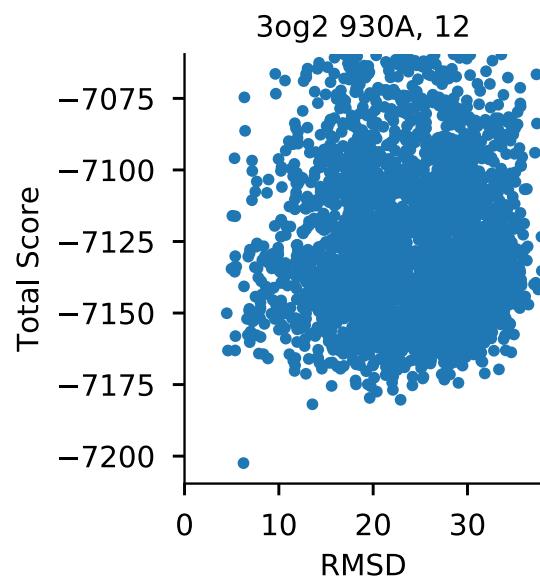


3 top scoring with sub-angstrom accuracy  
6 top scoring < 1.5A

# Best Result in Current Benchmark

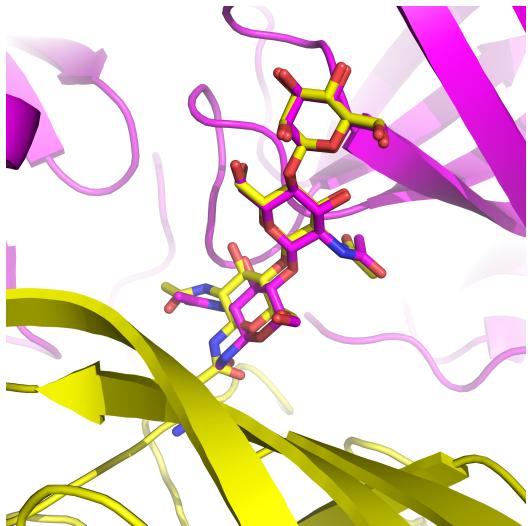


# Modeling 12 residues



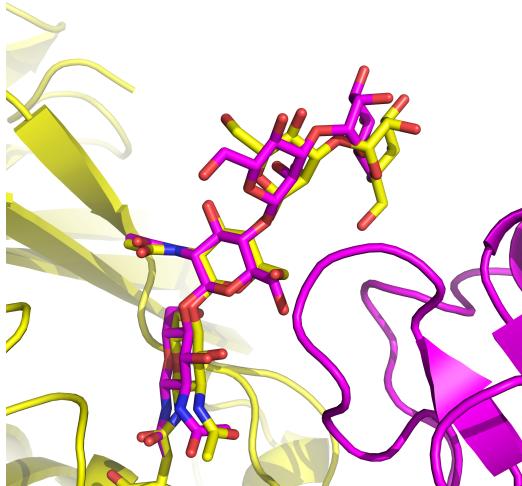
## Some more examples ...

4nyq



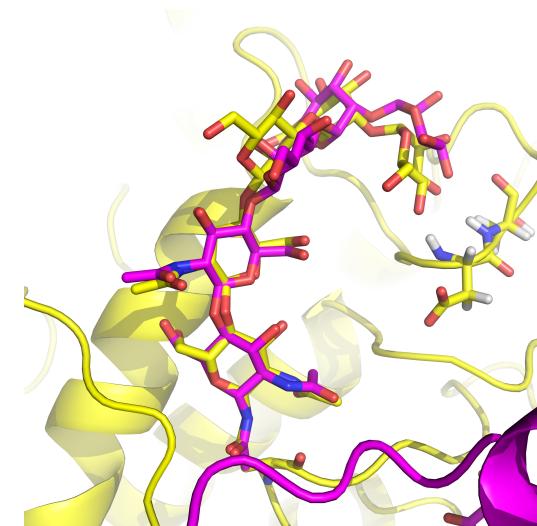
1.03 Å

3pxf



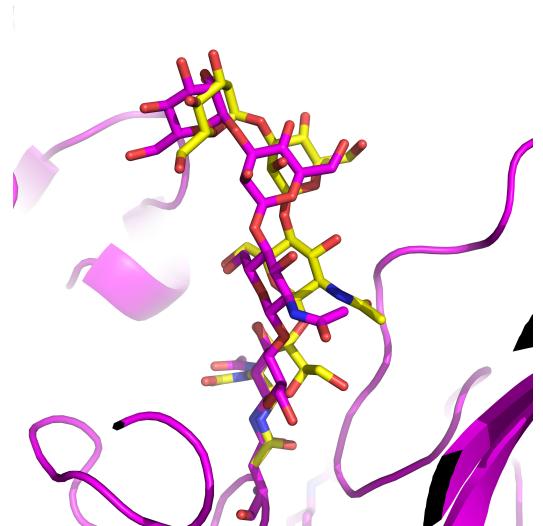
1.33 Å

1gai

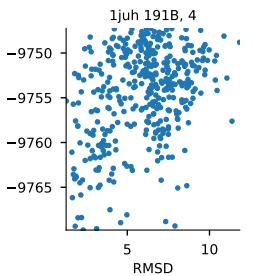
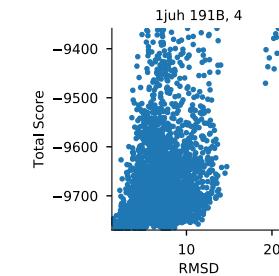
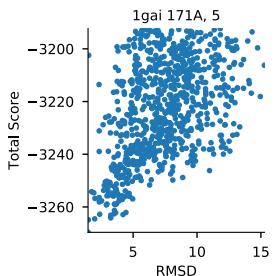
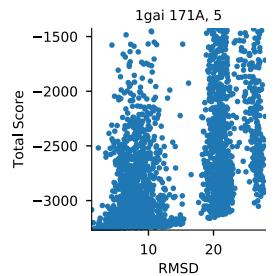
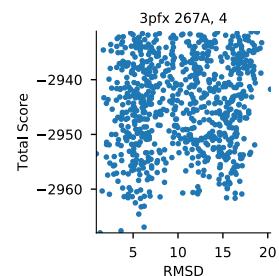
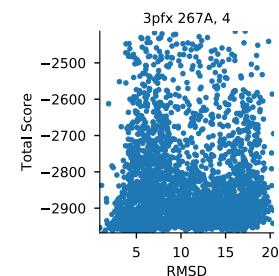
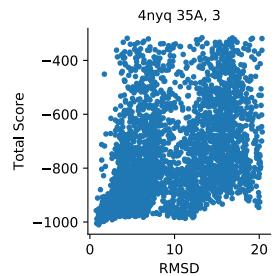


1.48 Å

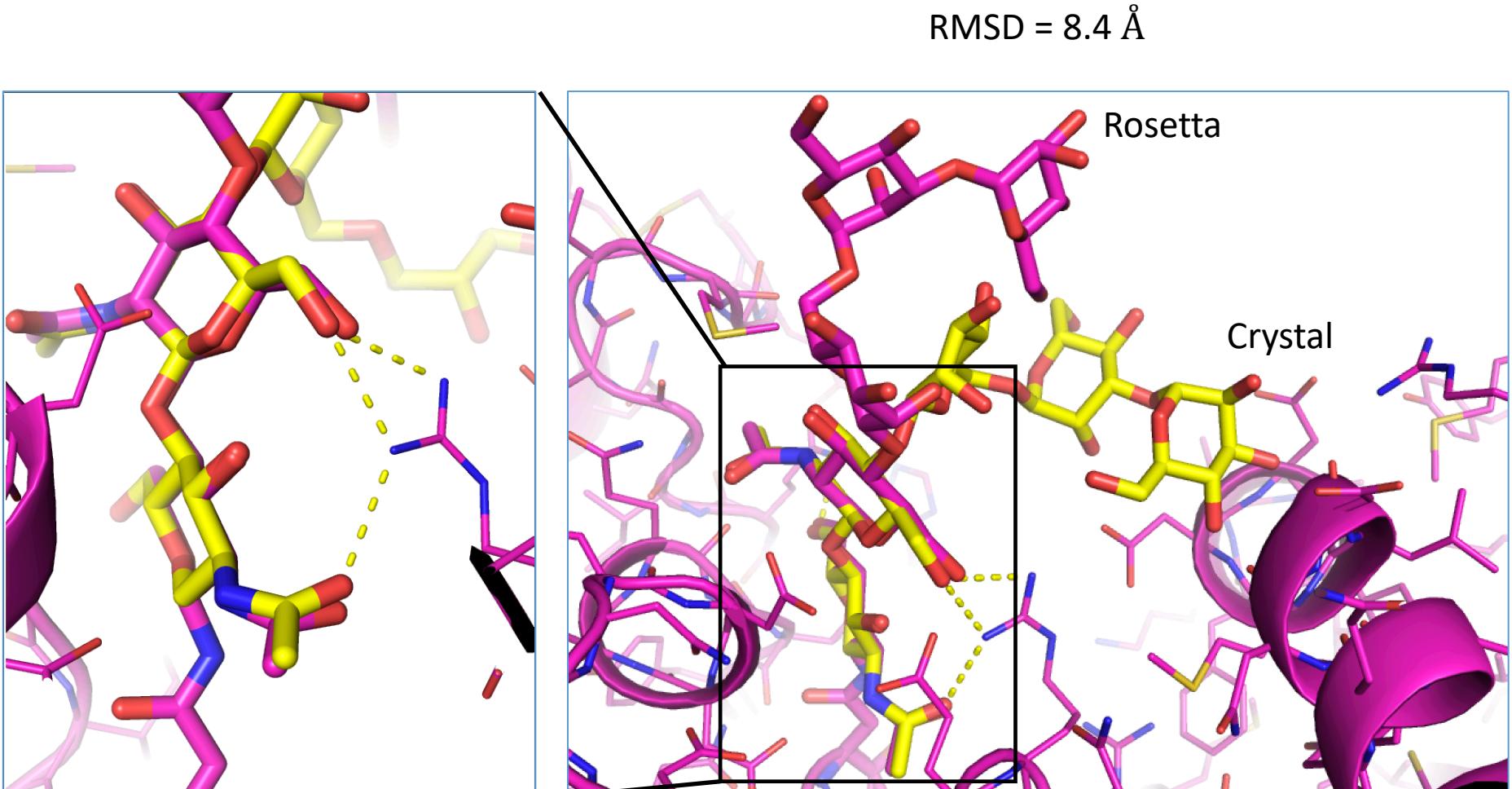
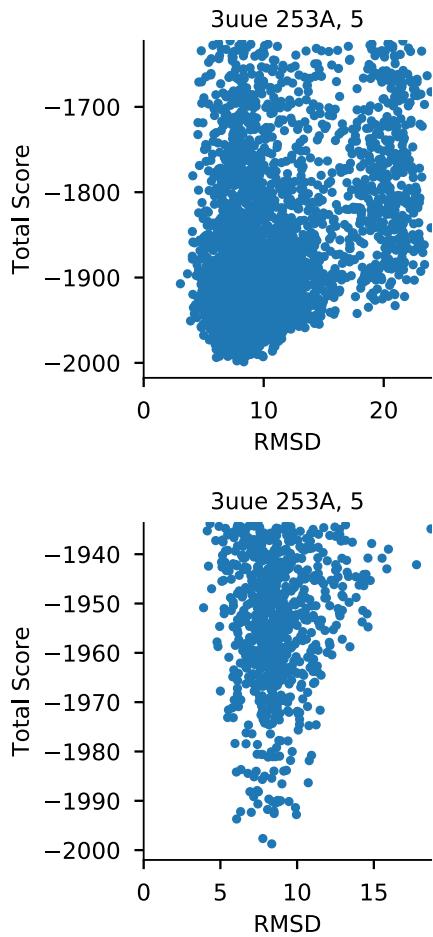
1juh



2.45 Å



# Bent conformation pose a problem



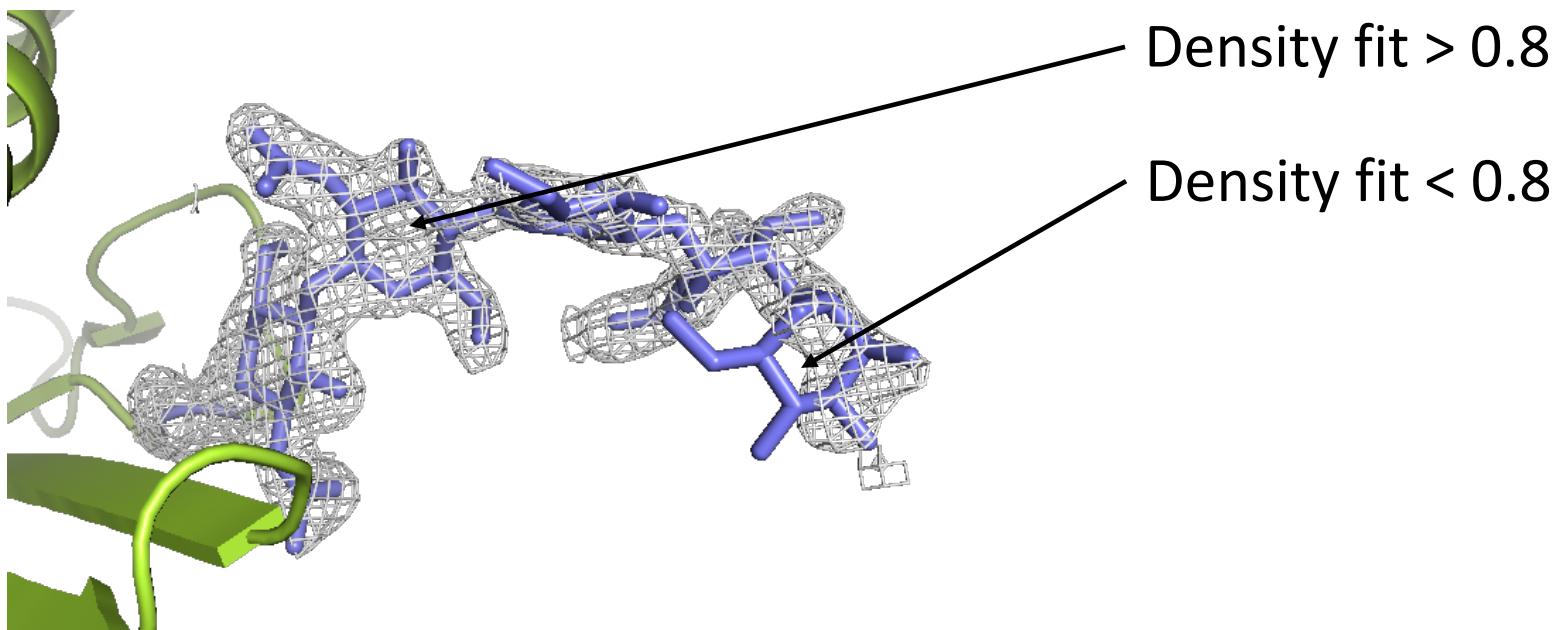
- Current protocol might not capture this very well
- Scoring of protein-glycan interactions inadequate

## Benchmark: Building into density maps

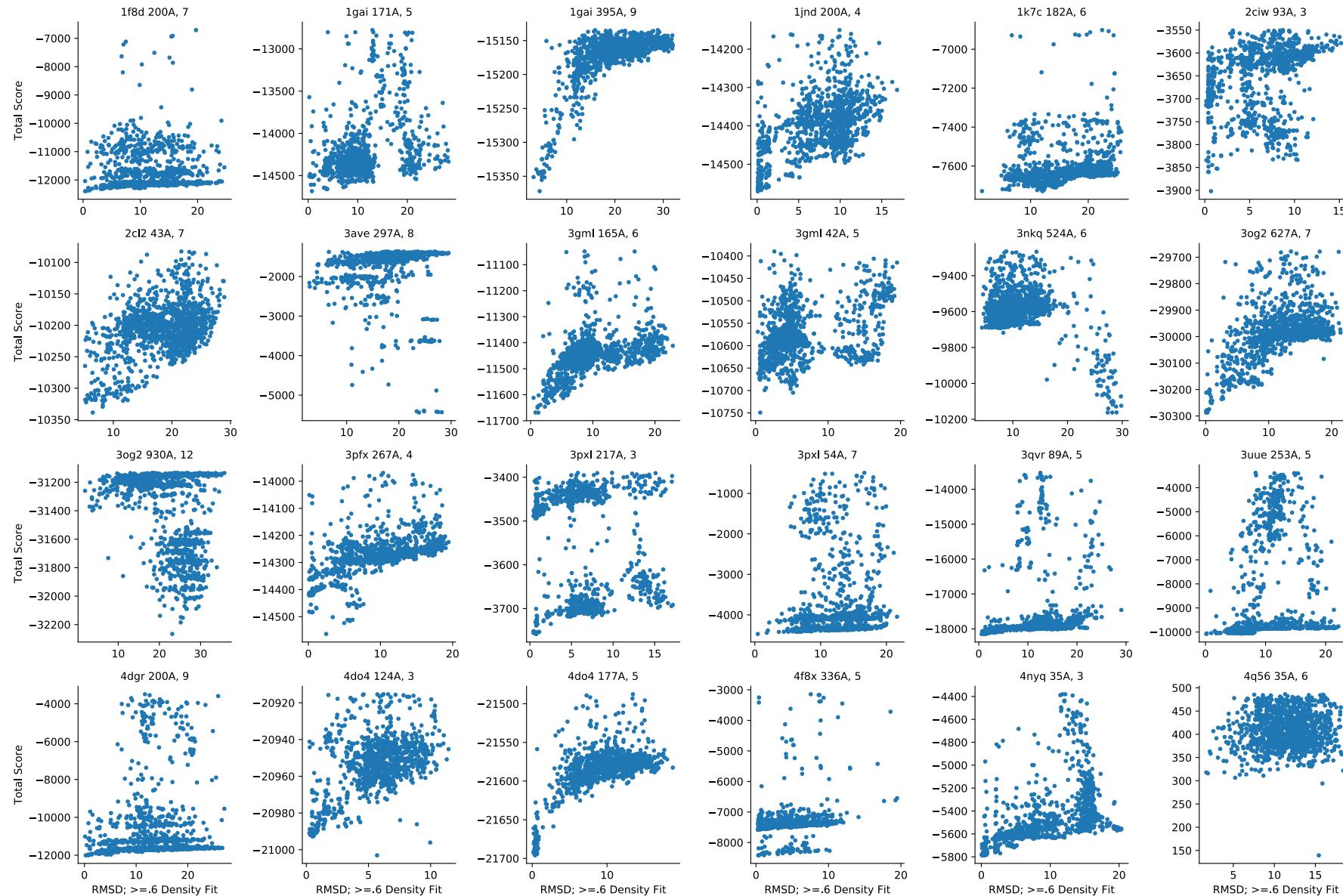
---

**fast\_elec\_dens=25**

Rcon data; **nstruct=750**

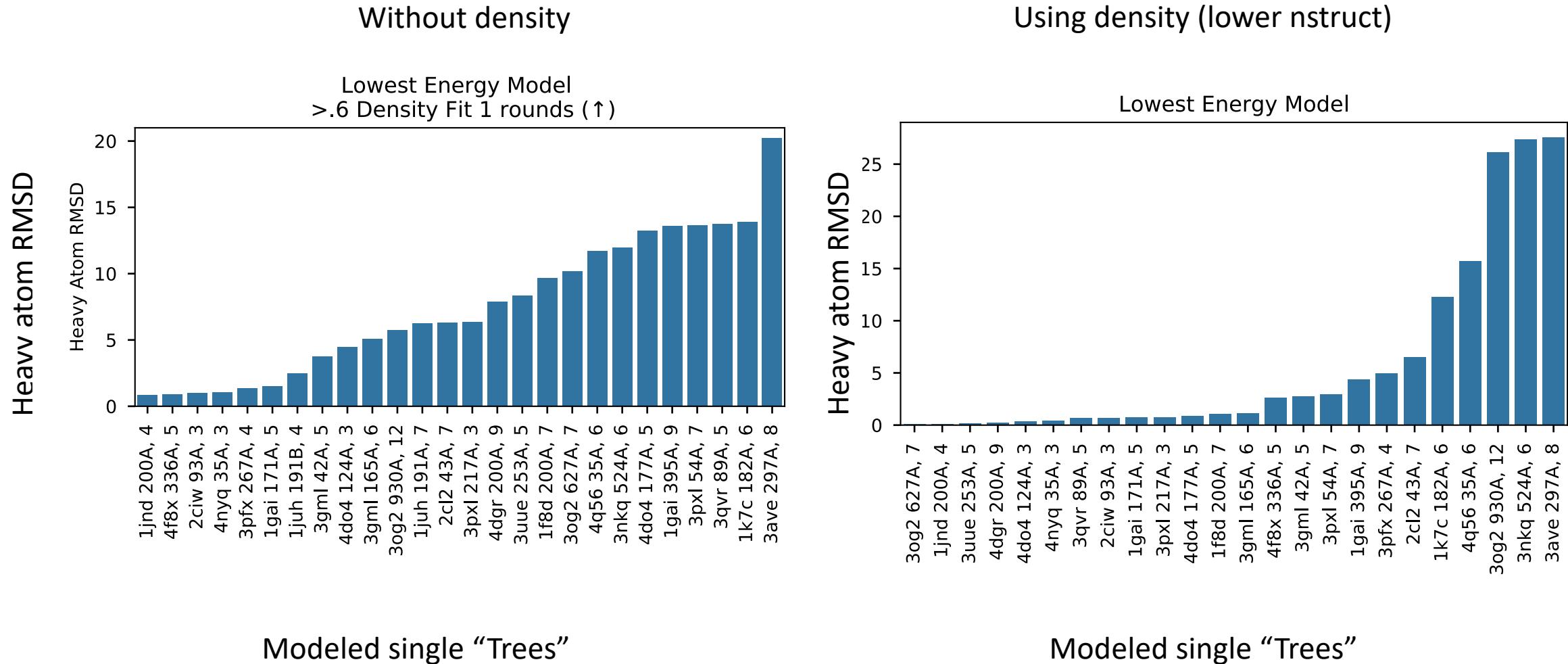


# Building into electron density looks promising – but its not perfect, yet



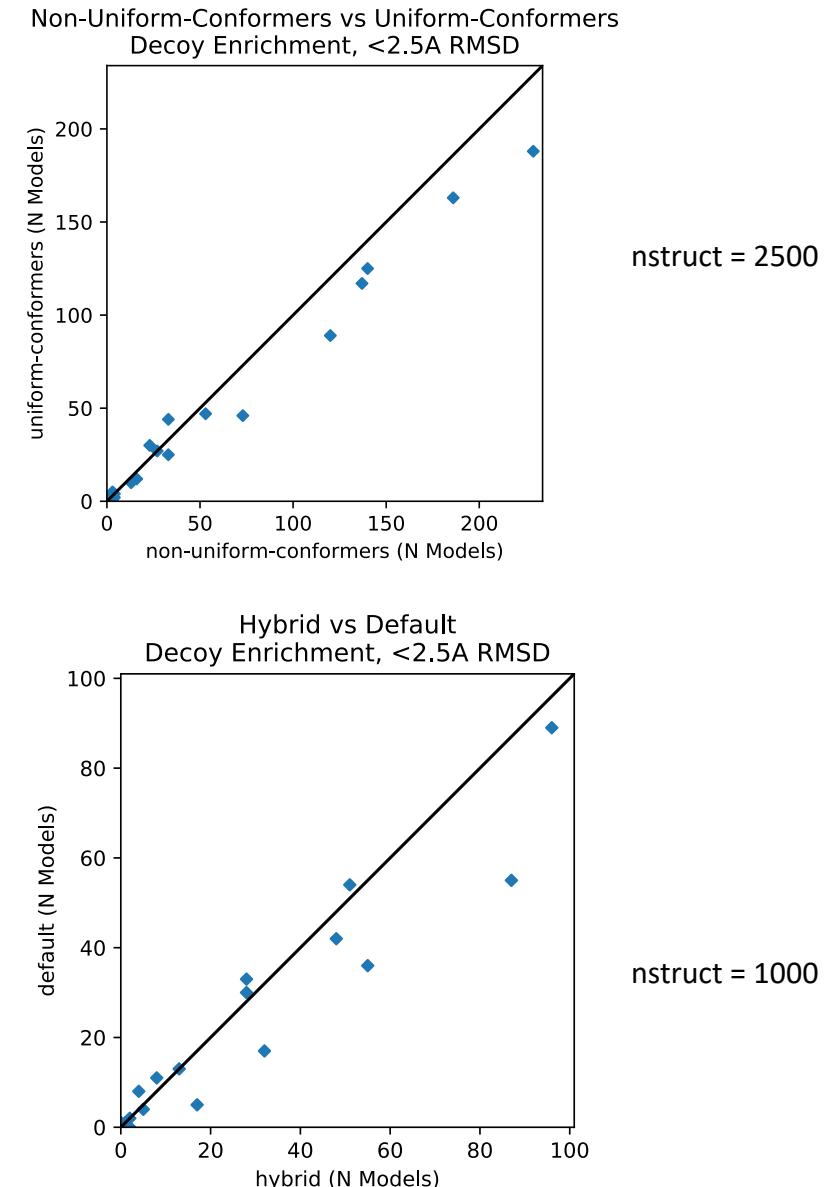
Rcon Results  
nstruct = 750

# Preliminary data: 12/24 trees have sub-angstrom accuracy



## Current Benchmarking

- Parameter/Feature optimization:
  - Shear Sampling
  - Conformer sampling on Gaussian
  - Hybrid Sampling
  - $kT$ , scoring
- Improved sampling of near-natives



## All New Tools from Project

---

- Sequon Creation:
  - GlycanSequonCreator
  - SequenceMotifMover
  - SequenceMotifOperation
  - ResfileCommandOperation
- Glycosylation:
  - SimpleGlycosylateMover
- Residue Selection:
  - GlycanResidueSelector
  - GlycanLayerSelector
  - GlycanSequonSelector
- Modeling:
  - LinkageConformerMover
  - GlycanTreeMinMover
  - GlycanTreeSampler
  - GlycanTreeModeler
- Frameworks:
  - BBSampler Framework
  - **SimpleMetric Framework**
    - 6 metric types
    - ~20 implemented metrics
- Apps:
  - **rosetta\_scripts\_jd3**
  - sugar\_coat
- Etc:
  - RosettaScripts in PyRosetta

## Summary

---

1. Preliminary glycan sampling yields low-energy conformations, similar to native structures
2. Glycan symmetry and RMSD calculations implemented
3. Solving glycoprotein structures automatically is within reach

### New tools in Rosetta

1. Importing glycan structures from PDB drastically improved ([Thanks Frank and Brandon for code contribution!](#))
2. *Sequence motif*-based Movers
3. *Fragment* framework extended to go beyond phi, psi and omega
4. Many glycan-related ResidueSelectors
5. SimpleMetrics system introduced

- 1) **Create a Glycan sequence motif in a protein**, which is recognized by Glycosyltransferase
- 2) **Add a common glycan to the protein** using the SimpleGlycosylateMover
- 3) **Model the glycan** using the GlycanTreeModeler
- 4) **Model the glycan using experimental density**, density fitting tools, and SimpleMetrics

## Acknowledgements

---

- Sebastian Raemisch
- Jason Labonte (JHU)
- Chris Bahl (Harvard)
- Maxim Shapavolov (Dunbrack Lab)
- Jesper Palleson (IU)
- Frank Dimaio (UW)

