Automating the Rational Design of Glycomimetics

Robert J. Woods

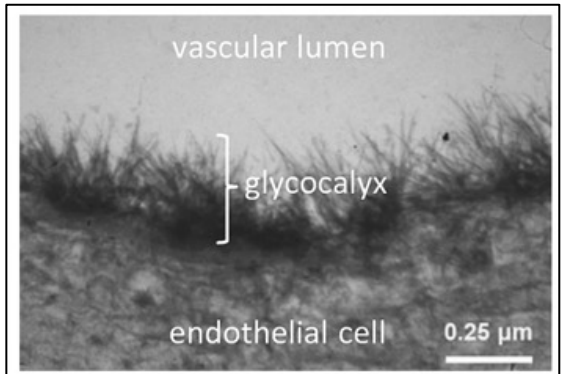
Complex Carbohydrate Research Center
University of Georgia

GLYCAM-Web: glycam.org

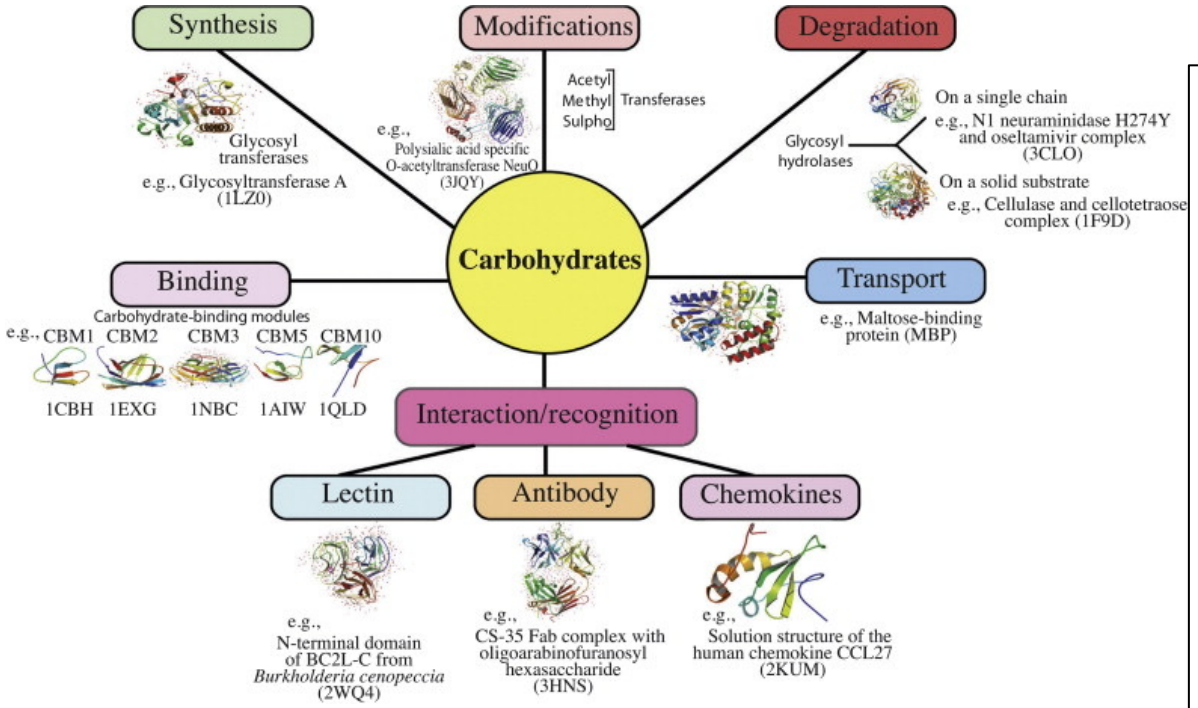
Carbohydrate Recognition in Human Health

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

Carbohydrate-protein interactions 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

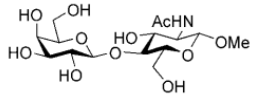
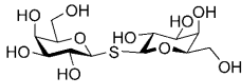
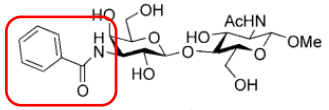
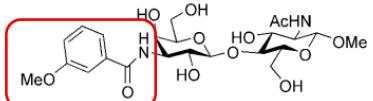
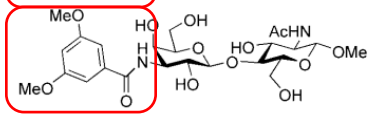
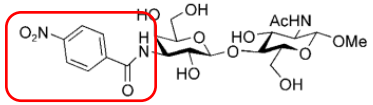
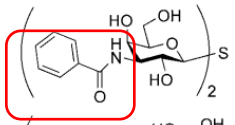
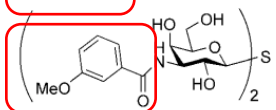
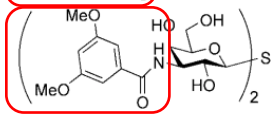
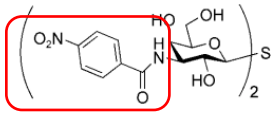
Glycan binding proteins (GBPs)

- Essential for 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

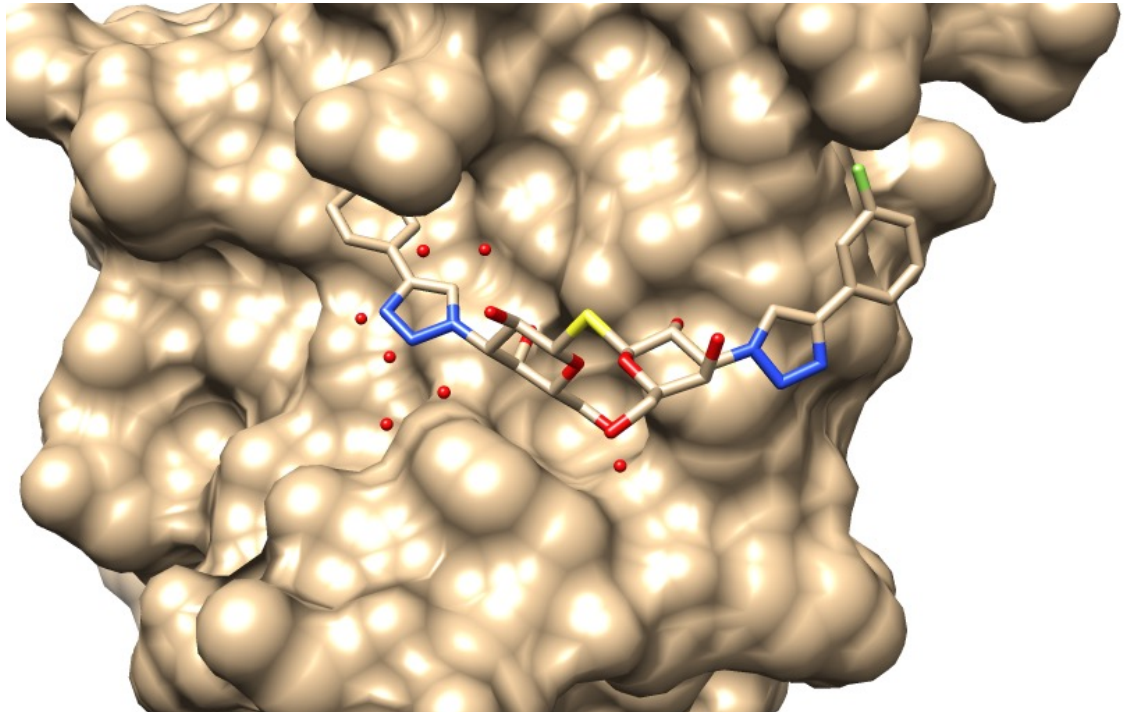
- Synthesize and degrade glycans
- Exploited by hosts and pathogens

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

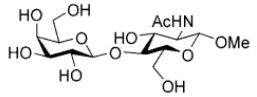
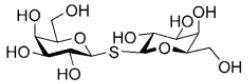
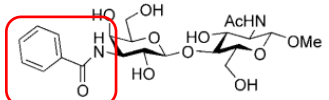
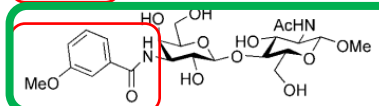
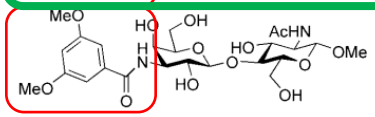
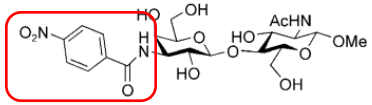
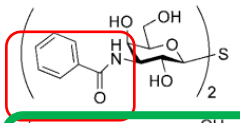
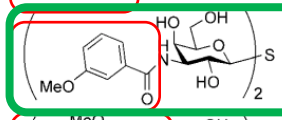
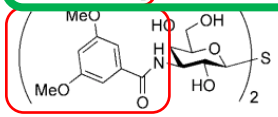
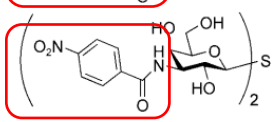
[a] Compounds **9**, **2a–d**, and **3** are included for reference.

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



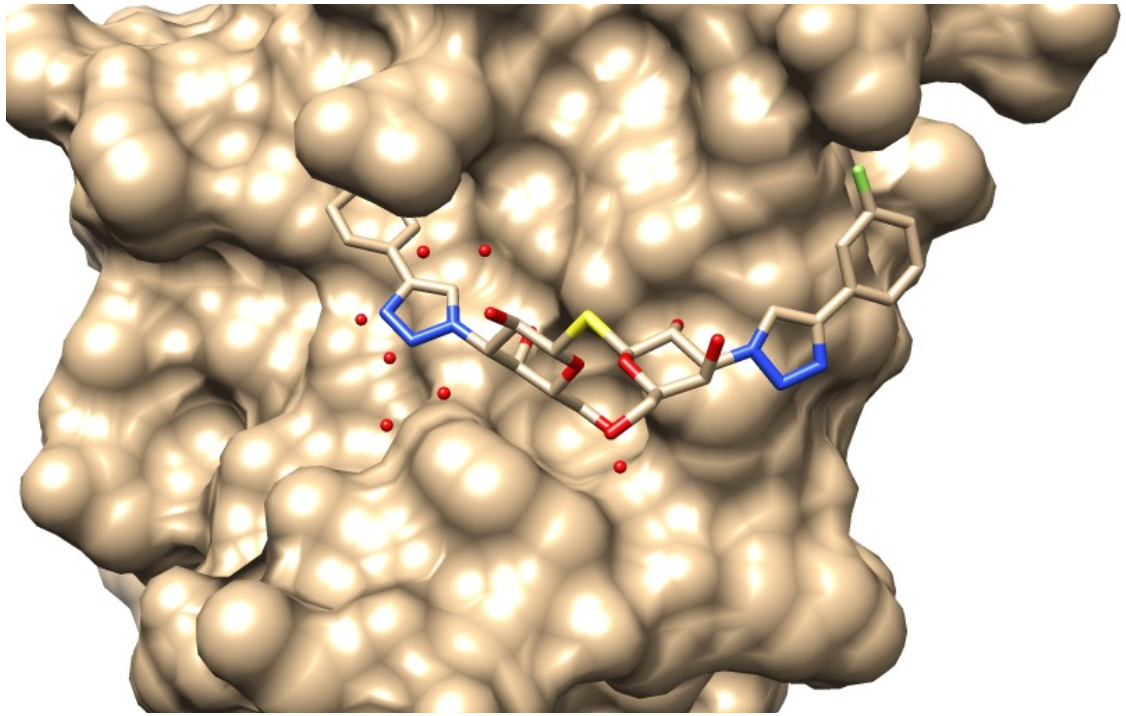
- Exploit the specificity of the endogenous carbohydrate ligand
- Employ the native carbohydrate ligand as a basis for rational design
- Examples: Relenza® and Tamiflu®
- Review: Magnani and Ernst (2009) *Discov. Med.* **8**, 247-252

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?

Benefits of Automated Virtual 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

Accessible to non-experts in computational chemistry

Enhanced user access to sophisticated modeling tools

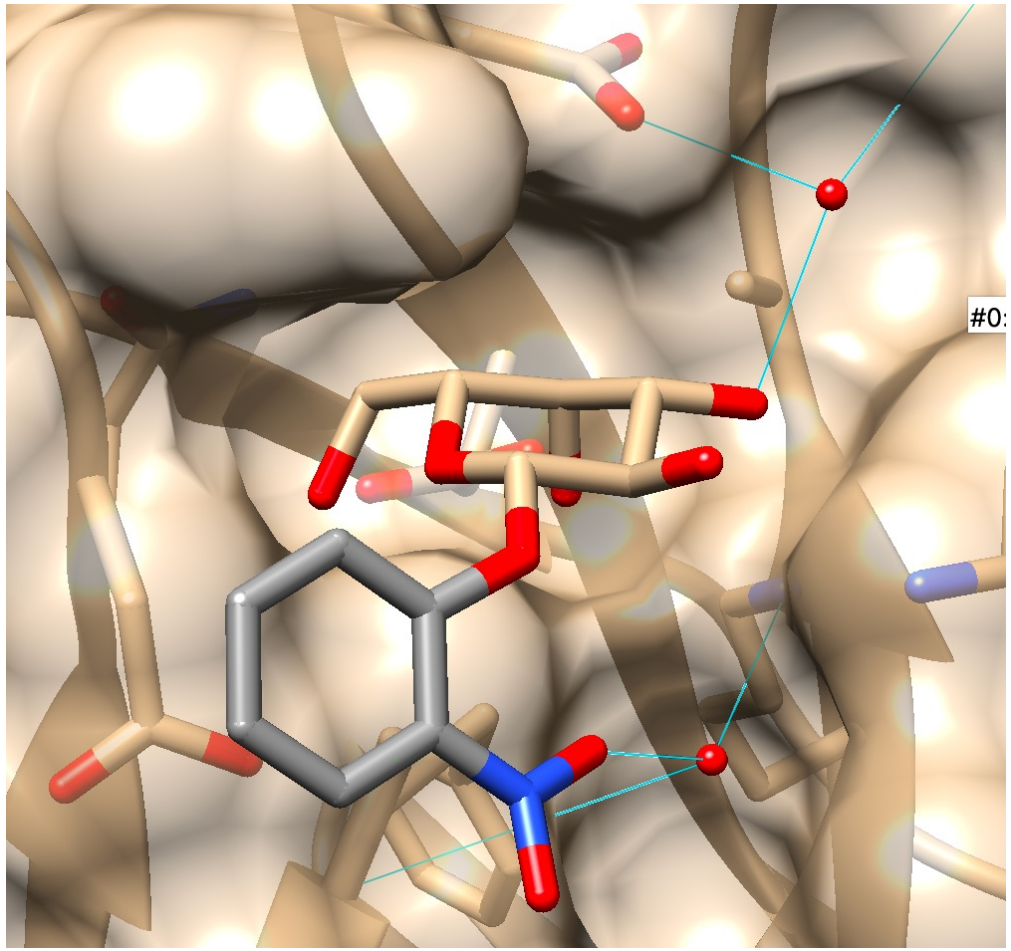
Agnostic – any anchored (co-crystalized) ligand could be modified

Motivation: Translate modeling technology into the experimental laboratory

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.)

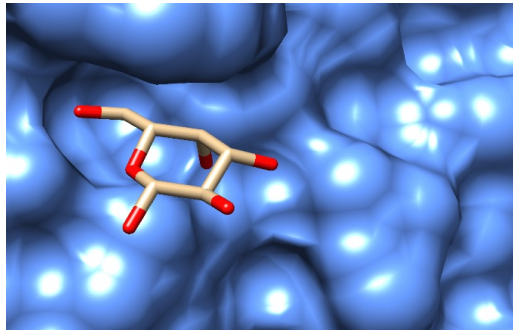
- **Our Project:**
  - Develop a high-throughput virtual screening pipeline for glycomimetic discovery
  - Automate this and create an online tool for glycomimetic screening
  - Apply it to Influenza and other disease targets

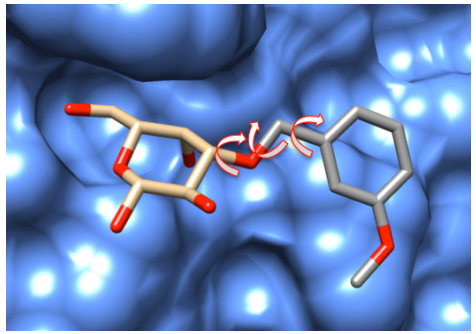


PDB 6AOY[5] (FmlH + ONPG)

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

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

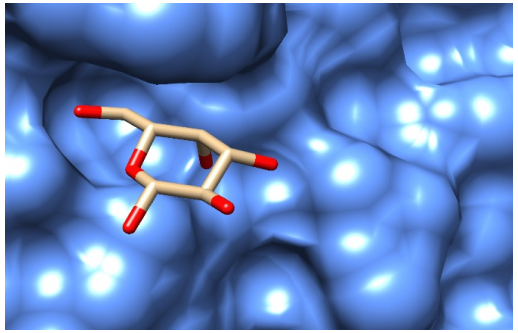
Perform MD of receptor-carbohydrate co-complex
- explicit water
- AMBER GAFF force field
- 50 ns

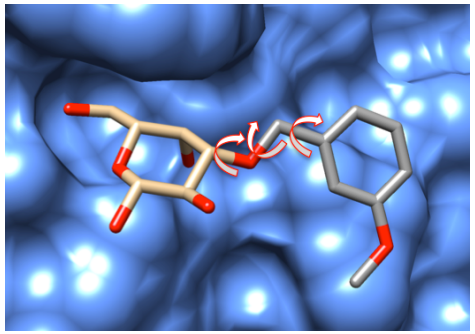
Calculate GIST desolvation energy and docking scoring for all rotamers of each analog in each MD snapshot

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

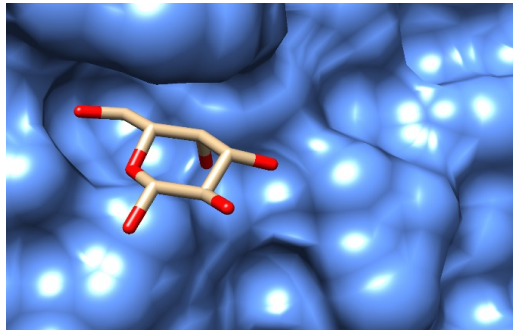
Calculate GIST desolvation energy and docking scoring for all rotamers of each analog in each MD snapshot

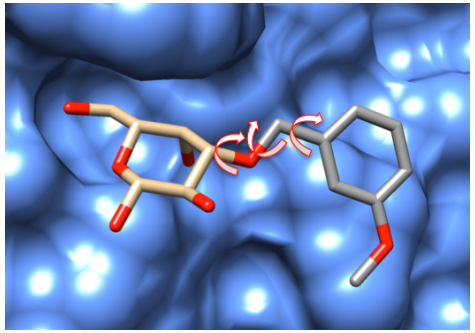
Compute ligand conformational entropy

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 ~1500 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

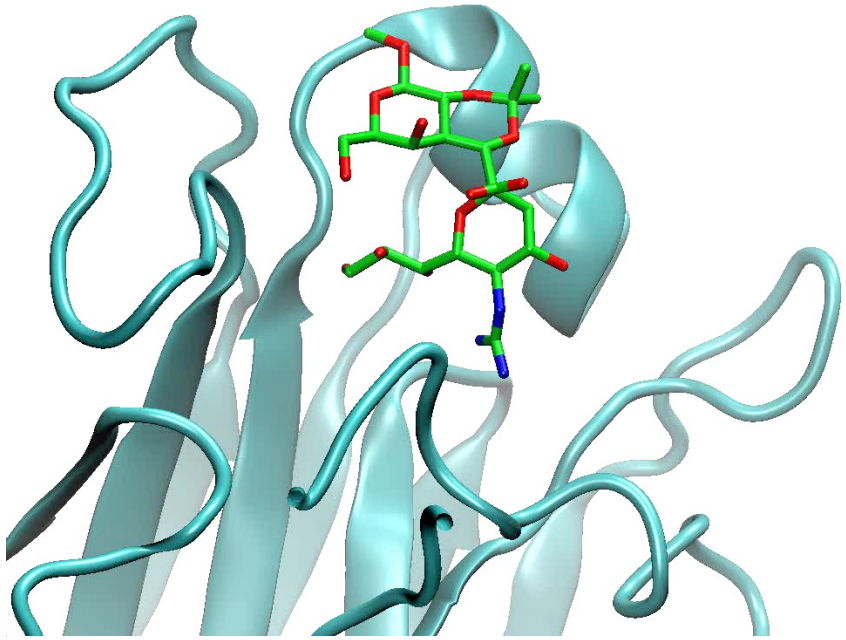
Compute ligand conformational entropy

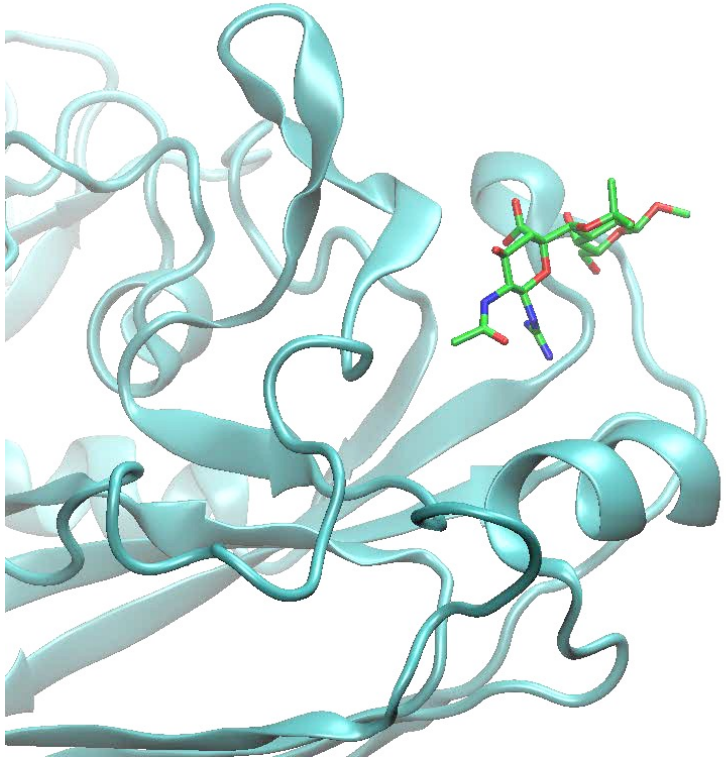
Return best pose and estimated binding free energy for each analog

Binding Energies:
- Multiple methods implemented
  - AutoDock VINA-Carb [1]
  - AMBER/GLYCAM MM-GBSA/PBSA

Molecular Dynamics Can Discriminate Strong from Weak

Which inhibitors should we simulate?





Putative Influenza Hemagglutinin Inhibitor 1	Putative Influenza Hemagglutinin Inhibitor 2

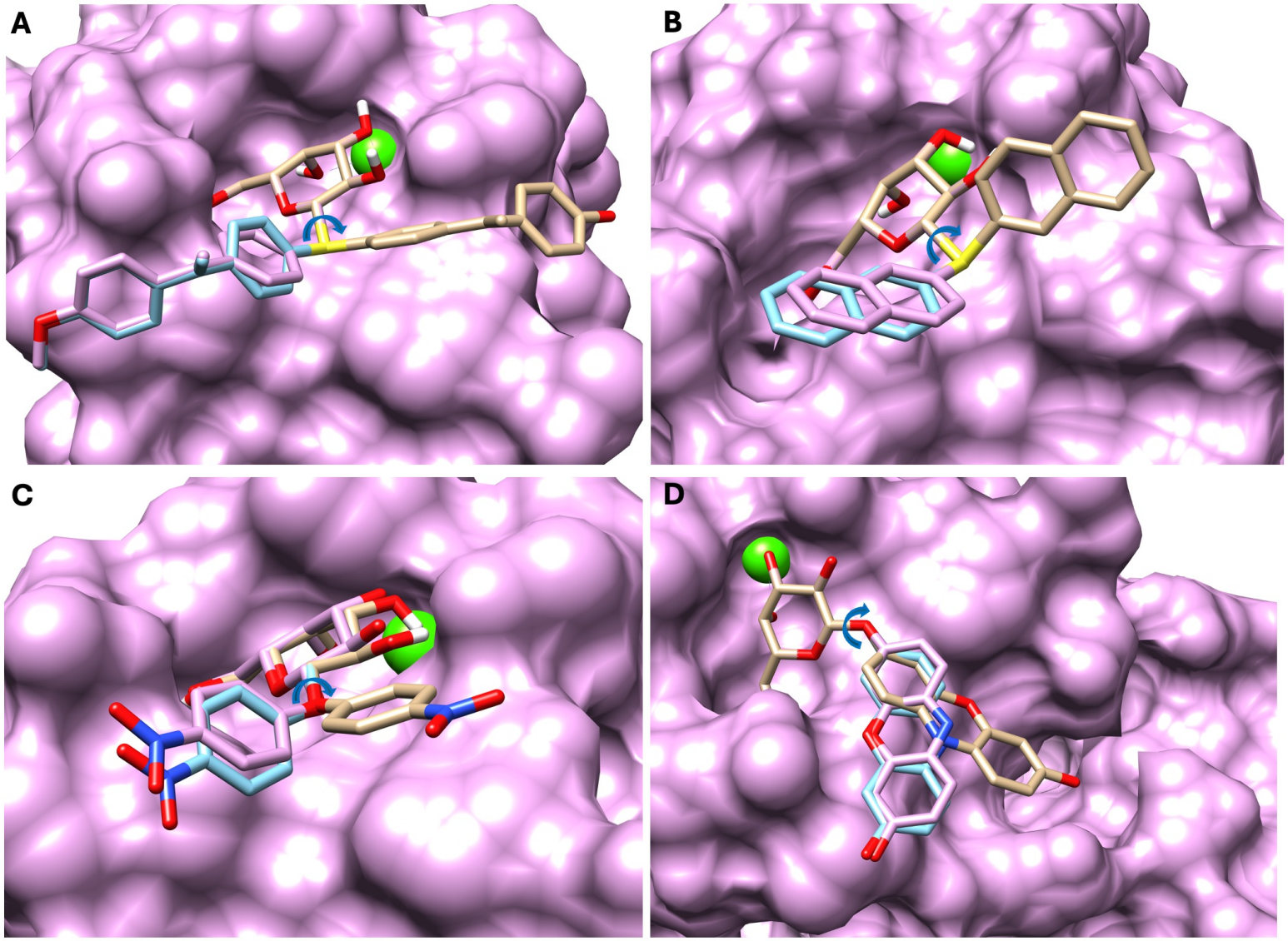
Test and Application Datasets

- Binding energies were computed from MD simulations of the glycomimetic complexes using AutoDock VINA-Carb, as well as the MM-PBSA/MM-GBSA) approximations
- VC and MM-GBSA/PBSA energies were augmented by ligand conformational entropies.
- Test data set of 58 glycomimetics with known co-crystal structures and binding energies – *includes the moiety in the crystal structure*
- Application dataset of 73 glycomimetics with reported solution affinities, employing known crystal structures that – *includes the carbohydrate, but not the moiety in the crystal structure*

Case Studies

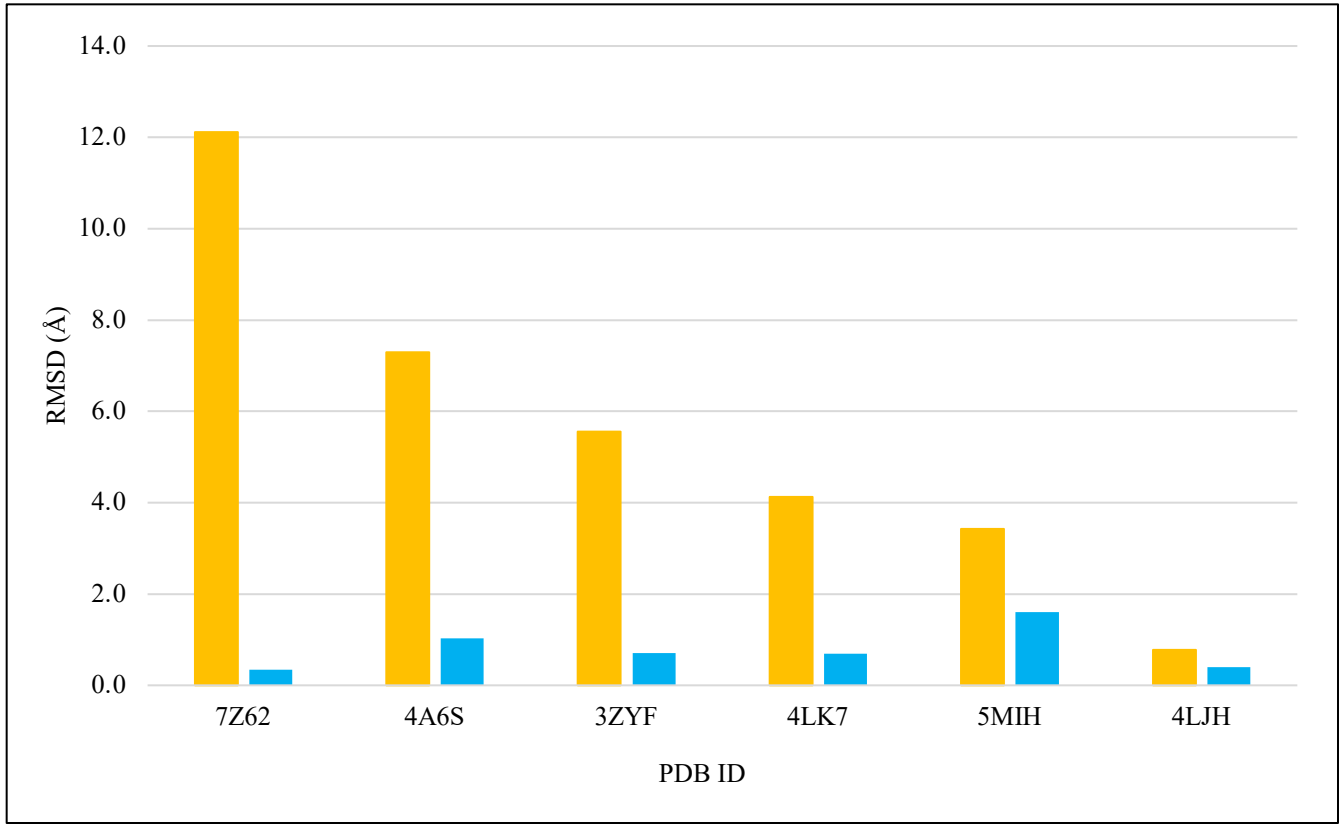
Carbohydrate Binding Protein	Endogenous Ligand	Function	Number of Reported Mimetics	Number of co-crystal structures
DC-SIGN	High-mannose N-glycans	Pathogen Recognition	13	1
Galectin-1	Beta-galactosides	Cell-cell / matrix interactions	11	0
Galectin-3	Beta-galactosides	Cell adhesion, growth, apoptosis, 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

LecA & LecB: Need for VC Scoring Function Correction



GA screening results for LecA systems **with (cyan) and without (beige) energy terms for the *exo*-anomeric effect** compared to the experimental structures (magenta). A, PDB ID: 7Z62. B, PDB ID: 4A6S. C, PDB ID: 3ZYF. D, PDB ID: 4LK7. **The anomeric ϕ torsion angle is indicated with a blue arrow.**

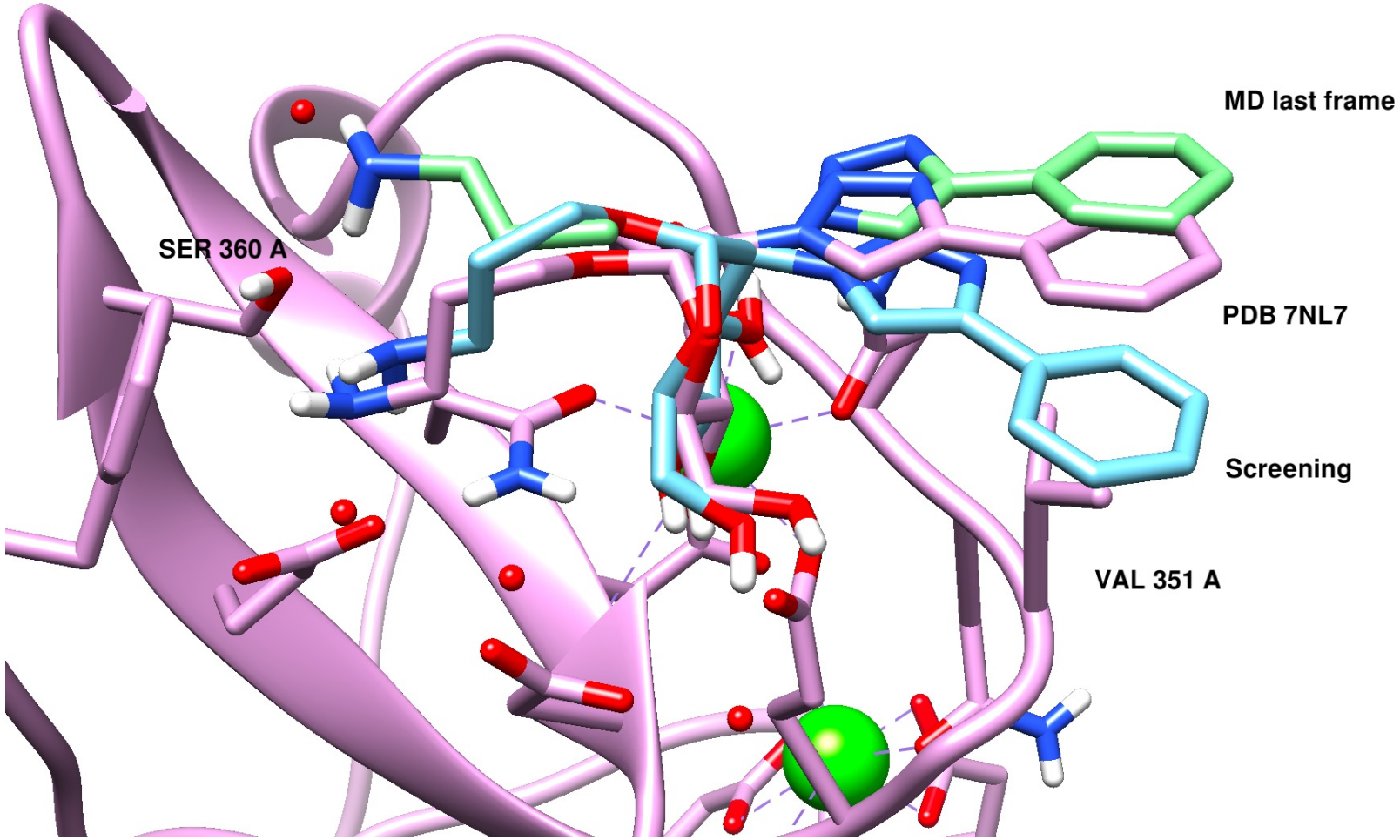
LecA & LecB: Need for VC Scoring Function Correction



RMSD values for the glycomimetic moieties from GA screening of LecA-glycomimetic complexes with (blue) and without (yellow) inclusion of torsion terms for the exo-anomeric effect in the ligand.

Success Example: DC-SIGN

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



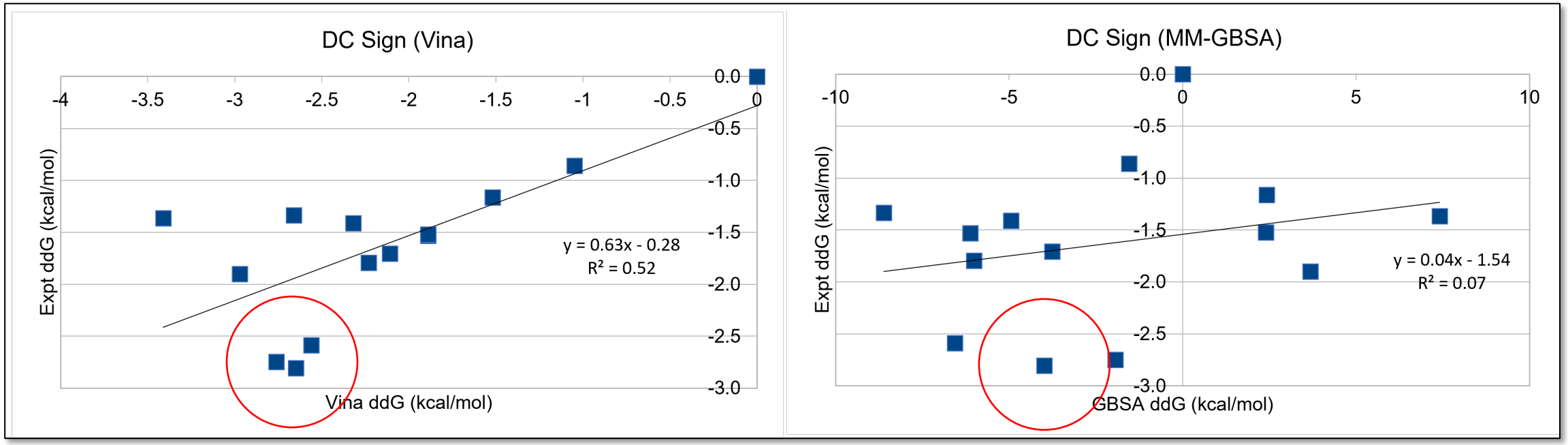
Initial computational screening reproduces crystal structure

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

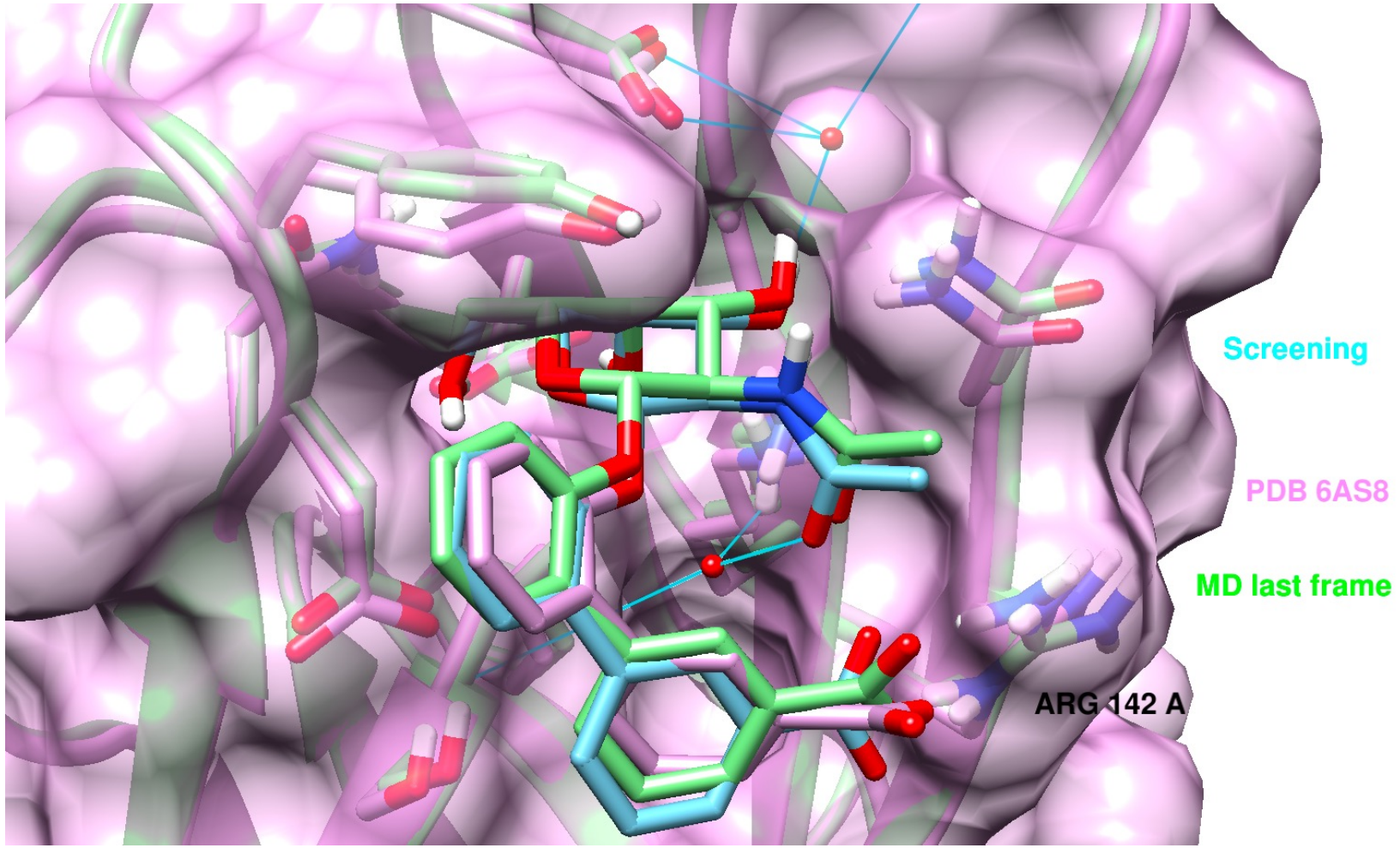


Vina-Carb: $R^2 = 0.52$

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

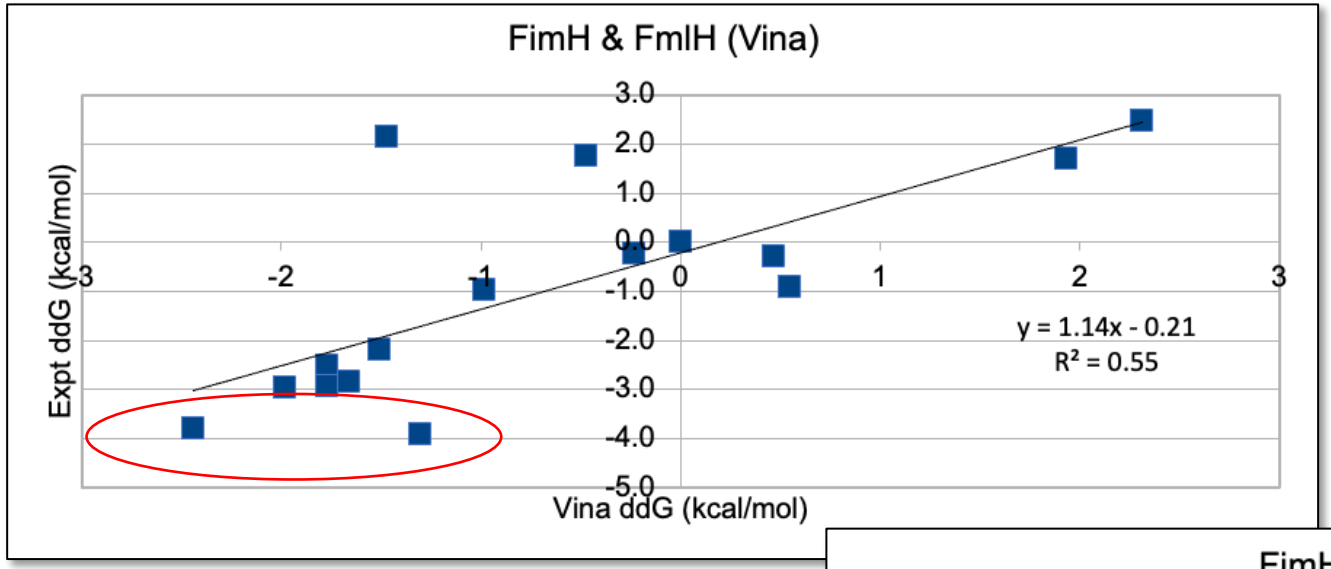


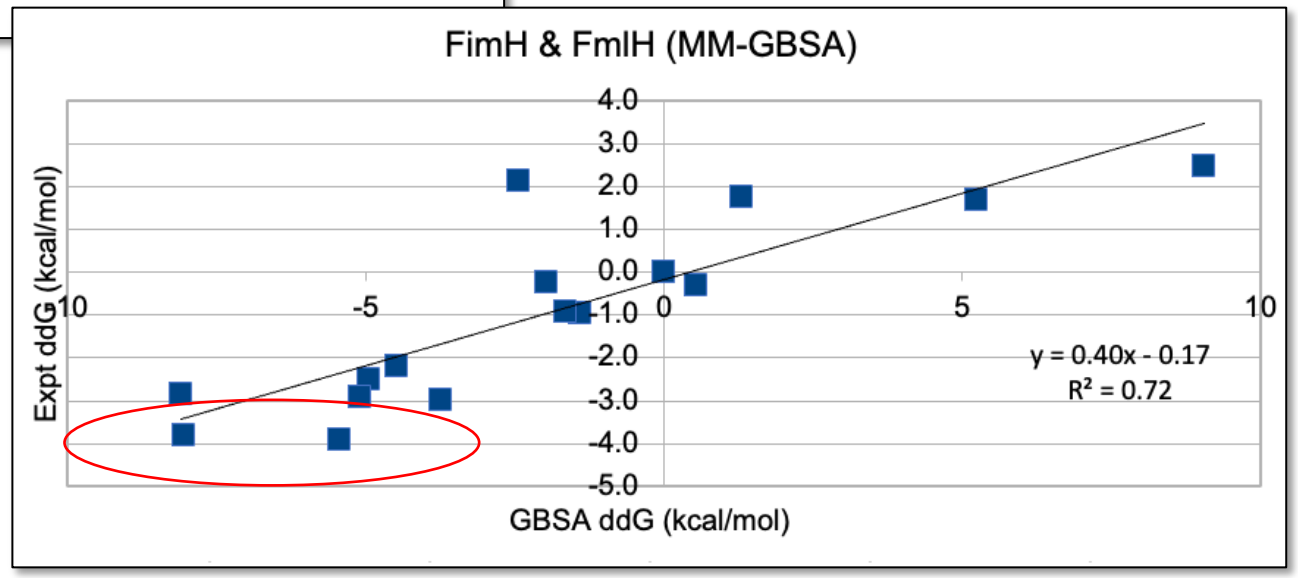
Initial computational screening reproduces crystal structure

MD simulation reproduces crystal structure

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

FimH & FmlH: Computed versus Experimental Affinity





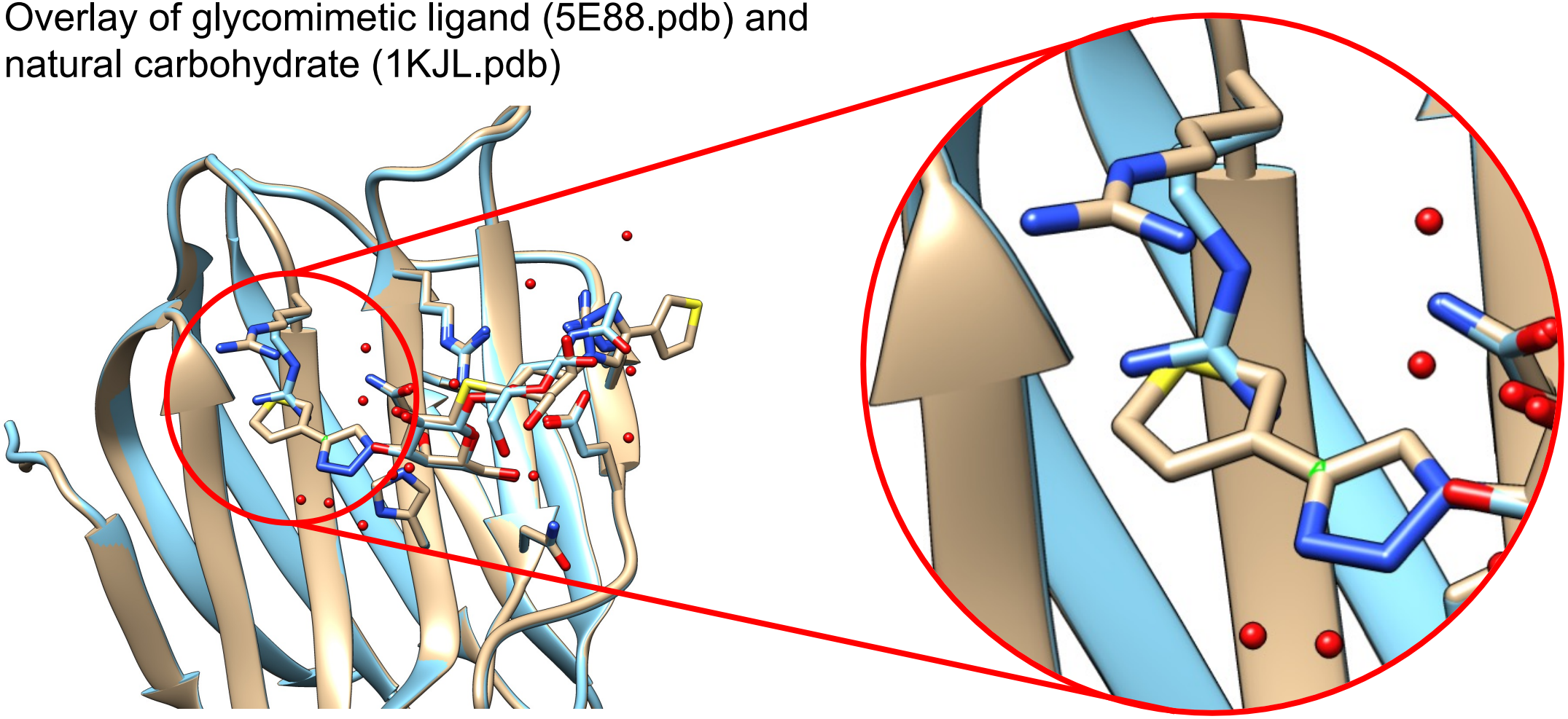
Vina-Carb: $R^2 = 0.55$

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.

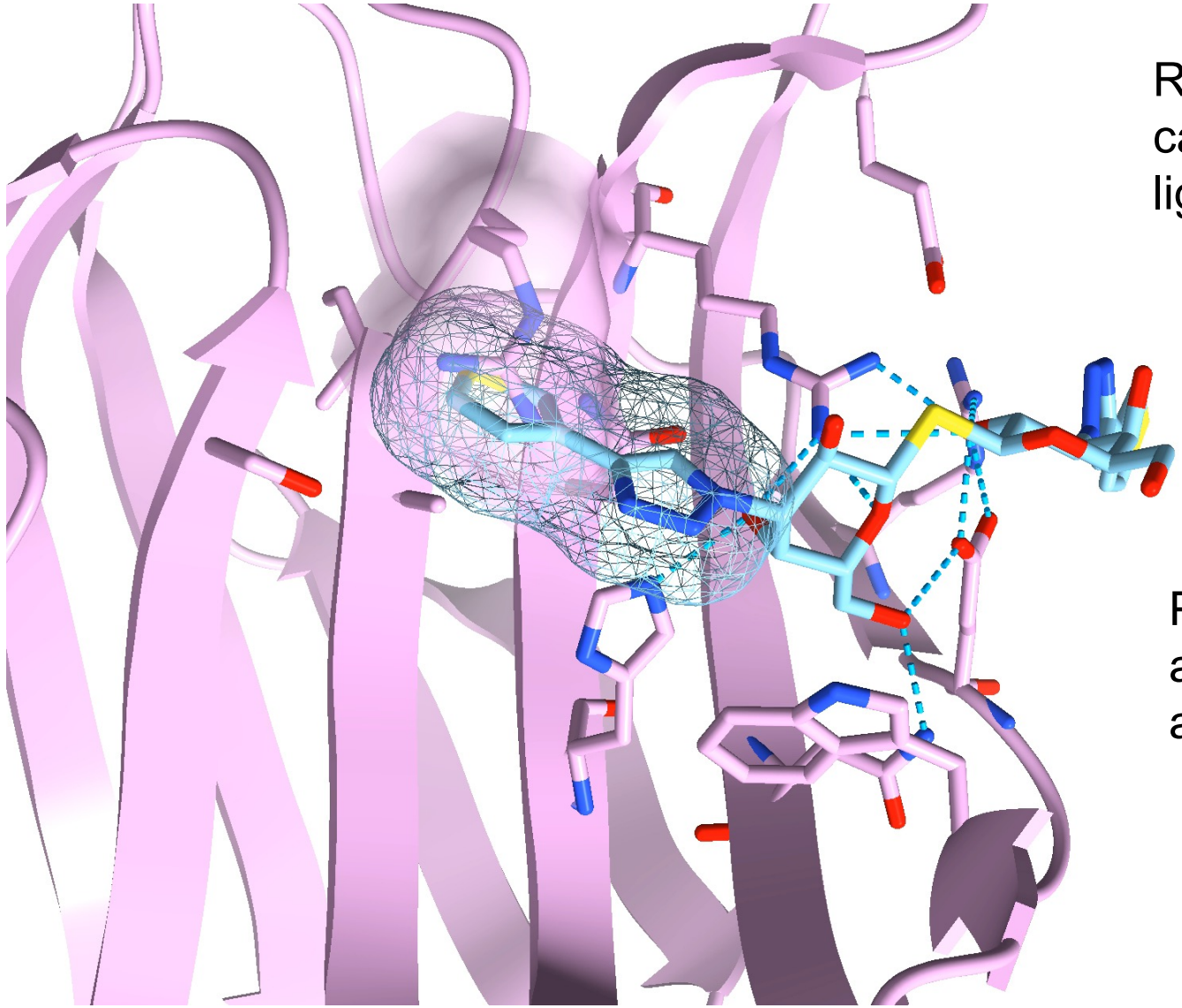
Problem Example: Galectin-3

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



Requirement for induced fit in ARG 144 causes prediction error.

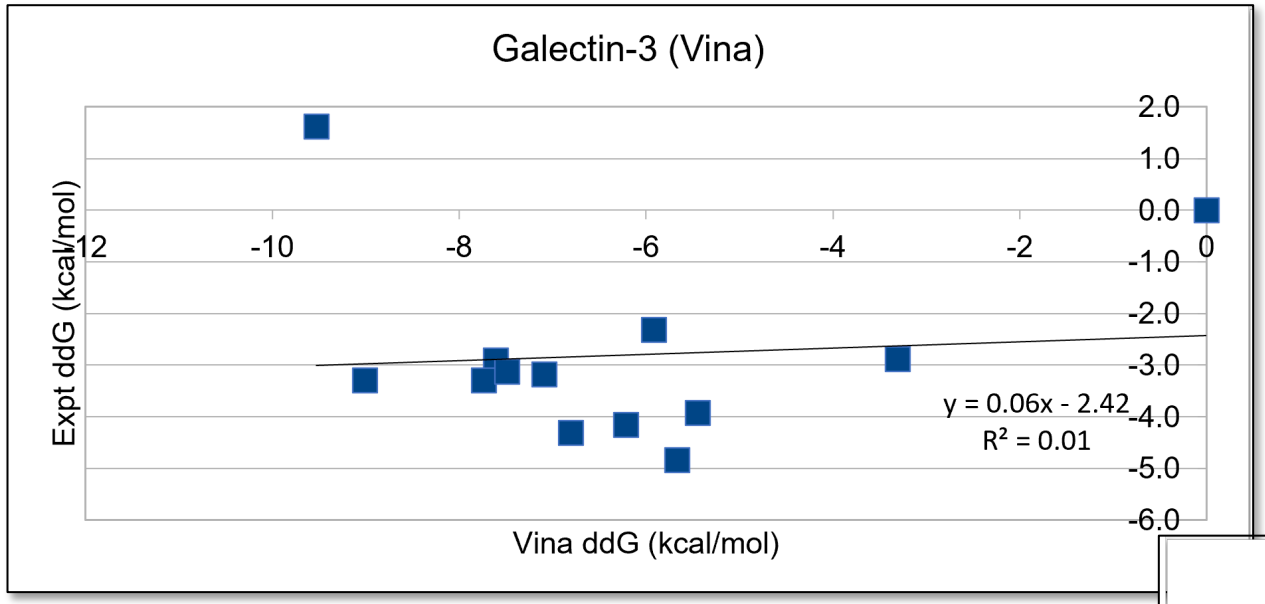
Problem Example: Galectin-3



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

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

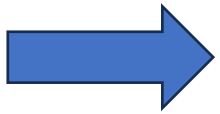
Problem Example: Galectin-3



Vina-Carb: $R^2 = 0.01$

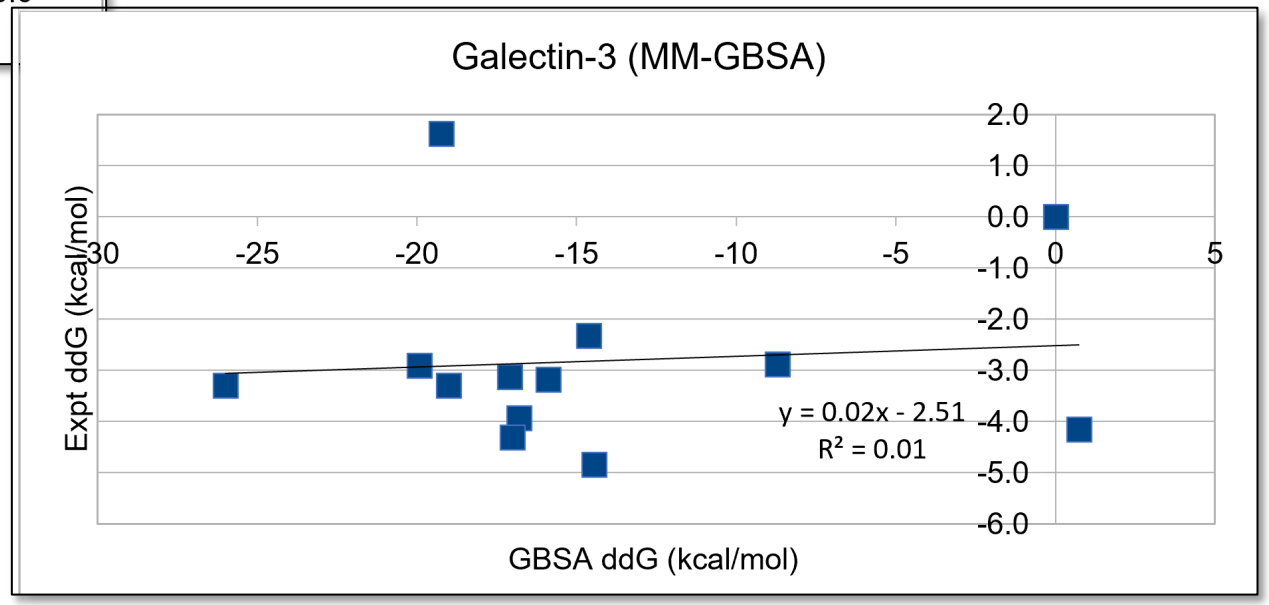
MM-GBSA: $R^2 = 0.01$

Incorrect Starting Geometry

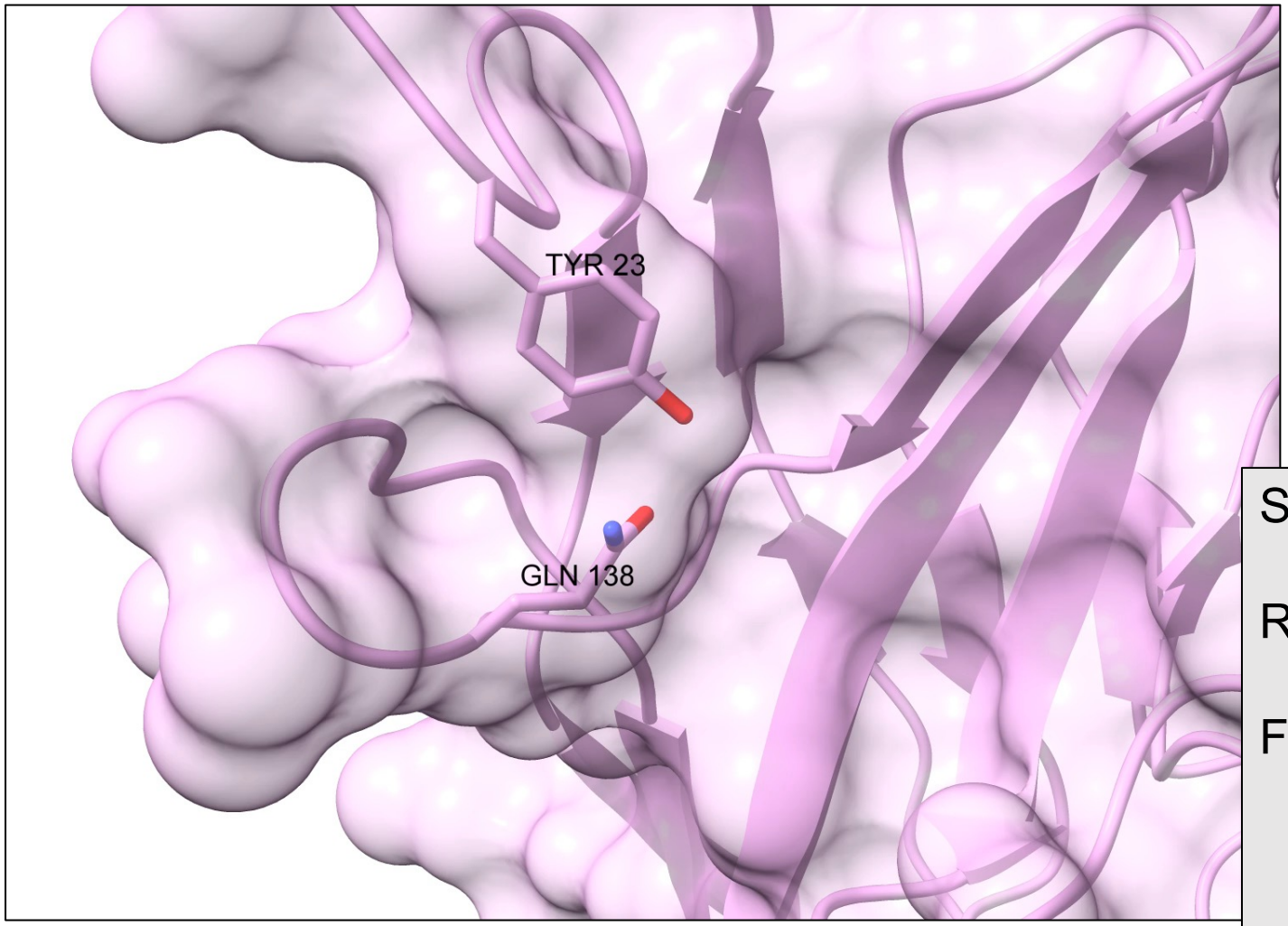


"GIGO" Principle

Erroneous Energies

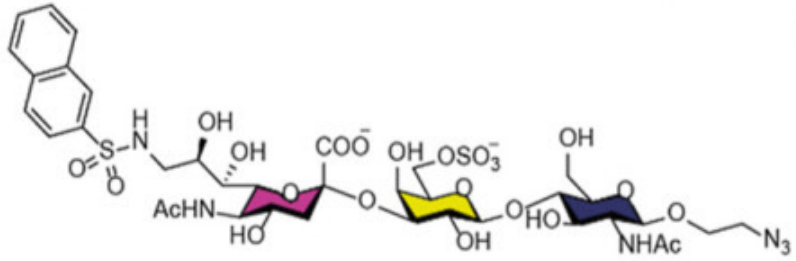


The Problem of Induced Fit in the Backbone: Siglec-8



Apo Protein: 7qu6.pdb

Co-Crystal with Glycomimetic: 7qui.pdb



Screening Protocol:

Rigid Protein
- Typical

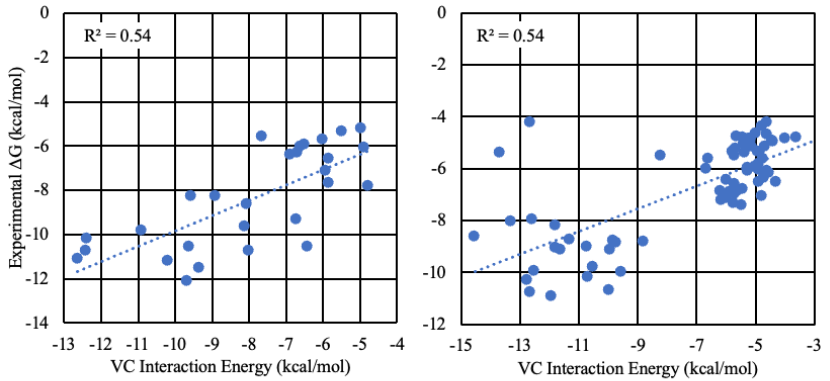
Flexible Protein
- Side chains: Employ a rotamer library, Backbone: changes in protein fold are highly problematic for docking

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

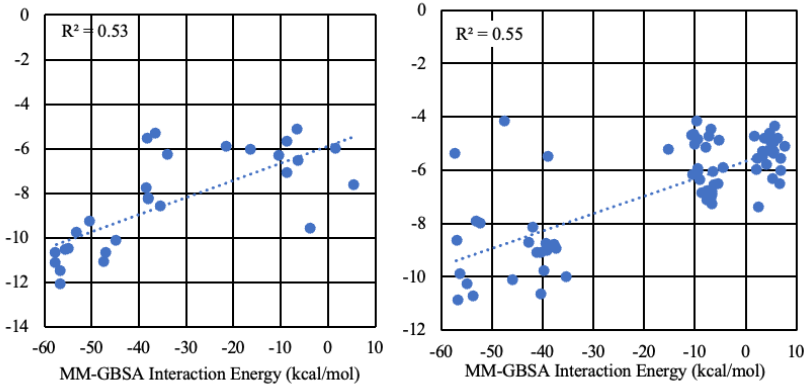
Overall Performance

Test Dataset | Application Dataset

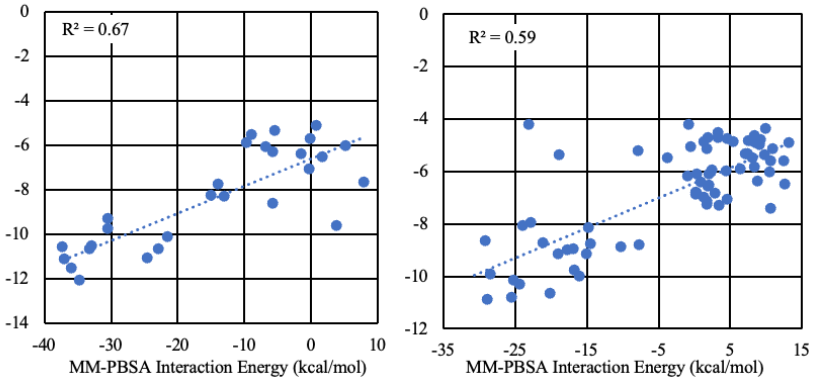
Post-MD VC



Post-MD MM-GBSA



Post-MD MM-PBSA



Average theoretical interaction energies from MD data versus experimental values.

Left column, data from the Test dataset for moieties with RMSD values < 2 Å (n = 28)

All theoretical values include conformational entropy.

Summary of Performance

Correlation coefficients (R^2) for the Test dataset from VC, MM-GBSA and MM-PBSA analyses post-MD or pre-MD with energy minimization only (**with/without ligand conformational entropies**).

Method	MD RMSD < 2 Å	MD RMSD 2–3 Å	MD RMSD > 3 Å	MD All Poses	EMin GA	EMin X-tal
MM-PBSA	0.67/0.62	0.32/0.28	0.00/0.03	0.45/0.41	0.22/0.18	0.26/0.21
MM-GBSA	0.53/0.49	0.35/0.32	0.08/0.14	0.35/0.32	0.27/0.22	0.28/0.23
VC	0.54/0.40	0.37/0.23	0.12/0.11	0.42/0.26	0.18/0.08	0.17/0.08

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	Grayson Miller
Lachele Foley	Oliver C. Grant

Xiao et al., *J. Chem. Info. Model.*, **2025**, In Press.

NIH National Institute of General Medical Sciences

NIH NATIONAL CANCER INSTITUTE