Automating the Rational Design of Glycomimetics

Robert J. Woods

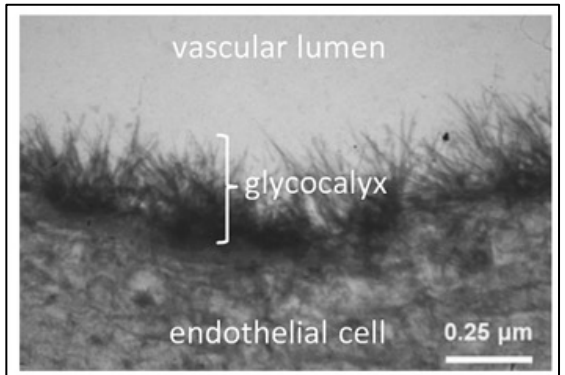
Complex Carbohydrate Research Center University of Georgia

GLYCAM-Web: glycam.org

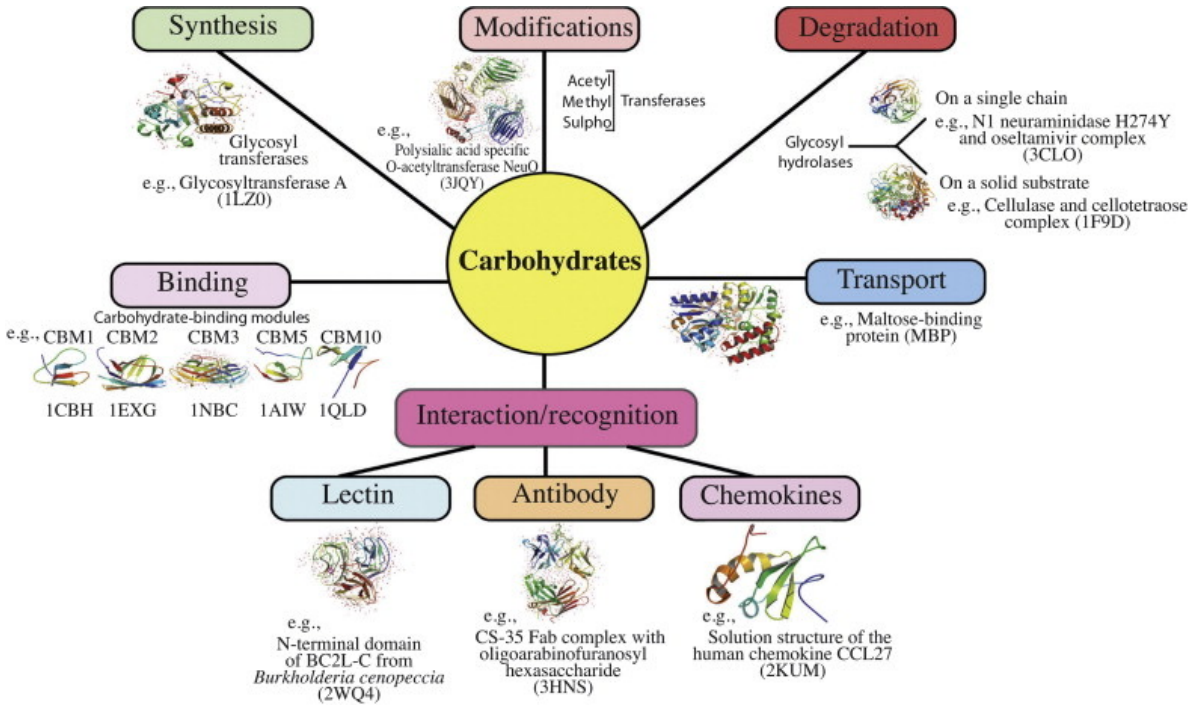
Carbohydrate Recognition in Human Health

Mammalian cells are covered in a complex forest of glycoconjugates (**the glycocalyx**).

Carbohydrate-protein are key for cell-cell and host-pathogen recognition and are therefore potentially important therapeutic targets.



Wiesinger A, Peters W, Chappell D, Kentrup D, Reuter S, Pavenstädt H, et al. (2013) Nanomechanics of the Endothelial Glycocalyx in Experimental Sepsis. PLoS ONE 8(11): e80905



Pérez and Tvaroška (2014) Adv. Carbohydr. Chem. Biochem. 71, 9-136

Glycan binding proteins (GBPs)

- Essential to normal cell growth and development
- Used by pathogens (viruses and bacteria) to adhere to host cells
- Transport carbohydrates for catabolism
- Modulate protein folding and secretion

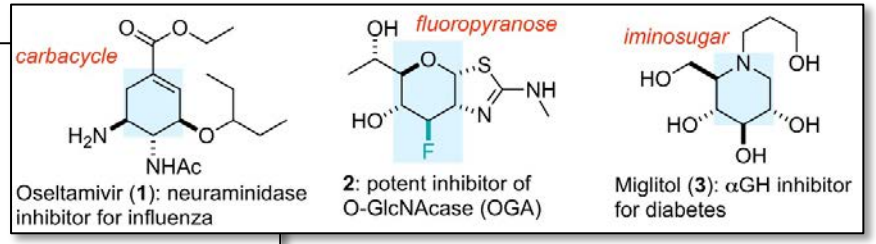
Carbohydrate-processing enzymes

- Synthesize and degrade glycans
- Exploited by hosts and pathogens

Approaches to Inhibiting Carbohydrate Binding

Glycomimetics:
Meanwell, et al. *Commun. Chem.*, 4, 96 (2021).

- Convert an endogenous carbohydrate into a drug



Traditional drug-like molecules:
Shanina, et al. (2022) *Commun. Chem.*, **5**, 64

- Discover a drug that is specific for a carbohydrate binding site

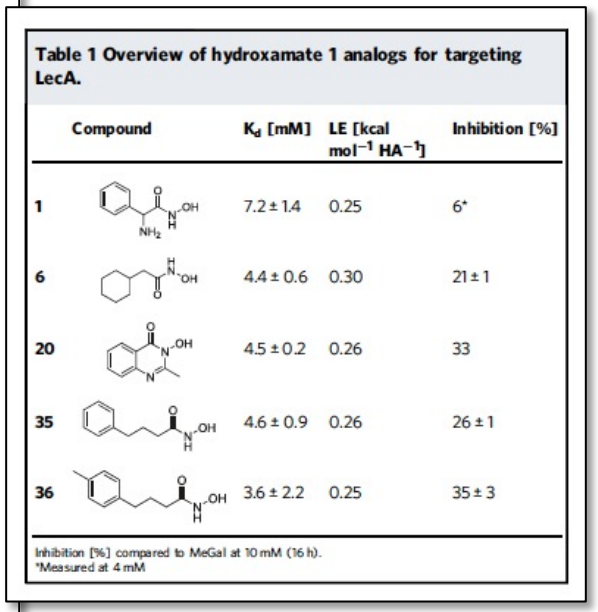


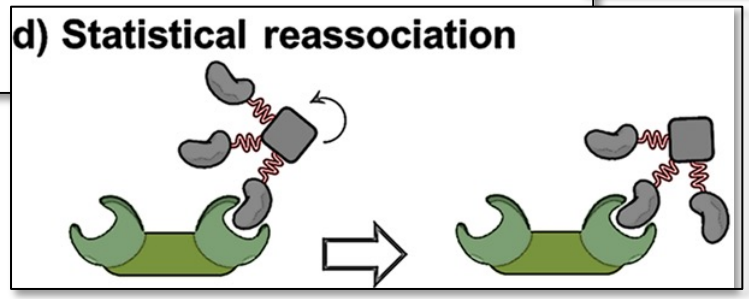
Table 1 Overview of hydroxamate 1 analogs for targeting LecA.

Compound	K_d [mM]	LE [kcal mol^{-1} HA^{-1}]	Inhibition [%]
1	7.2 ± 1.4	0.25	6*
6	4.4 ± 0.6	0.30	21 ± 1
20	4.5 ± 0.2	0.26	33
35	4.6 ± 0.9	0.26	26 ± 1
36	3.6 ± 2.2	0.25	35 ± 3

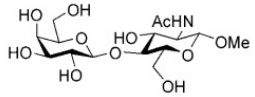
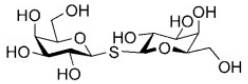
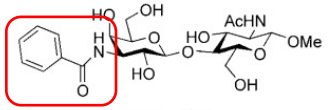
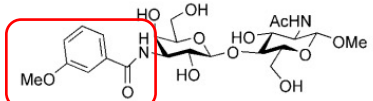
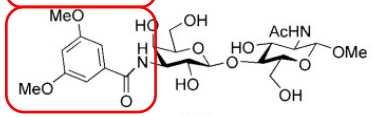
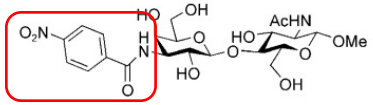
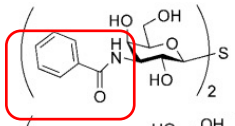
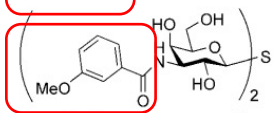
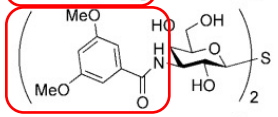
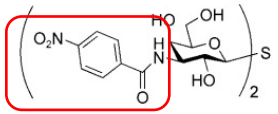
Inhibition [%] compared to MeGal at 10 mM (16 h).
*Measured at 4 mM

Multivalent glycoconjugates (dendrimers):
Cecioni, et al. (2015) *Chem. Rev.*, **115**, 525

- Amplify apparent affinity by avidity

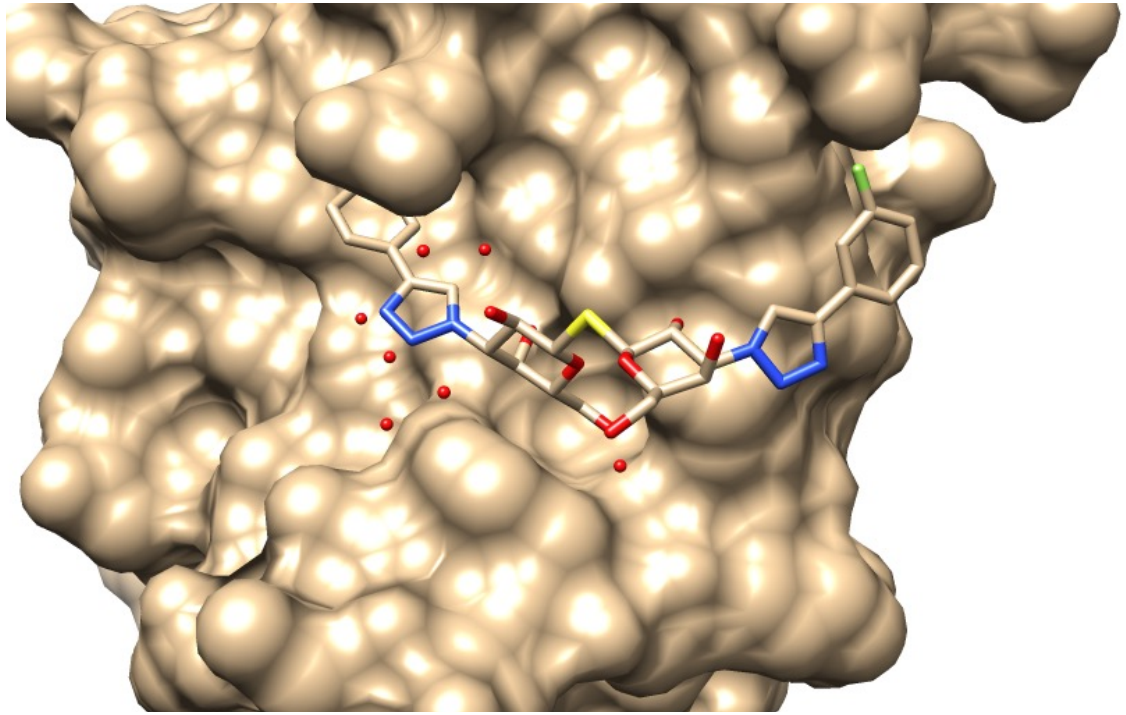


Glycomimetics as a Therapeutic Strategy

	Compound	K_d [nM]	Relative activity[a]
9		69 000[5b]	1
3		43 000	1.6
2a		6700[5b]	10
2b		2500[5b]	28
2c		1100[5b]	63
2d		950[5b]	73
4a		3000	23
4b		61	1130
4c		50	1380
4d		33	2090

[a] Compounds **9**, **2a–d**, and **3** are included for reference.

Cumstey et al. *Angew. Chem. Int. Ed.* **44**, 5110-5112



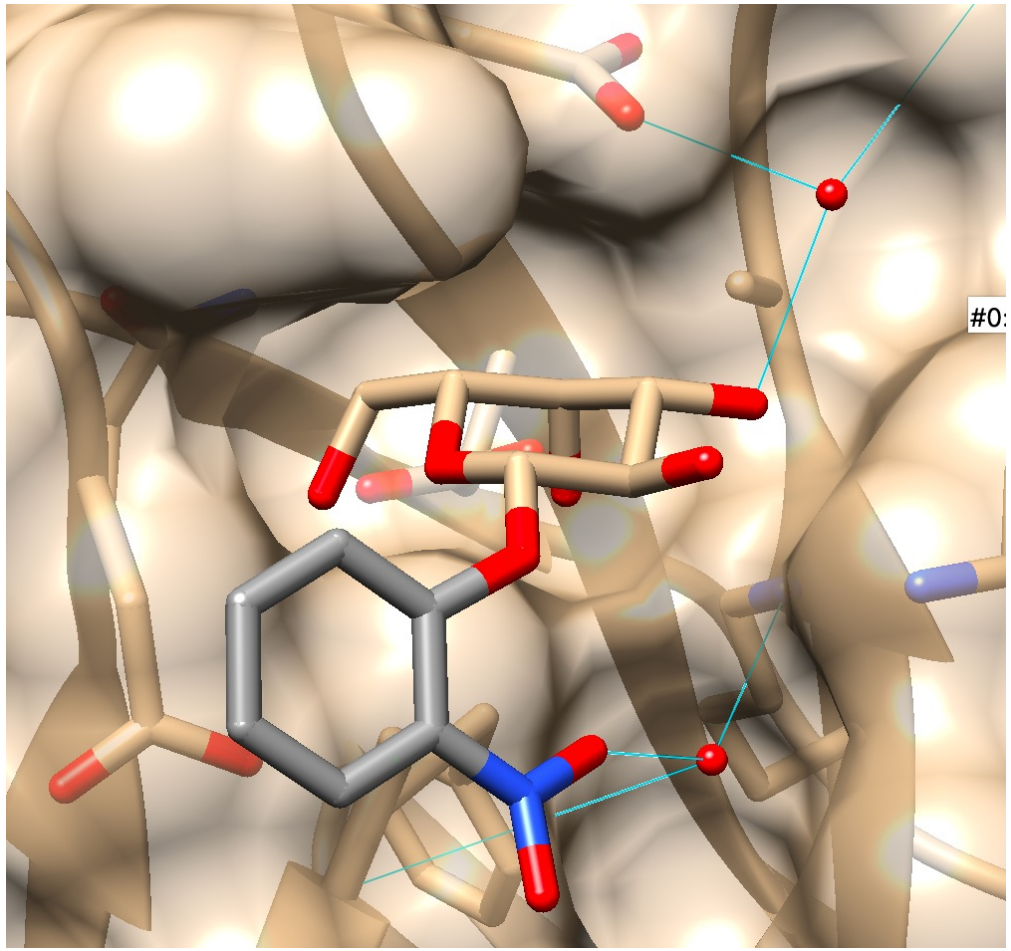
- Exploit the specificity of the endogenous carbohydrate ligand
- Employ the native carbohydrate ligand as a basis for rational design
- Examples: Relenza® and Tamiflu®
- Review: Magnani and Ernst (2009) *Discov. Med.* **8**, 247-252

How to choose the "R" groups?

Inhibiting Protein-Carbohydrate Interactions

- **Glycomimetic compounds:**
  - Contain a carbohydrate core plus drug-like modifications
  - Enhanced binding affinity
  - Enhanced drug-like characteristics (membrane permeability, half life, etc.)

- **Our Project:**
  - Develop a high-throughput virtual screening pipeline for glycomimetic discovery.
  - Automate this and create an online tool for glycomimetic screening
  - Apply it to Influenza and other disease targets



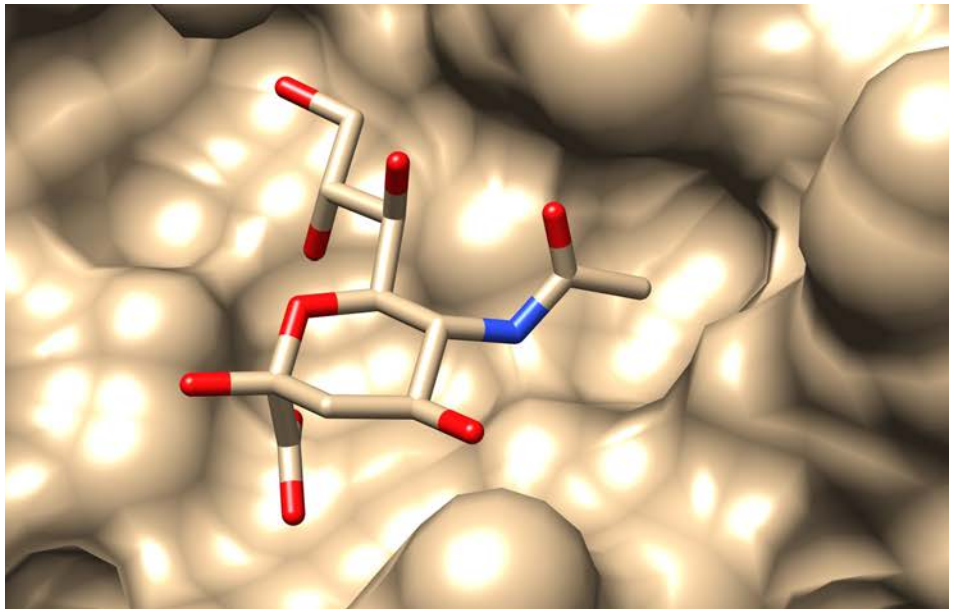
PDB 6AOY[5] (FmlH + ONPG)

[Kalas, et al. (2018).. *Proceedings of the National Academy of Sciences*, *115*(12), E2819-E2828.

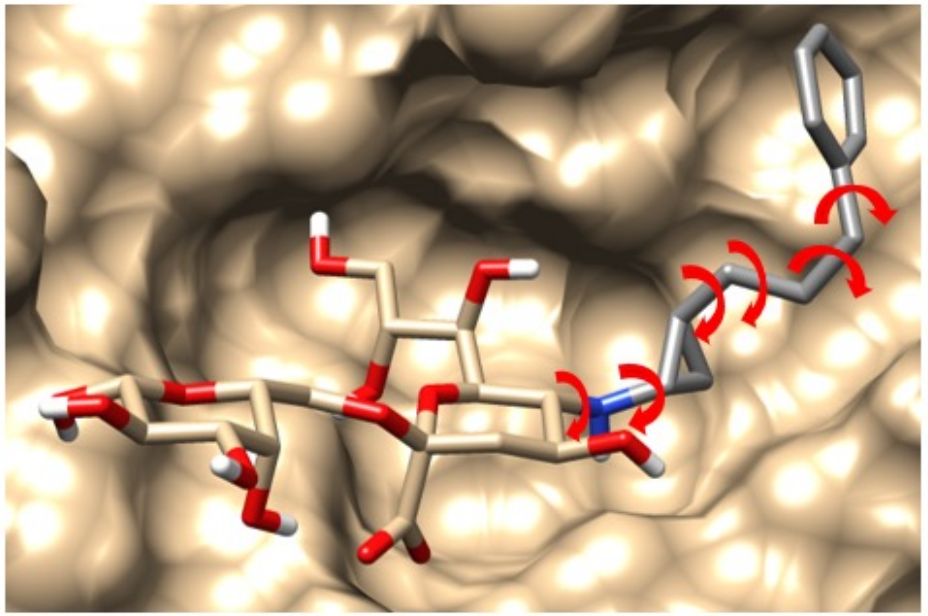
Moiety Grafting and Conformational Sampling

Graft drug-like moieties onto bound carbohydrate and look for optimal orientation

- All rotatable bonds in the chemical moiety are identified and rotated
  - Number of Rotamers $= \prod_{i=1}^{N} \frac{360}{\theta_i}$, 6 bonds, sampled at 10^{o} increments = 2.2×10^{9} rotamers!
- A genetic algorithm is employed for conformational sampling

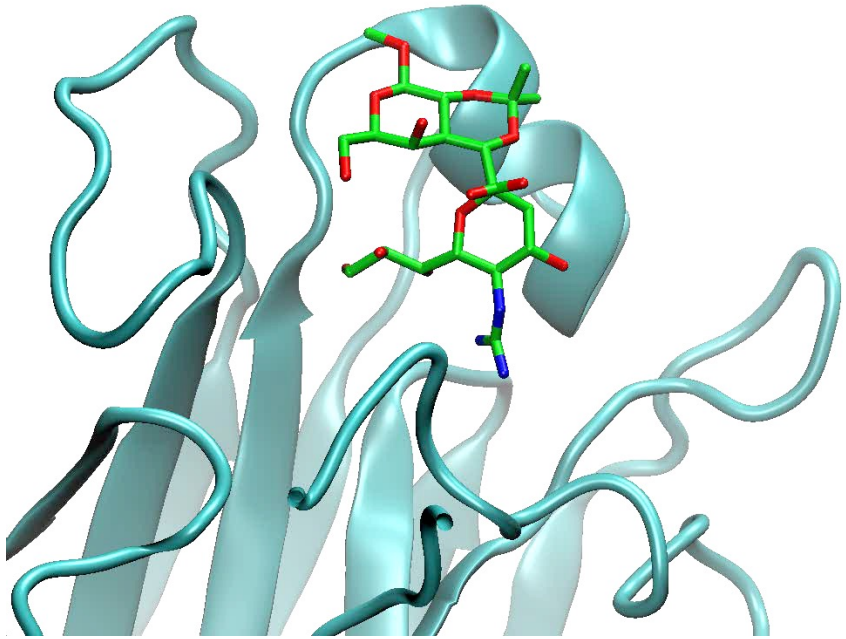


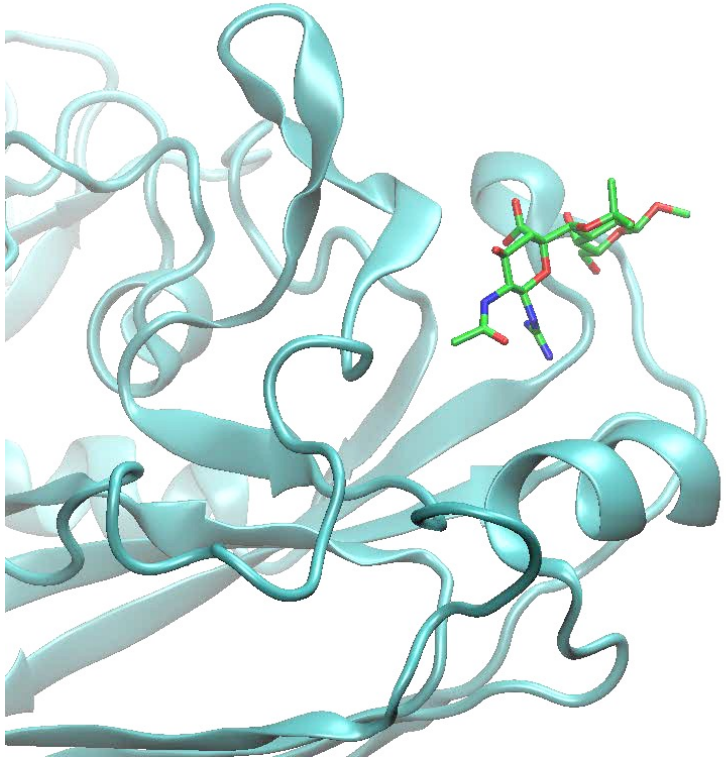
Moiety Grafting
Rotamer Sampling



Molecular Dynamics Can Discriminate Strong from Weak

Which inhibitors should we simulate?





Putative Influenza Hemagglutinin Inhibitor 1	Putative Influenza Hemagglutinin Inhibitor 2

Automated Virtual Glycomimetic Screening

Robust and **reproducible** data sets.

Expandible in response to user demand/scientific developments.

Standardized simulation conditions, otherwise highly prone to user error.

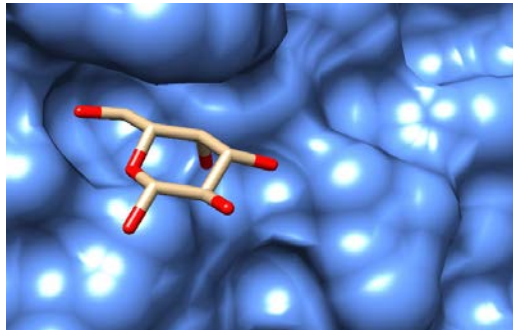
Eliminates complex software installation and training.

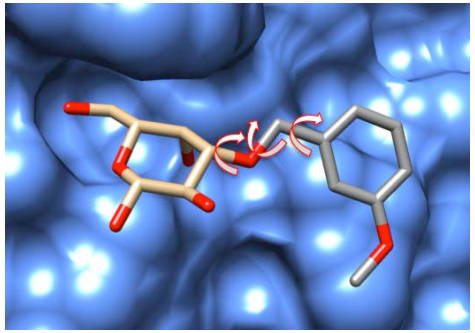
Accessibility to non-experts in computational chemistry.

Enhanced user access to sophisticated modeling tools.

Motivation: Translate modeling technology into the experimental laboratory

Virtual Glycomimetic Screening





Upload protein-carbohydrate PDB file

Select position in ligand to modify

Select library of analogs to screen

Perform conformational search for each analog
- rotamer generation
- clash checking
- energy minimization

Pass MD snapshots at 5 ns intervals

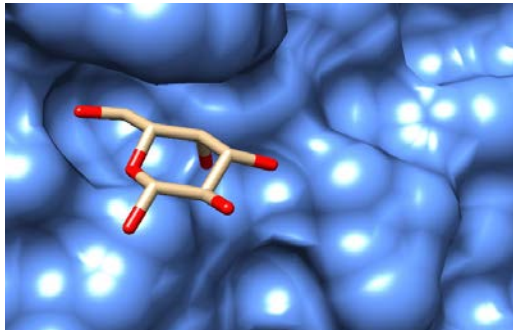
Perform MD of receptor-carbohydrate co-complex
- explicit water
- AMBER GAFF force field
- 50 ns

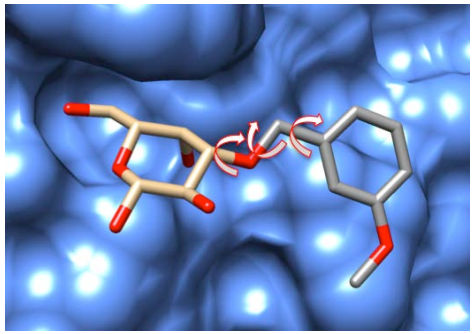
Calculate GIST desolvation energy and docking scoring for all rotamers of each analog in each MD snapshot

Compute ligand conformational entropy

Return best pose and estimated binding free energy for each analog

Virtual Glycomimetic Screening





Upload protein-carbohydrate PDB file

Select position in ligand to modify

Select library of analogs to screen

Perform conformational search for each analog
- rotamer generation
- clash checking
- energy minimization

Pass MD snapshots at 5 ns intervals

Perform MD of receptor-carbohydrate co-complex
- explicit water
- AMBER GAFF force field
- 50 ns

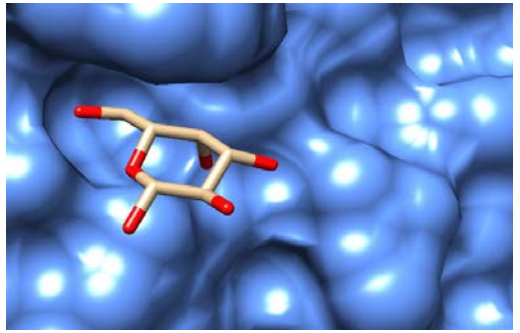
Calculate GIST desolvation energy and docking scoring for all rotamers of each analog in each MD snapshot

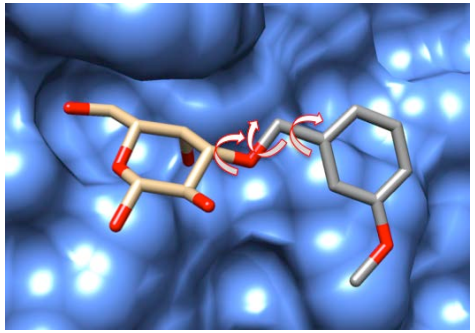
Compute ligand conformational entropy

Return best pose and estimated binding free energy for each analog

Moiety Library:
- Designed for conjugation at NH/OH groups in carbohydrates
- Moieties scraped from chemical catalogs and PubChem
- Converted from SMILES to 3D
- Currently ~1000 moieties

Virtual Glycomimetic Screening





Upload protein-carbohydrate PDB file

Select position in ligand to modify

Select library of analogs to screen

Perform conformational search for each analog
- rotamer generation
- clash checking
- energy minimization

Pass MD snapshots at 5 ns intervals

Perform MD of receptor-carbohydrate co-complex
- explicit water
- AMBER GAFF force field
- 50 ns

Calculate GIST desolvation energy and docking scoring for all rotamers of each analog in each MD snapshot

Compute ligand conformational entropy

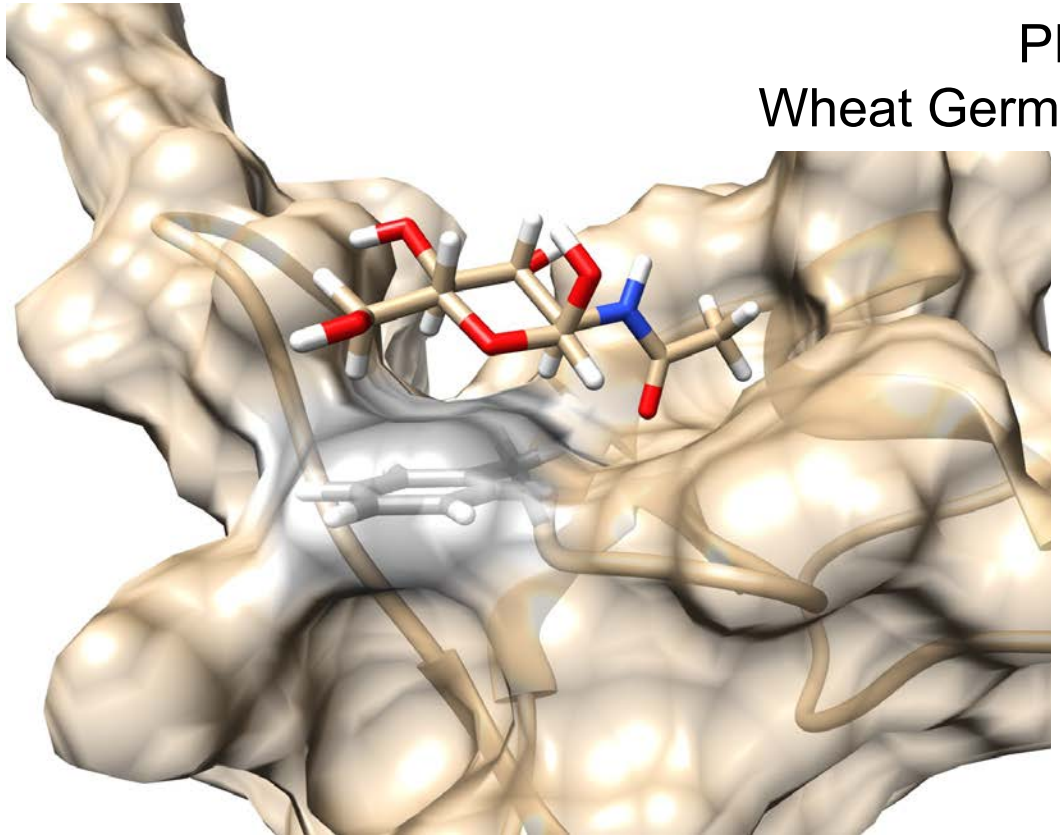
Return best pose and estimated binding free energy for each analog

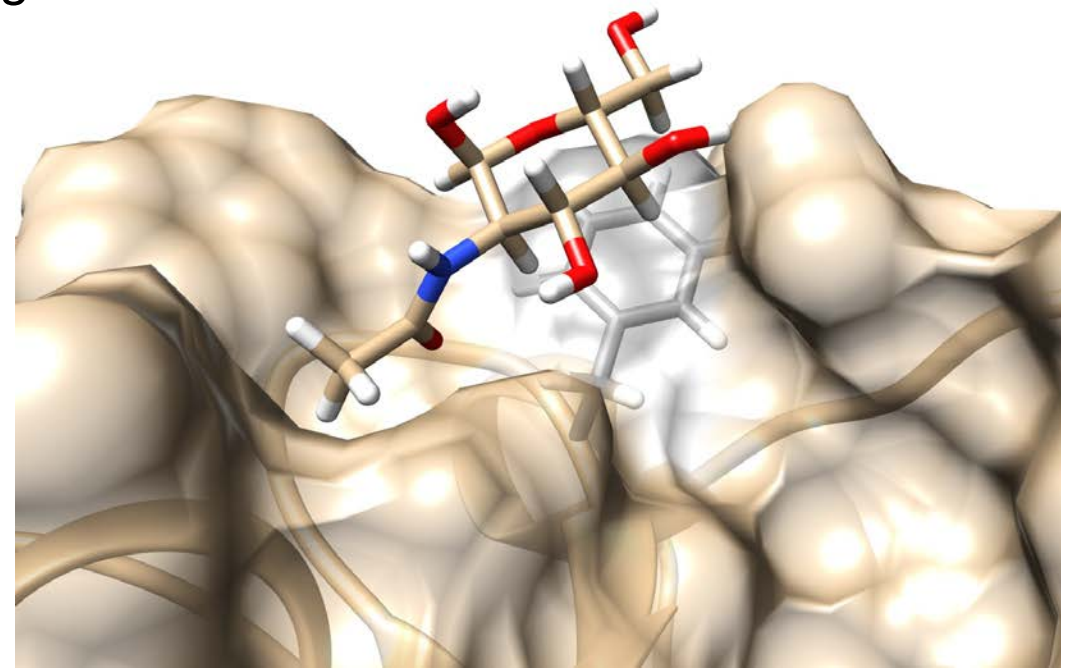
Binding Energies:

- Multiple methods implemented
  - AutoDock VINA-Carb [1]
  - AMBER/GLYCAM MM-GBSA
- New functional forms added (CH-π)

CH-π Interactions in Carbohydrate Binding

PDB 2UVO
Wheat Germ Agglutinin + GlcNAc





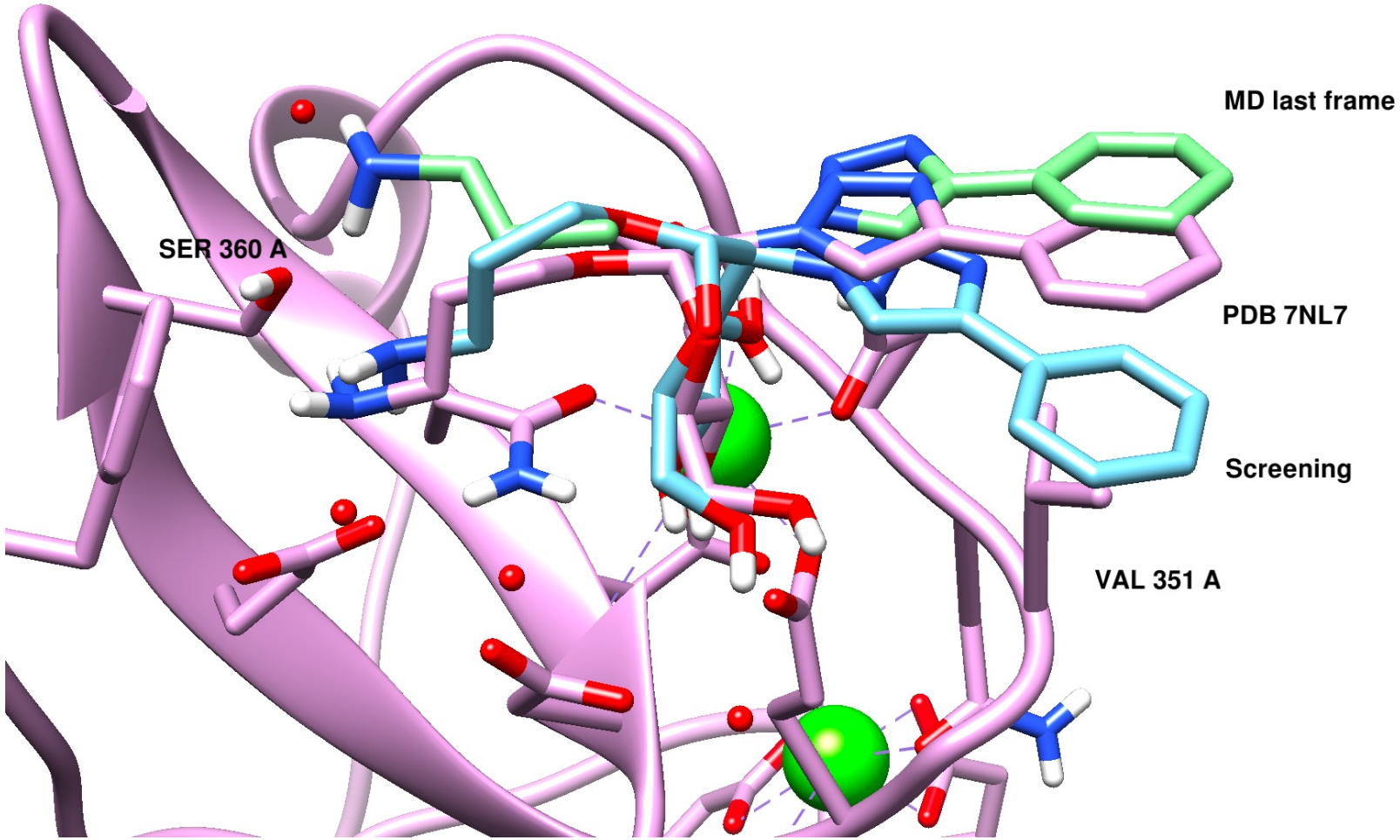
AutoDock Vina for Carbohydrates (Vina-Carb): Nivedha, et al. (2016) *J. Chem. Theory Comput.*, *12*, 892-901.
AutoDock Vina with CH−π: Xiao, Y., & Woods, R. J. (2023) *J. Chem. Theory Comput.*, *19*(16), 5503-5515.

Case Studies

Carbohydrate Binding Protein	Endogenous Ligand	Function	Number of Reported Mimetics	Number of co-crystal structures
DC-SIGN	High-mannose N-glycans	Pathogen Recognition	13	1
Galectin-1	Beta-galactosides	Cell-cell / matrix interactions	11	0
Galectin-3	Beta-galactosides	Cell adhesion, growth, apoptosis, etc	12	3
FimH	Terminal mannoses	*E.coli* adhesin, urinary tract infection	7	8
FmlH	Gal/GalNAc	*E.coli* adhesin, UTI	9	7
Siglec-2	Neu5Ac/Gcα2-6Gal	B cell activation	42	0
Siglec-4	Sialylated gangliosides	Axon regeneration	25	0
Siglec-7	Neu5Acα2-8Neu5Ac	Natural killer cell inhibition	22	1
Siglec-8	6'-sulfo-sLe^x/LacNAc	Mast cell/eosinophil apoptosis	11	1
LecA	Glycosphingolipid Gb3	*Pseudomonas* host cell invasion	28	8
LecB	Fucose glycoconjugates	Biofilm formation	22	7
Cholera Toxin	GM1 gangliosides	Host cell invasion	11	7

Success Example: DC-SIGN

A lectin involved in immunity. Exploited for infection by HIV and COVID.



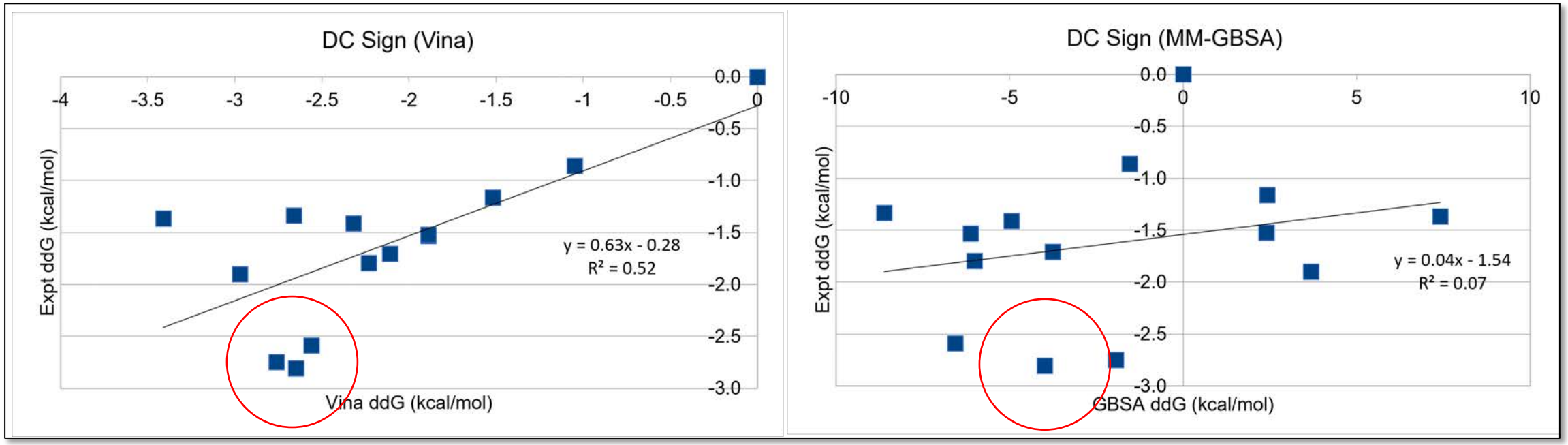
Initial computational screening reproduces crystal structure

MD simulation reproduces crystal structure

Cramer, et al. *Journal of the American Chemical Society* **2021,** *143* (42), 17465-17478
Geijtenbeek, et al. (2003).. *Dendritic Cells and Virus Infection*, 31-54.

Statistical Correlation to Experimental Affinity

Vina-Carb with CH-π significantly outperformed MM-GBSA in this system

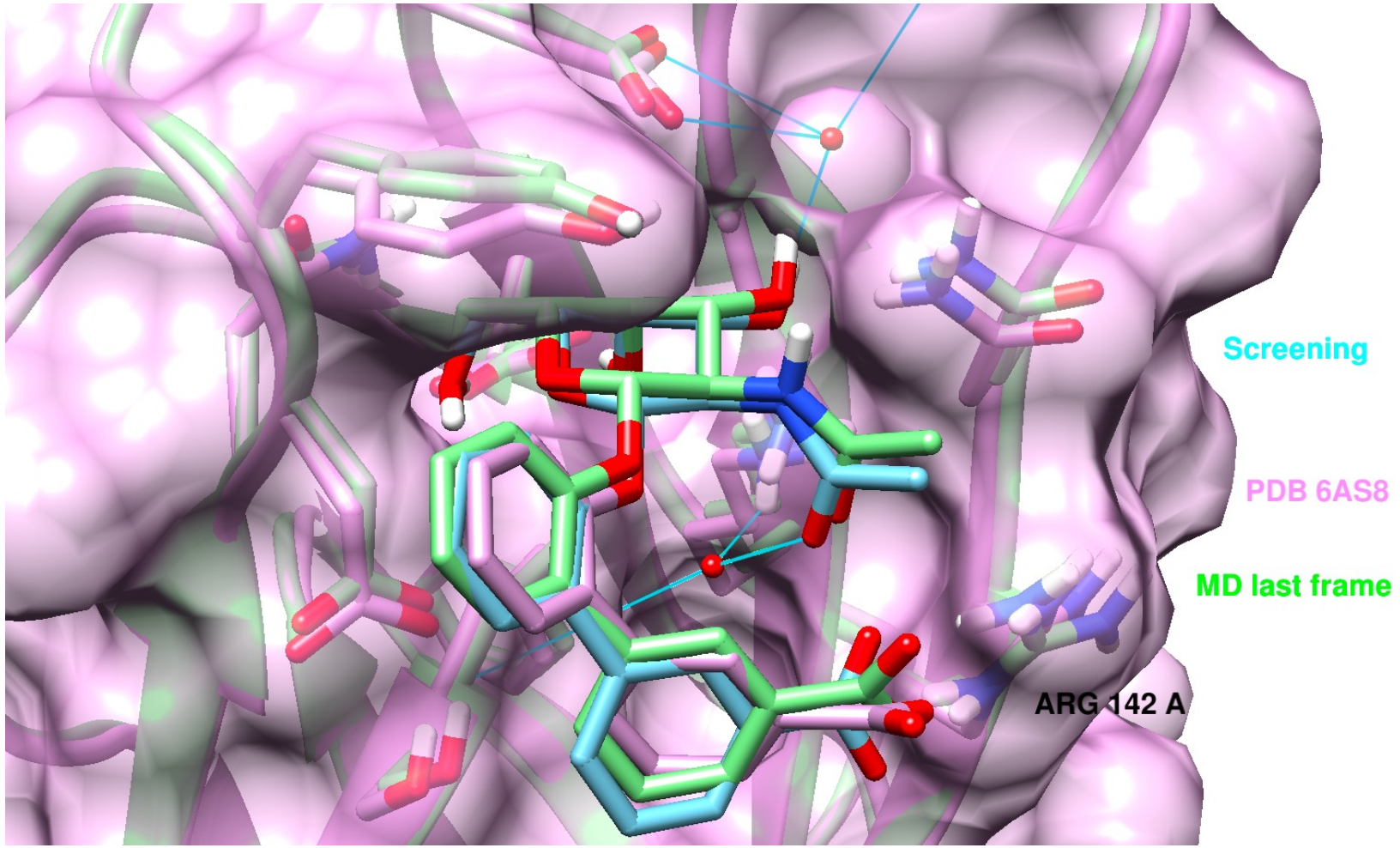


Vina-Carb: $R^2 = 0.52$

MM-GBSA: $R^2 = 0.07$

Success Example: FimH/FmlH

FimH (*E. coli*) binds to Gal epitopes on human epithelial cells, causing urinary tract infections

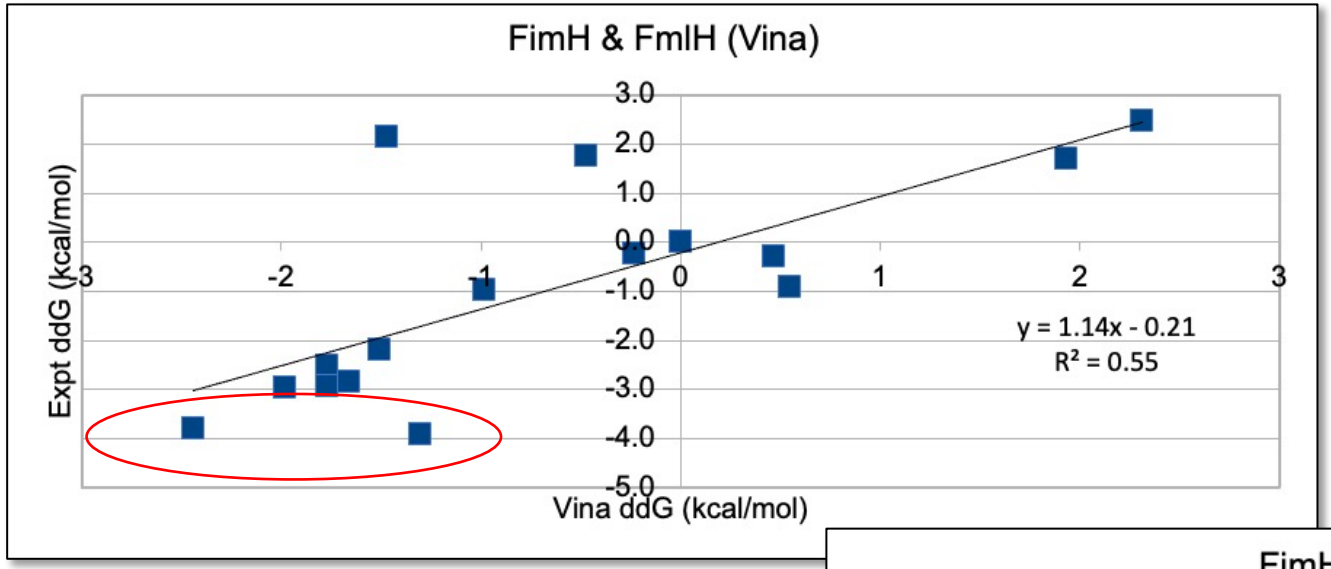


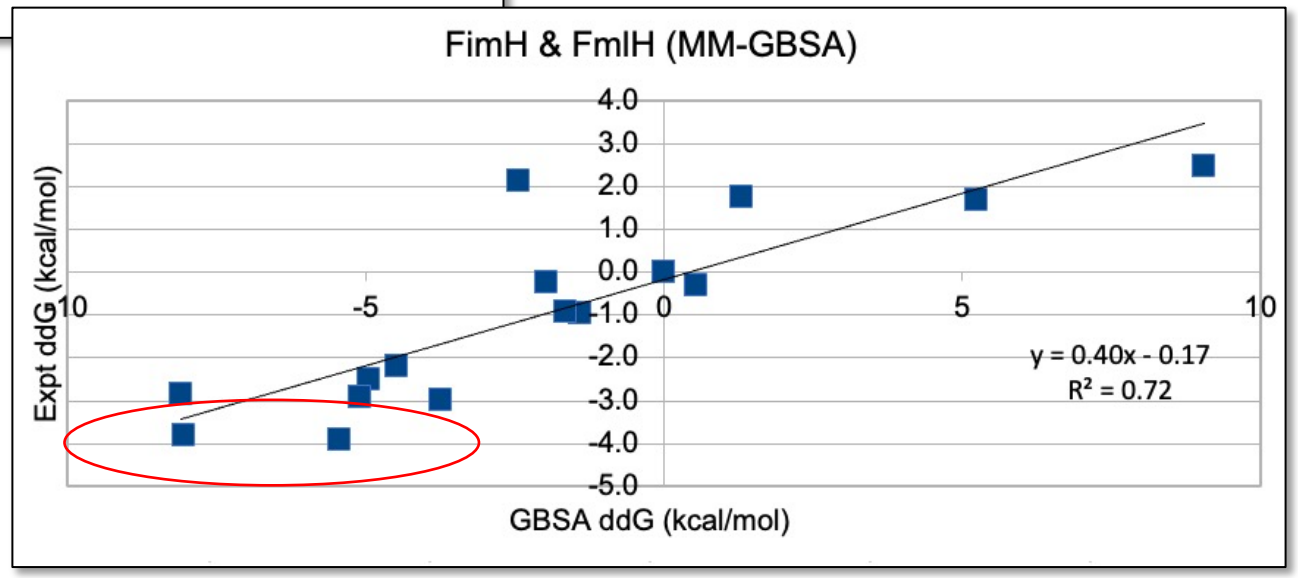
Initial computational screening reproduces crystal structure

MD simulation reproduces crystal structure

Kalas, et al. (2018). *P.N.A.S.*, *115*(12), E2819-E2828.

FimH & FmlH: Computed versus Experimental Affinity





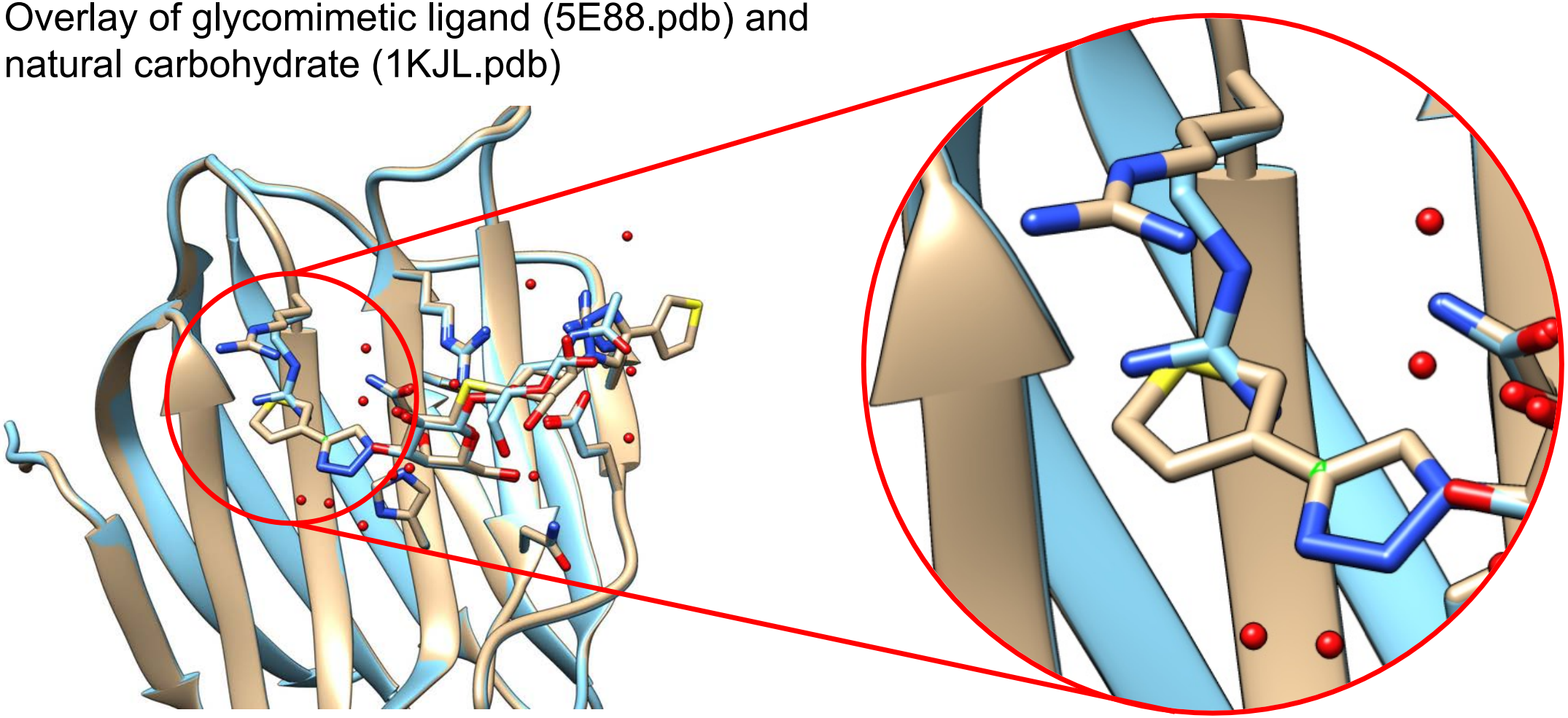
Vina-Carb: $R^2 = 0.55$

MM-GBSA: $R^2 = 0.72$

[10] Hartmann, M., & Lindhorst, T. K. (2011). The bacterial lectin FimH, a target for drug discovery–carbohydrate inhibitors of type 1 fimbriae-mediated bacterial adhesion. *European Journal of Organic Chemistry*, *2011*(20-21), 3583-3609.

Problem Example: Galectin-3

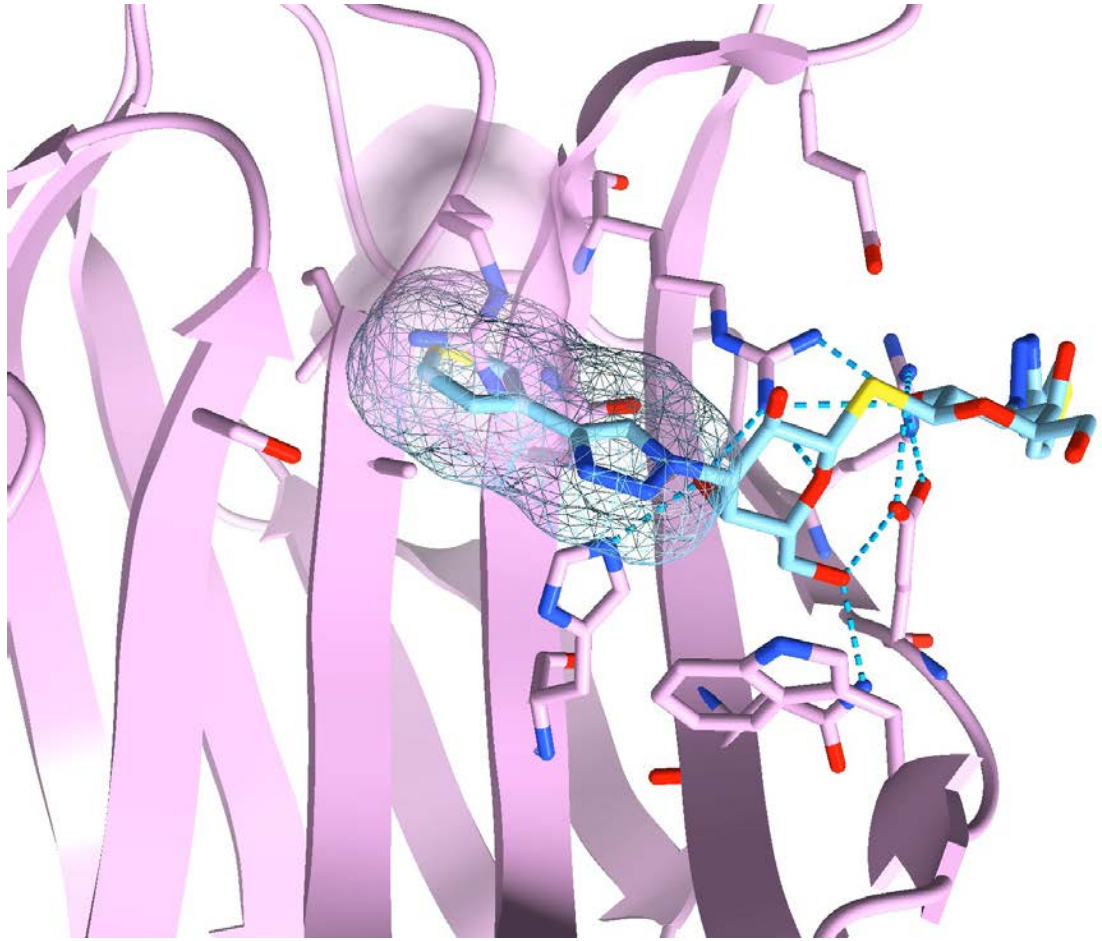
Overlay of glycomimetic ligand (5E88.pdb) and natural carbohydrate (1KJL.pdb)



Requirement for induced fit in ARG 144 causes prediction error.

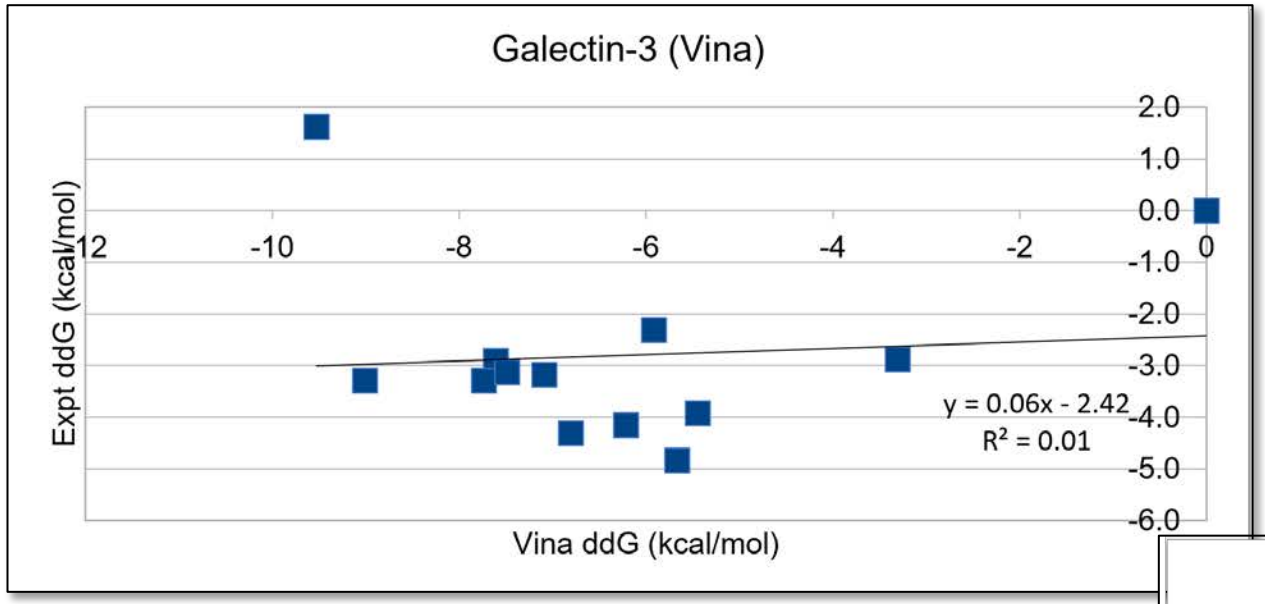
Problem Example: Galectin-3

Morphing of ARG 144 from natural carbohydrate (1KJL.pdb) to glycomimetic ligand (5E88.pdb)



Solution: employ screening with a rotamer library of the nearest amino acid residues

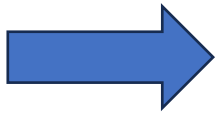
Problem Example: Galectin-3



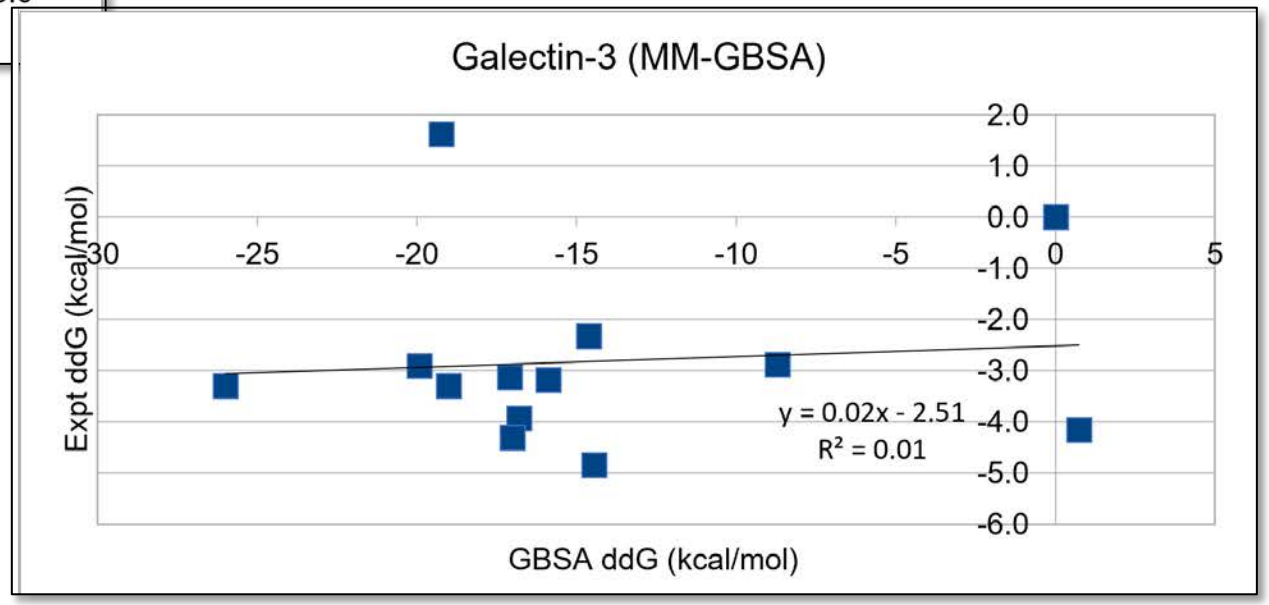
Vina-Carb: $R^2 = 0.01$

MM-GBSA: $R^2 = 0.01$

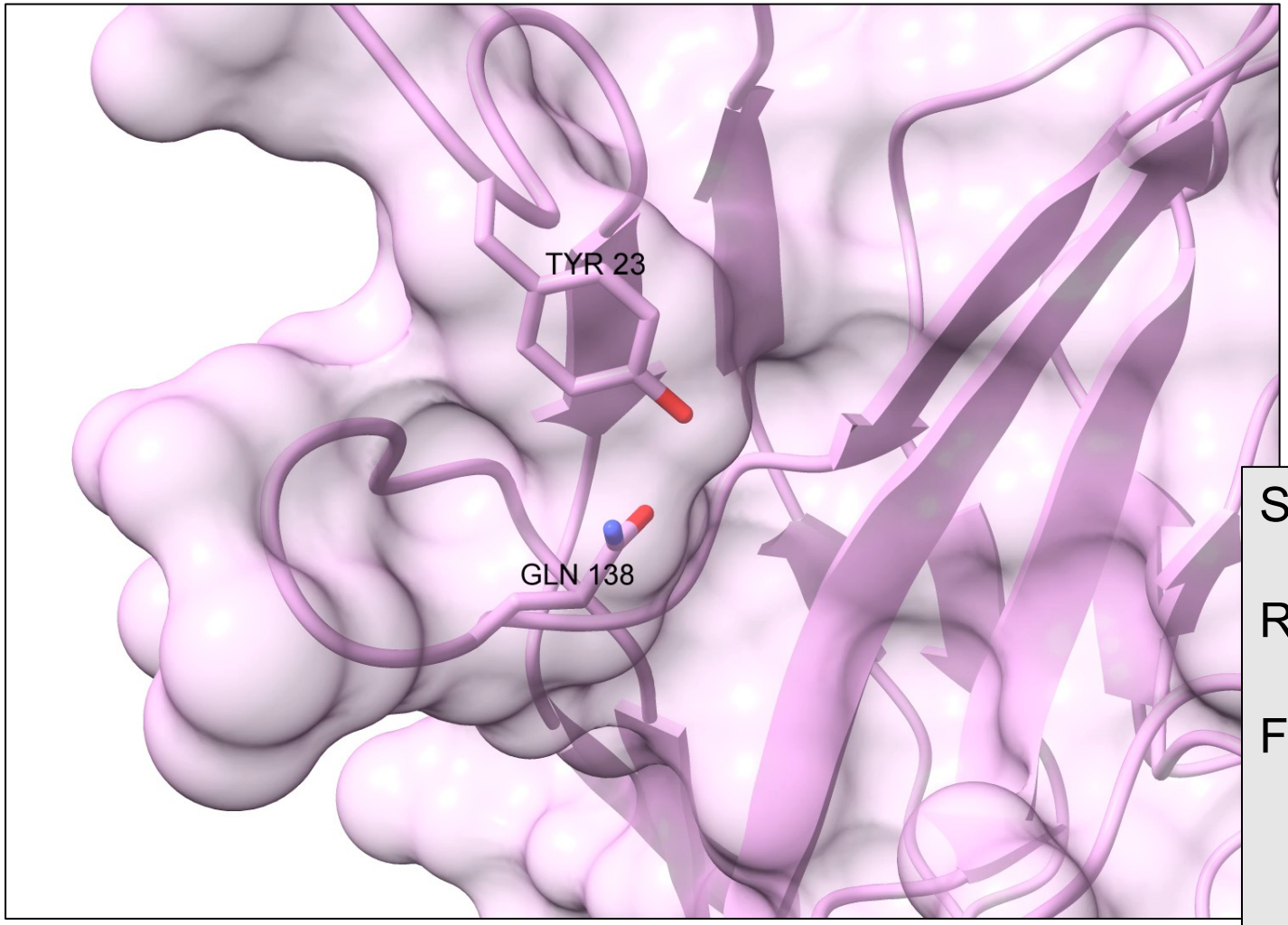
Incorrect Starting Geometry



Erroneous Energies

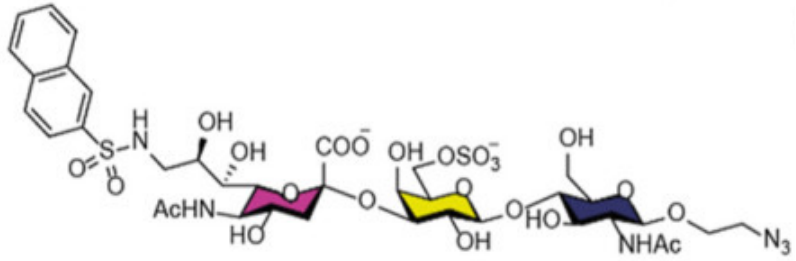


The Problem of Induced Fit in the Backbone: Siglec-8



Apo Protein: 7qu6.pdb

Co-Crystal with Glycomimetic: 7qui.pdb



Screening Protocol:

Rigid Protein
- Typical

Flexible Protein
- Side chains: Employ a rotamer library, Backbone: changes in protein fold are highly problematic for docking

Lenza, et al. *J. Am. Chem. Soc., Au.* (2022) 3:204-215

Results

Protein	Number of Mimetics	Number of crystal structures with mimetics	R^2 after MD (Vina-Carb/pi)	R^2 after MD (MM-GBSA)
DC-SIGN	13	1	0.52	0.07
Galectin-1	11	0	0.01	0.01
Galectin-3	12	3	0.01	0.01
FimH	7	8	0.55	0.72
FmlH	9	7	0.55	0.72
Siglec-2	42	0	0.15	0.14
Siglec-4	25	0	0.15	0.16
Siglec-7	22	1	0.35	0.38
Siglec-8	11	1	0.57	0.61
LecA	28	8	slope < 0	slope < 0
LecB	22	7	0.05	0.02
Cholera Toxin	11	7	0.15	0.28

Results

Protein	Number of Mimetics	Number of crystal structures with mimetics	R^2 after MD (Vina-Carb/pi)	R^2 after MD (MM-GBSA)
DC-SIGN	13	1	0.52	0.07
Galectin-1	11	0	0.01	0.01
Galectin-3	12	3	0.01	0.01
FimH	7	8	0.55	0.72
FmlH	9	7	0.55	0.72
Siglec-2	42	0	0.15	0.14
Siglec-4	25	0	0.15	0.16
Siglec-7	22	1	0.35	0.38
Siglec-8	11	1	0.57	0.61
LecA	28	8	slope < 0	slope < 0
LecB	22	7	0.05	0.02
Cholera Toxin	11	7	0.15	0.28

induced fit

Results

Protein	Number of Mimetics	Number of crystal structures with mimetics	R^2 after MD (Vina-Carb/pi)	R^2 after MD (MM-GBSA)
DC-SIGN	13	1	0.52	0.07
Galectin-1	11	0	0.01	0.01
Galectin-3	12	3	0.01	0.01
FimH	7	8	0.55	0.72
FmlH	9	7	0.55	0.72
Siglec-2	42	0	0.15	0.14
Siglec-4	25	0	0.15	0.16
Siglec-7	22	1	0.35	0.38
Siglec-8	11	1	0.57	0.61
LecA	28	8	slope < 0	slope < 0
LecB	22	7	0.05	0.02
Cholera Toxin	11	7	0.15	0.28

No induced fit

Results

Protein	Number of Mimetics	Number of crystal structures with mimetics	R^2 after MD (Vina-Carb/pi)	R^2 after MD (MM-GBSA)
DC-SIGN	13	1	0.52	0.07
Galectin-1	11	0	0.01	0.01
Galectin-3	12	3	0.01	0.01
FimH	7	8	0.55	0.72
FmlH	9	7	0.55	0.72
Siglec-2	42	0	0.15	0.14
Siglec-4	25	0	0.15	0.16
Siglec-7	22	1	0.35	0.38
Siglec-8	11	1	0.57	0.61
LecA	28	8	slope < 0	slope < 0
LecB	22	7	0.05	0.02
Cholera Toxin	11	7	0.15	0.28

induced fit

Automated Virtual Glycomimetic Screening

- **Enables** the rapid, objective, standardized screening of relevant moieties
- **Facilitates** testing many scoring protocols (Vina-Carb, MM-GBSA)
- **Enables** the discovery of systemic problems
  - Galectins – missing force field terms (cation-π), induced side chain fit
  - Siglec-8 – induced fit in backbone
  - LecA/B – induced side chain fit
- **Benefits** from as much x–ray data and binding data as possible

Caveat 1 – the pdb is riddled with low quality structures for glycans

Agirre et al., (2015) *Nat. Chem. Biol.* **5**, 303

Caveat 2 – binding assays can give very different (1000x) K_D values

Ji Y, Woods RJ. (2018). *Adv Exp Med Biol*. **1104**, 259

Conclusions

Glycomimetic design is amenable to automation!

- Expect to see it at glycam.org in 2025

Predicted binding energies can (and need to) be improved

- Introduction of new physics in scoring functions
  - CH-π, cation-π
- Need to introduce new physics into AMBER force field for MD

Predicted binding poses can (and need to) be improved

- Induced fit in receptor, conserved waters

Need beta test users

- rwoods@ccrc.uga.edu

Acknowledgments

Underlying Science	Modeling Tool Development
Yao Xiao	Lachele Foley
Alex Lee	Dan Wentworth
Oliver C. Grant	Dave Montgomery
Lachele Foley	Oliver C. Grant

GLYCAM-Web: glycam.org

NIH National Institute of General Medical Sciences

NIH NATIONAL CANCER INSTITUTE