Supplementary Materials for

# CH-π interactions are required for human galectin-3 function

**Authors:** Roger C. Diehl, [‡,1] Rajeev S. Chorghade, [‡,1] Allison M. Keys,[2] Mohammad Murshid Alam,[1], Stephen A. Early,[3] Amanda E. Dugan,[1] Miri Krupkin,[4] Katharina Ribbeck,[4] Heather J. Kulik,[1,5] and Laura L. Kiessling[*1,6,7]

[‡]These authors contributed equally to this work

*To whom correspondence should be addressed: kiesslin@mit.edu

**Affiliations**:
[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2]Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

[4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[7]Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA 02142, USA

**This PDF file includes:**

## MATERIALS AND METHODS:

### Cloning and site directed mutagenesis

The coding sequence of human galectin-3 (Uniprot ID P17931) residues 113-250, a truncated construct known as galectin-3C (Gal3C), was obtained from NCBI (Gene ID: 3958). This sequence was ordered as a gene product from Integrated DNA Technologies (Coralville, IA) and inserted into a pET24a vector (Novagen, Madison, WI) by the Gibson assembly method.[1] This vector was amplified in Escherichia coli DH5α cells, then used to transform DE3 Tuner cells for protein expression. To add a noncleavable hexahistidine ($His_6$) tag and generate variants at W181, we used an inverse PCR based strategy. Primers were designed to amplify the entire construct, with the desired modification included as a 5' overhang. After completion of a PCR reaction on the wild type plasmid using these primers, the product was circularized with T4 polynucleotide kinase and T4 DNA ligase, both obtained from New England Biolabs (Ipswich, MA). The ligation reaction product was digested with the restriction enzyme DpnI to remove the template plasmid, and then the cleaned product was electroporated into DH5α cells. Mutants were verified by Sanger DNA sequencing. Galectin-3C used for mucin experiments was cloned with an N-terminal $His_6$ tag for purification and an N-terminal SUMO tag for improved expression. For hemagglutination experiments, we cloned a full-length pET24a-galectin-3 construct bearing an N-terminal hexahistidine tag, with E. coli codon optimization on residues 1-112.

### Protein expression and purification

Galectin-3C, full-length galectin-3, and all variants were expressed in *Escherichia coli* DE3 from a pET-24a vector. Liter-scale cultures were grown to mid-log phase ($OD_{600}$=0.3-0.5) and induced

with 0.4 mM isopropyl β-thiogalactoside (IPTG). Cultures were then grown for 4 hours at 37 °C and pelleted by centrifugation (20 min at 3000xg). After storage at -80 °C, pellets were resuspended in Ni-NTA Loading Buffer (20 mM sodium phosphate, 500 mM sodium chloride, 20 mM imidazole, pH 7.4), lysed by sonication (3 x 40 seconds) or French press (two cycles, maximum pressure 1500 bar), and cell debris was removed by centrifugation (30 min at 33,000 xG). Filtered supernatants were then loaded onto a Ni-NTA affinity column (BioRad, Hercules, CA), washed in Ni-NTA Loading Buffer, and then eluted on a gradient to Ni-NTA Elution Buffer (20 mM sodium phosphate, 500 mM sodium chloride, 500 mM imidazole, pH 7.4). Fractions were analyzed on a tris-glycine gel stained with Coomassie Brilliant Blue, and pooled and injected onto a Superdex 75 size exclusion column (GE Healthcare) in Galectin-3 Assay Buffer (60 mM sodium phosphate, 8 mM potassium phosphate, 390 mM sodium chloride, 5.5 mM potassium chloride, pH 7.4).

**Differential scanning fluorimetry**

Differential scanning fluorimetry assays were conducted in Galectin-3 Assay Buffer. Experiments were conducted with 10 μM Gal3C or a variant thereof, and 15x Sypro Orange dye (Thermo Fisher Scientific, Waltham, MA). Samples were heated at a rate of 1.5 °C/min from 15-95°C in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA), and the melting temperature ($T_m$) of the protein was measured using a Boltzmann analysis. Measurements were repeated in triplicate with separate stocks of protein to verify reproducibility of the results. For variants where SUMO tags were used for ITC experiments, both tagged and untagged constructs were tested in the DSF assay.

**Isothermal titration calorimetry**

Binding of Gal3C and variants to lactose was evaluated by isothermal titration calorimetry (ITC). A 10-50 mM solution of lactose was titrated into a 0.05-1 mM solution of galectin-3C on a MicroCal VP-ITC or a MicroCal Auto-iTC200 calorimeter (Malvern Panalytical, Malvern, UK) with a 1.4 mL cell volume with 2 μL, 5μL, and 10 μL injection sizes until the titrant comprised 15% of the total volume. We used a ligand-buffer titration to normalize measured binding enthalpies. Galectin-3 Assay Buffer was used for both protein and ligands to ensure that all variants remained in their native state. Experiments were conducted in triplicate with separate preparations of protein to ensure reproducibility.

**Bio-layer interferometry**

Binding of Gal3C and variants to LacNAc was determined using a Bio-Layer Interferometry instrument (ForteBio). Biotin-LacNAc was diluted to 5 μM in 1X PBS, pH 7.5, and loaded (120 s) onto Streptavidin biosensors (Forte Bio). The sensor was washed in 1X PBS for 60 s, and then baseline (60 s) established in (20 mM HEPES, pH 7.4, 150 mM sodium chloride, 0.1% Tween-20, 0.1% BSA). Wild type Gal3C or variant was diluted to 250, 100, 10, 1, and 0.1 ug/mL in (20 mM HEPES, pH 7.4, 150 mM sodium chloride, 0.1% Tween-20, 0.1% BSA) and allowed to associate with the biotin-LacNAc coated biosensor (450 s). Dissociation (450 s) was monitored by dipping biosensor in (20 mM HEPES, pH 7.4, 150 mM sodium chloride, 0.1% Tween-20, 0.1% BSA). Biosensors were regenerated between samples by triple-washing biosensors in 10 mM glycine pH 2.5 (5 s) followed by neutralization (5 s) in (20 mM HEPES, pH 7.4, 150 mM sodium chloride, 0.1% Tween-20, 0.1% BSA). The shake rate was set to 1000 rpm and binding assays were performed at room temperature. Data was plotted in Prism version 9.0.0 (GraphPad).

**Computational analysis of galectin-3 and variants**

We obtained 3D coordinates for wild type Galectin-3 from the Protein Data Bank (PDB)[2], PDB ID: 3ZSJ.[3] Three variants, each with a single amino acid mutation (E184A, N174A, H158A, and W181M) were generated using the PyMol v. 2.5.2 protein mutagenesis wizard. Protonation states of the wildtype and three variant structures were assigned using the H++ web server[4] with an internal dielectric constant of 10.0, a pH of 7.0, and all other default settings applied. The resulting total protein charge of the wildtype and H158A and W181M variants was +5 and the final charge of the E184A variant was +6. The only charge difference between the structures was due to the presence or absence of the negatively charged E184 residue (Table S1). The residue labels for the lactose ligand were manually updated to match the GLYCAM_06[5] force field naming conventions. The AMBER LEaP program[6] was used to solvate the protein in a 10 Å TIP3P[7] water box, neutralize the system using $Cl^-$ ions, and generate final topology and coordinate files for the system using the appropriate forcefields, i.e., ff14SB[8] for the protein, GLYCAM_06j-1[5] for the carbohydrate ligand, and TIP3P[7] for water. The final system sizes were 23,302 atoms for the wildtype, 23,295 for the E184A and H158A variants, and 23,298 for the W181M variant.

Molecular dynamics (MD) simulations were performed with AMBER18[6] using the GPU-accelerated particle mesh Ewald (PME)[9] molecular dynamics (PMEMD) module[10]. Each system was equilibrated using the following procedure. First, two minimizations were carried out: 1) a 2000 cycle minimization of the water molecules and hydrogen atoms, with a 200 kcal $mol^{-1}$ $Å^{-2}$ weight restraint applied to protein and ligand heavy atoms and 2) a 2000 cycle minimization of all atoms. Next, a series of equilibration steps were run using 2 fs timestep along with a

Langevin thermostat for temperature control with a collision frequency of 5.0 $ps^{-1}$, a 2 fs timestep, the SHAKE algorithm for constraining hydrogen atom bond lengths, periodic boundary conditions, and the PMEMD code.[10] These equilibration steps consisted of: a 10 ps controlled *NVT* equilibration to heat the system from 100K to 300K followed by a 1 ns *NPT* equilibration. Next, a 100 ns *NPT* equilibration step was carried out with free movement of all atoms except for the heavy atoms of the binding pocket residues, H158, R162, N174, W181, E184, and the lactose ligand, which have 200 kcal $mol^{-1}$ $Å^{-2}$ weight Cartesian coordinate restraints. In contrast, the second minimization step and first two equilibration steps permit unrestrained movement of all heavy atoms except the tryptophan residue 181 and the lactose ligand, which have 200 kcal $mol^{-1}$ $Å^{-2}$ weight Cartesian coordinate restraints applied. Finally, a 200 ns production run was performed to observe the dynamics of each system. This full MD procedure was carried out three times for each protein variant.

The distance between the center of mass of the lactose ligand and the center of mass of the binding pocket residues H158, R162, N174, W181, and E184 was computed for all snapshots and averaged over intervals of 1 ns (Figure 4, Figures S4-S5). All snapshot intervals corresponding to at least 1 ns with an averaged binding distance greater than 6.25 Å were removed from the production simulations as the lactose had dissociated from the binding pocket under these circumstances (Table S2). For each variant, the remaining snapshots from each of the three production runs were spliced together using the AMBER CPPTRAJ trajectory processing program[11] to form a single trajectory of galectin-3 bound to lactose. These bound trajectories were clustered using the k-means algorithm implemented in CPPTRAJ[11] to identify the major binding conformations observed by each variant and the wildtype.

Binding affinity values evaluated with classical interaction analysis were determined for each variant using evenly spaced snapshots taken every 200 - 400 ps, ensuring at least 1400-1500 snapshots were analyzed for each variant (Table S2). Calculations were performed using the molecular mechanics with generalized born and surface area solvation (MM-GBSA)[12] with a salt concentration of 0.1 using the AMBER MMPBSA.py script.[13] Residue-based energy decompositions were used to evaluate the binding energy contribution from each residue. Total entropy contributions were evaluated using normal mode analysis on one out of every 10 snapshots used for classical interaction analysis. The snapshots were evenly spaced and taken every 2 – 4 ns while the ligand was bound, ensuring at least 140-150 snapshots were analyzed for each variant, using normal mode analysis, with a maximum number of cycles=10000.

**Mucin binding assay**

Lyophilized MUCC2, MUC5ac, and MUC5b were provided by Prof. Katharina Ribbeck, whose group purified them from porcine sources as previously described.[14] These mucins were reconstituted to 10 μg/μL overnight at 4 °C in Mucin Assay Buffer (20 mM HEPES, 150 mM sodium chloride, pH 7.4) as previously described[15], then diluted to a concentration of 1 μg/μL in Mucin Assay Buffer. 1 μL of this mucin solution was then spotted onto activated PVDF membranes and allowed to air dry at room temperature for 1 hour. The membranes were then blocked with 5% bovine serum albumin (BSA) and 0.1% Tween 20 in Mucin Assay Buffer overnight at 4 °C. The membranes were then incubated for 2 hours at room temperature with 10 μg/mL galectin-3C (wild type or variant) bearing a C-terminal Strep-tag in Mucin Assay Buffer with 0.1% BSA and 0.05% Tween 20. For each variant, one membrane was incubated with the

above solution plus 50 mM lactose, while the other was incubated with a solution not containing lactose.

After incubation and washing, the blots were incubated for 1 hour at room temperature with 1:5000 anti-$His_6$-HRP antibody (Thermo Fisher Scientific MA121315HRP) and developed for 5 minutes with Clarity ECL reagent (BioRad). Binding was visualized using chemiluminescence on a ChemiDoc imager (BioRad). The intensity of the chemiluminescence signal for each spot was quantified using the Fiji image suite, expressing the volume of each spot as a percentage of the highest intensity spot.

**Hemagglutination microscopy assay**

Blood was collected from C57bl/6 mice through retro-orbital bleeding. The blood was heparinized, and then washed three times with PBS by centrifuging 5 min at 100xg after each resuspension. A 6.25% suspension of the final pellet of mouse red blood cell (RBC) pellet was prepared in PBS, and 1 µL of this suspension was treated with a 9 µL solution of wild type or variant full-length galectin-3 in PBS on a glass sample dish, to afford a 160x final dilution. The dish was covered and incubated 30 minutes at room temperature, after which 20x magnification images were taken with a Nikon A1R confocal microscope (Tokyo, Japan). Percentage agglutination was quantified by manually counting single and agglutinated RBCs in the central 317x317 µm region of the drop.

**Hemagglutination plate assay**

Blood collected and processed as above was diluted to 1%, and 50 µL was applied to U-bottomed 96-well plate containing 50 µL of serial dilutions of 400 µg/mL protein as measured by

absorbance at 280 nm. Protein used in this assay was filtered through a 100 kDa molecular weight cutoff centrifugal filter immediately before use. Wells containing 50 μL PBS and 50 μL 1% RBC suspension were used as controls. Plates were incubated overnight at room temperature, and then wells were scored visually. A uniform suspension in the well was scored as a positive reading (agglutination) and any sign of a pellet on the bottom of the well was scored as a negative reading (no agglutination). The lowest concentration of protein sufficient to cause agglutination was recorded.

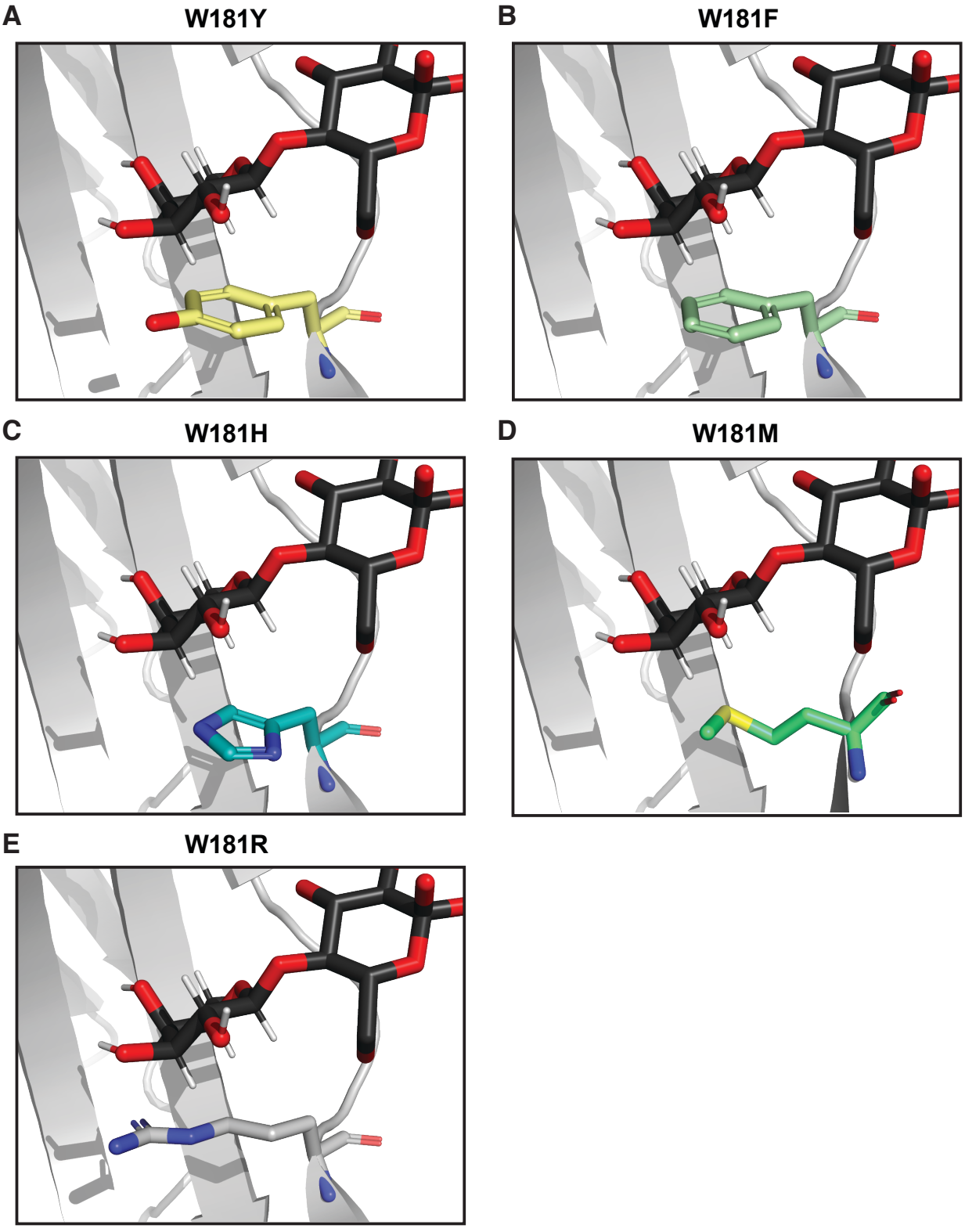


Figure S1. Close-up views of the modeled complexes of β-galactose and residue 181 in the following variants of galectin-3: (A) W181Y, (B) W181F, (C) W181H, (D) W181M, (E) W181R.

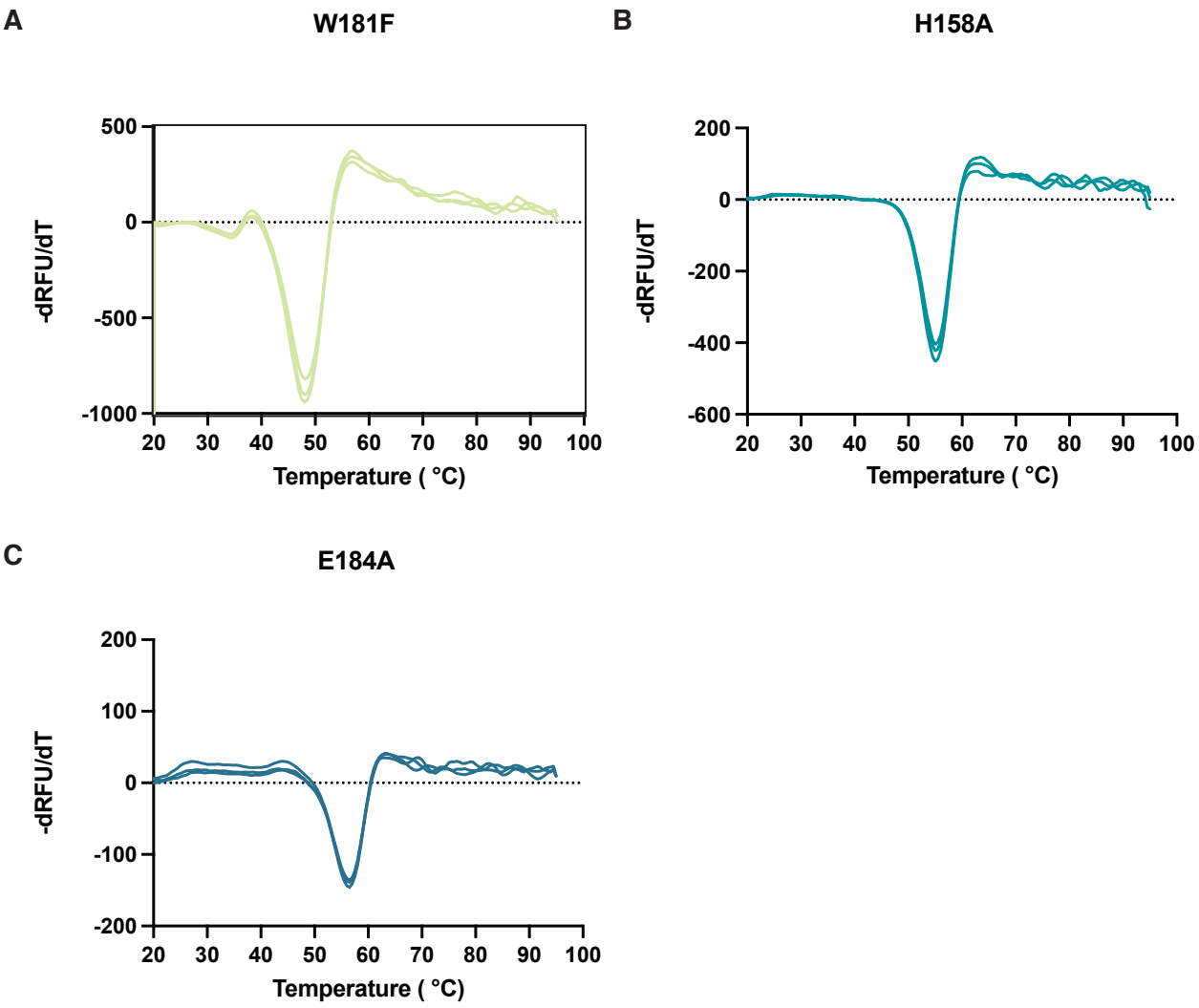


Figure S2. Raw DSF traces of select Gal3C variants. (A) W181F, (B) H158A, (C) E184A. All DSF experiments were performed in biological triplicate.

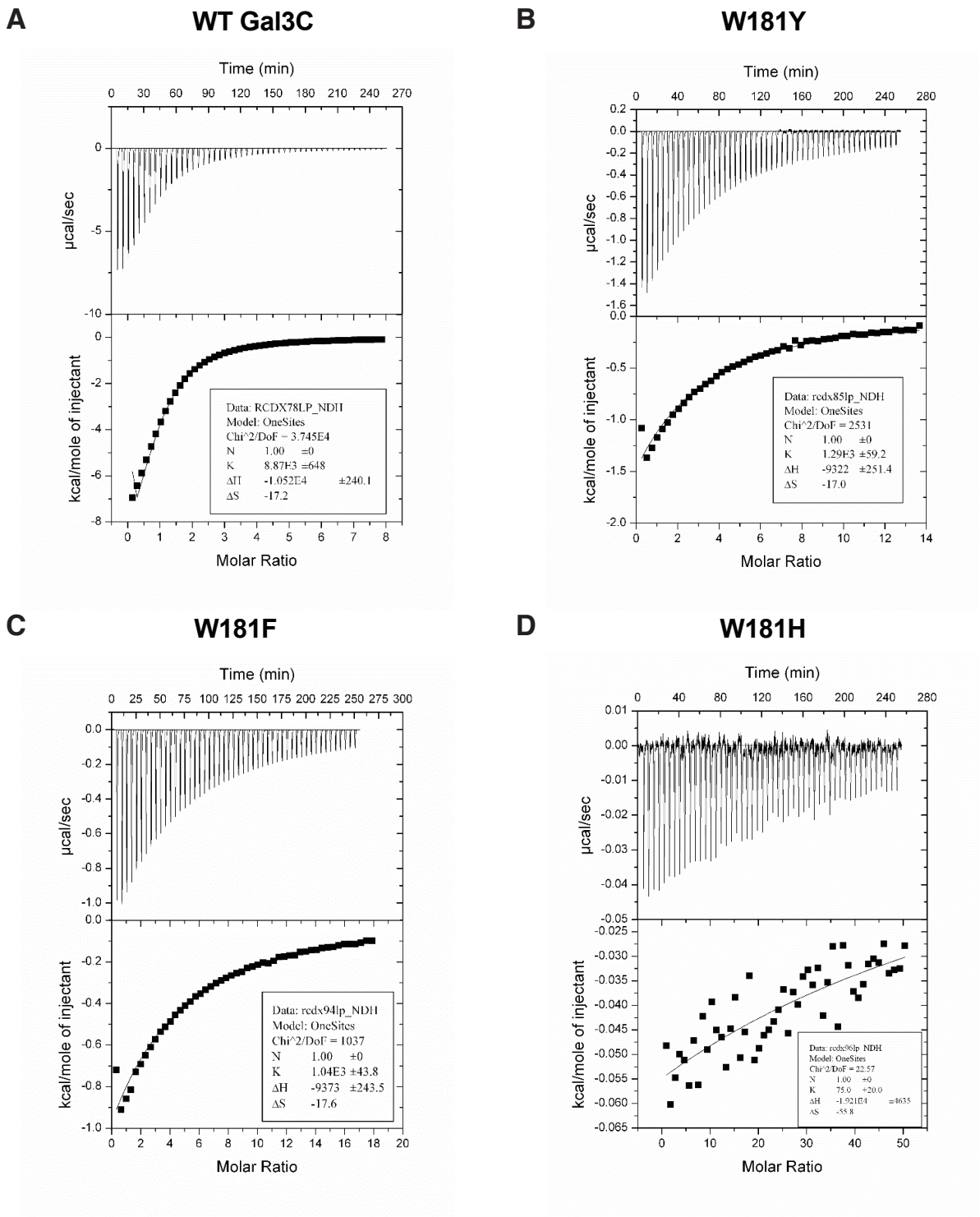


Figure S3. Representative ITC traces of Gal3C W181 variants. (A) WT Gal3C, (B) W181Y (C) W181F, (D) W181F. All experiments were performed in biological triplicate.

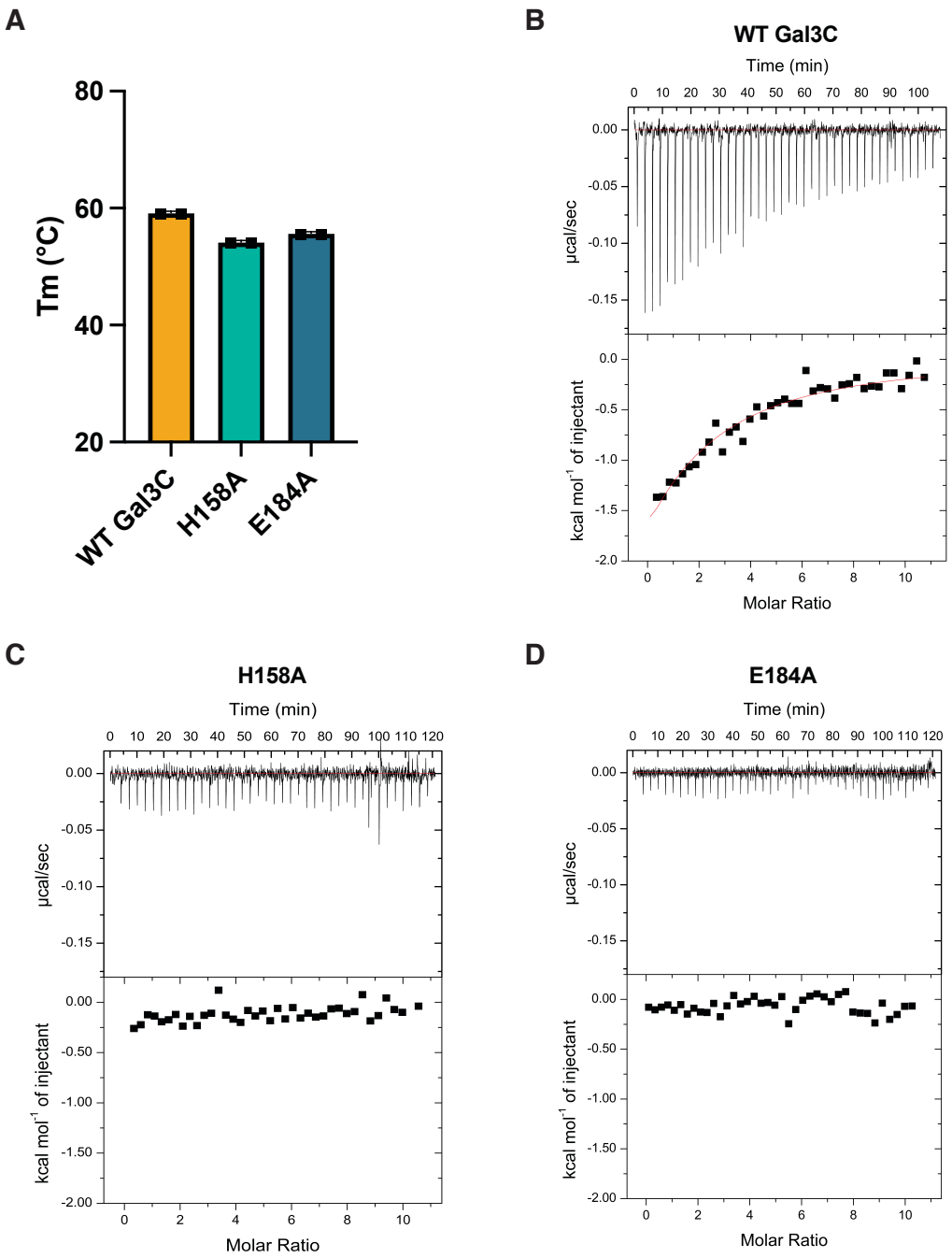


Figure S4. Gal3C variants with abrogated hydrogen bonding interactions are stably folded but do not bind lactose as determined by isothermal titration calorimetry (ITC). (A) $T_m$ of Gal3C H158A and E184A variants compared to WT Gal3C as determined by DSF with Sypro Orange. Both H158A and E184 variants are stably folded. Bars and whiskers indicate mean ± SEM from $n = 3$ independent biological replicates. $T_m$ data for WT Gal3C reproduced from Figure 2 for comparison with H158A and E184A variants. Representative ITC to determine binding affinity of (B) WT Gal3C, (C) H158A, and (D) E184A to lactose. No detectable lactose binding was observed for H158A or E184A. Experiments were conducted at 25 °C in galectin-3 assay buffer.

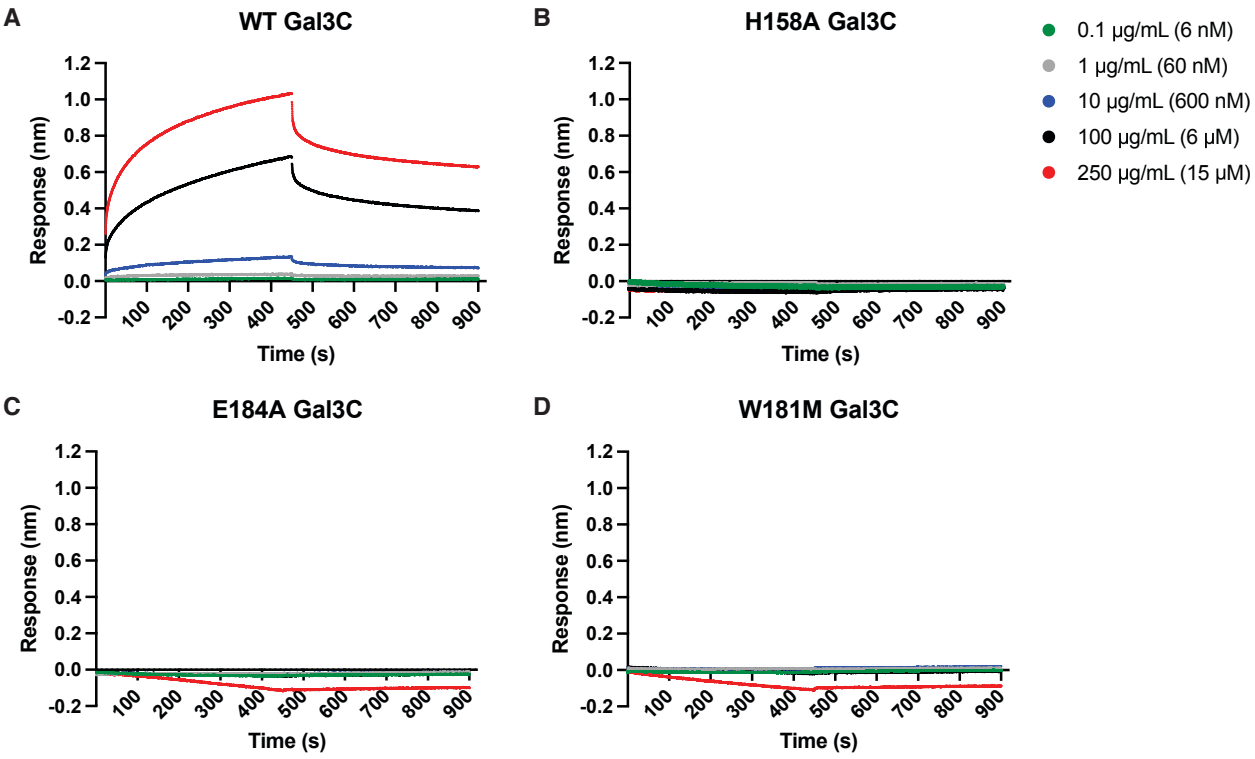


Figure S5. Binding of Gal3C and variants to LacNAc by BLI. (A) WT Gal3C binds strongly to LacNAc, while the (B) H158A, (C) E184A, and (D) W181M variants do not as determined by BLI. Plots depict representative traces from $n = 3$ independent biological replicates. Streptavidin-functionalized biosensors were loaded with 5 µM biotinylated-LacNAc, and touched into solutions with Gal3C concentrations depicted in figure legend.

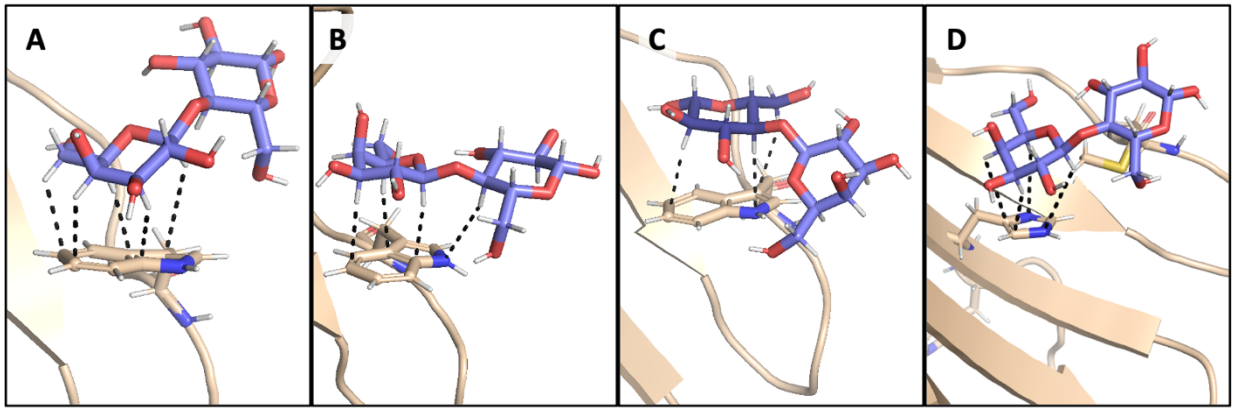


Figure S6. (A) Wild type CH-π interaction between W181 and the C-H groups on the glucose molecule. (B) Binding interaction with lactose rotated, such that the galactose maintains a CH-π interaction with W181 but the glucose is moved out of the binding pocket. (C) Inverted binding interaction with the CH-π interaction formed between the hydrogens on the glucose ß-face and W181. (D) Alternate CH-π interaction formed between H158 and the lactose CH groups involved in the wild type CH-π interaction in the W181M variant protein simulation.

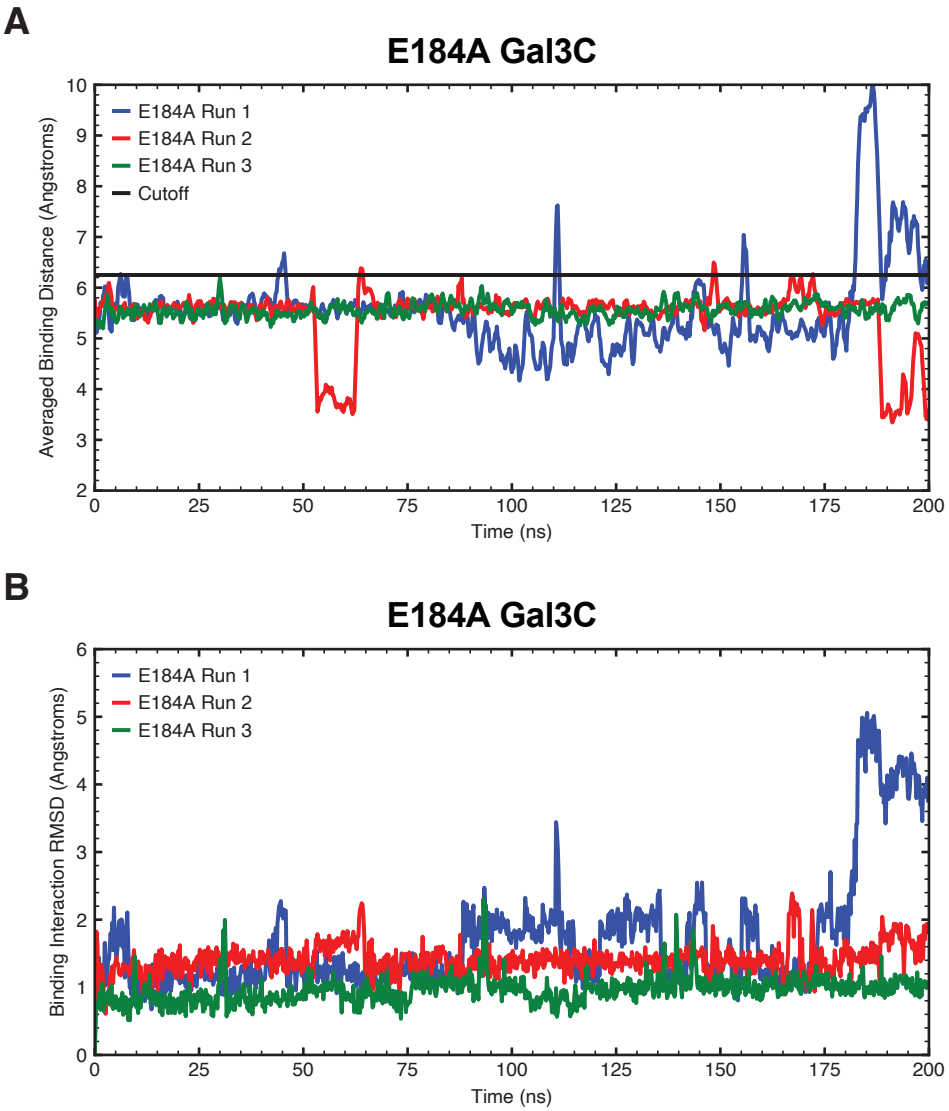


Figure S7. (A) Running average binding distances between the E184A variant protein and the lactose ligand using a 1 ns window. The binding distance is the distance between the center of mass of the lactose ligand and the center of mass of the binding pocket residues H158, R162, W181, and E184 in Å. All 200 ns productions run are shown in blue, red, and green. The cutoff value of 6.25 Å which is used to determine if the ligand is bound to galectin-3 is shown in black. (B) Total RMSD in Å of all atoms in the ligand and the interacting E184A protein residues, H158, R162, W181, and A184, for all three simulations.

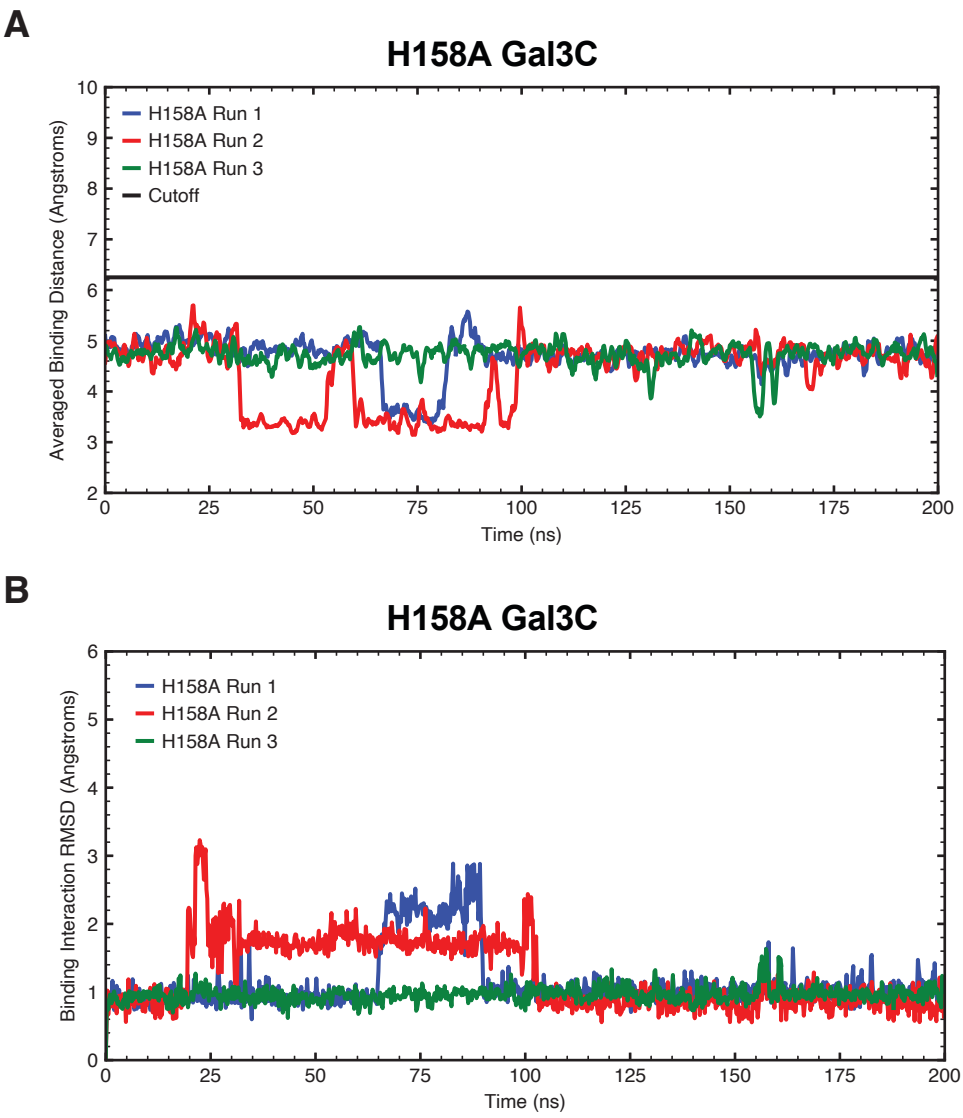


Figure S8. (A) Running average binding distances between the H158A variant protein and the lactose ligand using a 1 ns window. The binding distance is the distance between the center of mass of the lactose ligand and the center of mass of the binding pocket residues A158, R162, W181, and E184 in Å. All 200 ns productions run are shown in blue, red, and green. The cutoff value of 6.25 Å which is used to determine if the ligand is bound to galectin-3 is shown in black. (B) Total RMSD in Å of all atoms in the ligand and the interacting H158A protein residues, A158, R162, W181, and E184, for all three simulations.

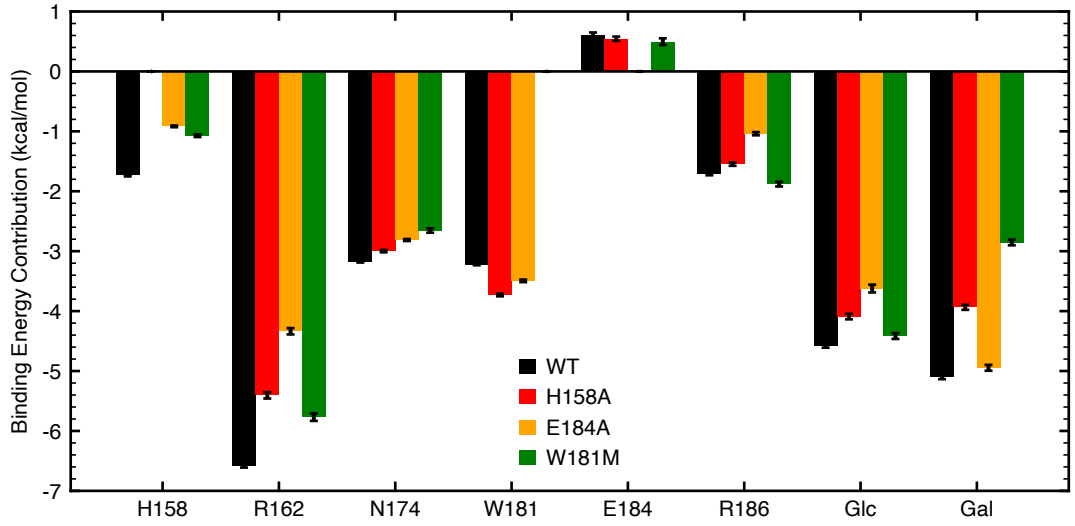


Figure S9. Comparison of the binding enthalpy contributions of glucose (Glc), galactose (Gal), and amino acids H158, R162, N174, W181, E184, and R186 for the wild type and each variant protein. Amino acids are labeled using single letter naming conventions (His: H, Glu: E, Trp: W, Asn: N, Met: M, Ala: A, Arg: R). Bars and whiskers indicate mean ± SEM.

| Protein | $K_d$ (mM) | ΔG (kcal/mol) |
|---|---|---|
| Wild type Gal3C | 0.11 | -5.41 ± 0.06 |
| W181Y | 0.71 | -4.29 ± 0.04 |
| W181F | 0.66 | -4.34 ± 0.15 |
| W181H | 39 | -2.17 ± 0.49 |
| W181M | n.d | n.d |
| H158A | n.d | n.d |
| E184A | n.d | n.d |

Table S1. Association constants and ΔG values for Gal3C variants binding to lactose as determined by ITC. Values represent an average from 3 independent biological replicates.

| | Wildtype | | E184A | | H158A | | W181M | |
|---|---|---|---|---|---|---|---|---|
| | pKa | Q | pKa | Q | pKa | Q | pKa | Q |
| **TYR118** | 11.77 | 0 | 11.84 | 0 | 11.68 | 0 | 11.69 | 0 |
| **ARG129** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **LYS139** | 10.60 | 1 | 10.44 | 1 | 10.59 | 1 | 10.49 | 1 |
| **ARG144** | 11.46 | 1 | 11.12 | 1 | 11.68 | 1 | 11.25 | 1 |
| **ASP148** | <0.00 | -1 | <0.00 | -1 | 4.31 | -1 | <0.00 | -1 |
| **ARG151** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **ASP154** | 3.78 | -1 | 3.35 | -1 | 3.57 | -1 | 3.61 | -1 |
| **H158** | 6.23 | 0 | 5.60 | 0 | **A158** | **0** | 6.20 | 0 |
| **R162** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **GLU165** | <0.00 | -1 | <0.00 | -1 | 0.03 | -1 | <0.00 | -1 |
| **ARG168** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **ARG169** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **CYS173** | >12.00 | 0 | >12.00 | 0 | >12.00 | 0 | >12.00 | 0 |
| **LYS176** | 10.10 | 1 | 9.89 | 1 | 9.98 | 1 | 10.11 | 1 |
| **ASP178** | 3.03 | -1 | 3.12 | -1 | 3.12 | -1 | 3.08 | -1 |
| **ARG183** | >12.00 | 1 | 11.94 | 1 | >12.00 | 1 | >12.00 | 1 |
| **E184** | 1.27 | -1 | **A184** | **0** | 1.74 | -1 | 1.25 | -1 |
| **GLU185** | 2.09 | -1 | 2.52 | -1 | 2.36 | -1 | 2.41 | -1 |
| **R186** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **GLU193** | 1.41 | -1 | 1.34 | -1 | 1.45 | -1 | 1.33 | -1 |
| **LYS196** | 10.29 | 1 | 10.30 | 1 | 10.30 | 1 | 10.31 | 1 |
| **LYS199** | 11.87 | 1 | 11.83 | 1 | 11.84 | 1 | 11.83 | 1 |
| **GLU205** | <0.00 | -1 | <0.00 | -1 | <0.00 | -1 | <0.00 | -1 |
| **ASP207** | 2.95 | -1 | 2.89 | -1 | 2.94 | -1 | 2.95 | -1 |
| **HIS208** | 8.12 | 1 | 8.12 | 1 | 8.17 | 1 | 8.13 | 1 |
| **LYS210** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **ASP215** | 1.93 | -1 | 1.93 | -1 | 1.94 | -1 | 1.95 | -1 |
| **HIS217** | 3.40 | 0 | 3.33 | 0 | 3.35 | 0 | 3.40 | 0 |
| **TYR221** | >12.00 | 0 | >12.00 | 0 | >12.00 | 0 | >12.00 | 0 |
| **HIS223** | <0.00 | 0 | <0.00 | 0 | <0.00 | 0 | <0.00 | 0 |
| **ARG224** | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 | >12.00 | 1 |
| **LYS226** | 10.67 | 1 | 10.63 | 1 | 10.66 | 1 | 10.66 | 1 |
| **LYS227** | 9.59 | 1 | 9.59 | 1 | 9.69 | 1 | 9.58 | 1 |
| **GLU230** | 2.04 | -1 | 1.94 | -1 | 1.98 | -1 | 2.02 | -1 |
| **LYS233** | 9.60 | 1 | 9.52 | 1 | 9.6 | 1 | 9.56 | 1 |
| **ASP239** | 0.67 | -1 | 0.35 | -1 | 0.67 | -1 | 0.33 | -1 |
| **ASP241** | 3.66 | -1 | 3.37 | -1 | 3.64 | -1 | 3.34 | -1 |
| **TYR247** | >12.00 | 0 | >12.00 | 0 | >12.00 | 0 | >12.00 | 0 |
| **Net** | -- | 5 | -- | 6 | -- | 5 | -- | 5 |

Table S2. Charge states (Q) of amino acids with physiologically relevant side chain pKa values. Mutated amino acids are bolded and labeled accordingly. pKa values and resulting amino acid residue charges were assigned by the H++ webserver.[4] Net protein charges are listed.

| | Production Run 1 | Production Run 2 | Production Run 3 | Total # of Snapshots |
|---|---|---|---|---|
| Wild type | 1-1000 | 1-1000 | 1-1000 | 3000 |
| E184A | 1-222, 231-553, 560-778, 784-912, 945-950 | 1-1000 | 1-1000 | 2899 |
| W181M | 1-931 | 1-68, 83-90, 396-405 | 1-366 | 1433 |
| H158A | 1-1000 | 1-1000 | 1-1000 | 3000 |

Table S3. Snapshot ranges that were included in the final trajectories of bound interactions for each variant. Snapshots were taken at 200 ps intervals. The total number of snapshots in each concatenated trajectory of binding interactions is listed in the final column.

| | WT | E184A | W181M | H158A |
|---|---|---|---|---|
| Entropy (kcal/mol) | -17.07 | -19.13 | -18.65 | -18.80 |
| Enthalpy (kcal/mol) | -25.75 | -21.92 | -20.24 | -22.19 |
| Binding free energy (kcal/mol) | -8.67 | -2.79 | -1.58 | -3.39 |

Table S4. Entropy, enthalpy, and binding free energy values all in units of kcal/mol computed using MMGBSA and normal mode analysis for the entropic contribution for the wild type and all three variant proteins (E184A, W181M, H158A). MMGBSA calculations were performed on at least 1400 snapshots and normal mode calculations were performed on out of every 10 snapshots used to compute MMGBSA, at least 140 snapshots for each mutant.

| | | van der Waals (kcal/mol) | Electrostatic (kcal/mol) | Polar Solvation (kcal/mol) | Non-Polar Solvation (kcal/mol) | Total Energy (kcal/mol) |
|---|---|---|---|---|---|---|
| Wild type | H158 | -0.56 | -3.87 | 2.8 | -0.11 | -1.74 |
| | R162 | -0.61 | -12.94 | 7.27 | -0.31 | -6.58 |
| | N174 | -0.22 | -3.24 | 0.32 | -0.04 | -3.18 |
| | W181 | -2.91 | -0.81 | 0.86 | -0.35 | -3.22 |
| | E184 | 0.03 | -14.67 | 15.48 | -0.23 | 0.61 |
| | R186 | -0.59 | 0.21 | 0.006 | -2.35 | 2.19 |
| E184A | H158 | -1.09 | -0.23 | 0.52 | -0.12 | -0.92 |
| | R162 | -0.88 | -11.68 | 8.58 | -0.36 | -4.34 |
| | N174 | -0.25 | -3.49 | 1.01 | -0.09 | -2.81 |
| | W181 | -3.08 | -0.68 | 0.65 | -0.38 | -3.49 |
| | **A184** | -0.38 | -0.31 | 0.32 | -0.05 | -0.42 |
| | R186 | -0.35 | -5.31 | 4.75 | -0.13 | -1.04 |
| W181M | H158 | -1.21 | -0.17 | 0.47 | -0.16 | -1.07 |
| | R162 | -0.75 | -12.38 | 7.67 | -0.31 | -5.77 |
| | N174 | -0.36 | -2.69 | 0.44 | -0.04 | -2.65 |
| | **M181** | -1.38 | 0.01 | 0.2 | -0.29 | -1.45 |
| | E184 | -0.04 | -12.76 | 13.5 | -0.21 | 0.5 |
| | R186 | -0.55 | -2.91 | 1.65 | -0.076 | -1.88 |
| H158A | **A158** | -0.3 | 0.19 | -0.19 | -0.05 | -0.35 |
| | R162 | -0.72 | -11.38 | 7.01 | -0.31 | -5.41 |
| | N174 | -0.37 | -3.01 | 0.45 | -0.06 | -3 |
| | W181 | -3.25 | -0.81 | 0.8 | -0.47 | -3.73 |
| | E184 | -0.12 | -11.61 | 12.49 | -0.21 | 0.55 |
| | R186 | -0.51 | -2.34 | 1.37 | -0.065 | -1.55 |

Table S5. Energy decomposition analysis (in kcal/mol) for the galectin-3 residues participating in the binding interaction with lactose, H158, R162, N174, W181, E184, and R186. The total energy contribution is decomposed into the van der Waals, electrostatic, polar solvation, and non-polar solvation energies. No entropic correction is computed for the residue-wise contributions.

**References:**


(1) Gibson, D. G.; Young, L.; Chuang, R.-y.; Venter, J. C.; Iii, C. A. H.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6* (5), 12-16. DOI: 10.1038/NMETH.1318.
(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235-242. DOI: 10.1093/nar/28.1.235 From NLM Medline.
(3) Saraboji, K.; Hakansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; Logan, D. T. The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry* **2012**, *51* (1), 296-306. DOI: 10.1021/bi201459p From NLM Medline.
(4) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **2005**, *33* (Web Server issue), W368-371. DOI: 10.1093/nar/gki464 From NLM Medline.
(5) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **2008**, *29* (4), 622-655. DOI: 10.1002/jcc.20820 From NLM Medline.
(6) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput Phys Commun* **1995**, *91* (1-3), 1-41. DOI: Doi 10.1016/0010-4655(95)00041-D.
(7) Price, D. J.; Brooks, C. L., 3rd. A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys* **2004**, *121* (20), 10096-10103. DOI: 10.1063/1.1808117 From NLM Medline.
(8) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255 From NLM Medline.
(9) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: Doi 10.1063/1.464397.
(10) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 3878-3888. DOI: 10.1021/ct400314y.
(11) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9* (7), 3084-3095. DOI: 10.1021/ct400341p.
(12) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect Drug Discov* **2000**, *18*, 113-135. DOI: Doi 10.1023/A:1008763014207.

(13) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* **2012**, *8* (9), 3314-3321. DOI: 10.1021/ct300418h.
(14) Kavanaugh, N. L.; Zhang, A. Q.; Nobile, C. J.; Johnson, A. D.; Ribbeck, K. Mucins suppress virulence traits of Candida albicans. *mBio* **2014**, *5* (6), 1-8. DOI: 10.1128/mBio.01911-14.
(15) Wagner, C. E.; Krupkin, M.; Smith-Dupont, K. B.; Wu, C. M.; Bustos, N. A.; Witten, J.; Ribbeck, K. Comparison of Physicochemical Properties of Native Mucus and Reconstituted Mucin Gels. *Biomacromolecules* **2023**, *24* (2), 628-639. DOI: 10.1021/acs.biomac.2c01016 From NLM Medline.