Title

Breaking Down the Bottlebrush: Atomically Detailed Structural Dynamics of Mucins.

Journal

Journal of chemical information and computer sciences, 64(20)

Authors

Kearns, Fiona

Rosenfeld, Mia

Amaro, Rommie

Publication Date

2024-10-28

DOI

10.1021/acs.jcim.4c00613

Peer reviewed

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

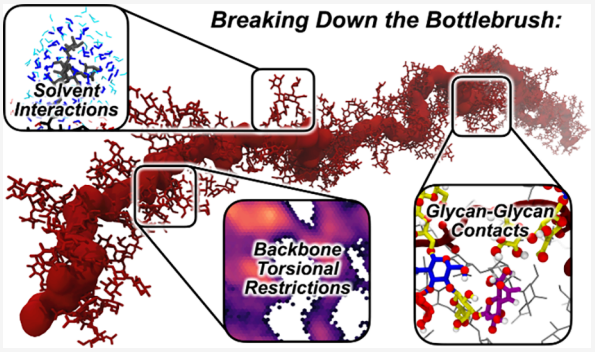
pubs.acs.org/jcim Article

Breaking Down the Bottlebrush: Atomically Detailed Structural Dynamics of Mucins

Fiona L. Kearns, Mia A. Rosenfeld, and Rommie E. Amaro*

Cite This: J. Chem. Inf. Model. 2024, 64, 7949−7965 Read Online

ABSTRACT: Mucins, the biomolecular components of mucus, are glycoproteins that form a thick physical barrier at all tissue−air interfaces, forming a first line of defense against pathogens. Structural features of mucins and their interactions with other biomolecules remain largely unexplored due to the challenges associated with their high-resolution characterization. Combining limited mass spectrometry glycomics and protein sequencing data, we present all-atom, explicitly solvated molecular dynamics simulations of a major respiratory mucin, MUC5B. We detail key forces and degrees of freedom imposed by the extensive *O*-glycosylation, which imbue the canonically observed bottlebrush-like structures to these otherwise intrinsically disordered protein backbones. We compare our simulation results to static structures observed in recent scanning tunneling microscopy experiments as well as other published experimental efforts. Our work represents the demonstration of a workflow applied to a mucin example, which we hope will be employed by other groups to investigate the dynamics and interactions of other mucins, which can inform on structural details currently inaccessible to experimental techniques.



INTRODUCTION

Mucin (MUC) proteins comprise the primary macromolecular component of mucus, a viscous colloid secreted from and with the fundamental role of protecting the epithelium (Figure 1A). The MUC family of proteins is characterized by very long (100−1000 nm, 200 kDa to 2.5 MDa) stretches called "mucin domains" that are rich in prolines, threonines, and serines; thus, they are also referred to as "PTS" domains.[1−11] PTS domains are composed of a linear protein backbone densely modified by *O*-glycans, resulting in over 80% of their total mass being composed of carbohydrates (Figure 1B−D).[1−12]

Mucin-type *O*-glycans are characterized by modification of Ser and Thr residues (rarely also Tyr) with an initiating α-*N*-acetylgalactosamine (α-GalNAc). Following initiation, mucin *O*-glycans elaborate into several core structures and are often terminated by sulfate, sialyl, or fucose moieties (Figure 1D). While N-linked glycosylation has a characteristic amino acid motif (N-X-S/T) for glycosite identification, no such motif for *O*-glycans exists; instead, *O*-glycans can occur on any Ser or Thr residue within the PTS domain.[3,13−15] Elaboration of *O*-glycan structure is nontemplated and controlled by expression levels of glycosyltransferases, sulfotransferases, and sulfatases,[16,17] all thus dependent on the relative abundance of acceptor and donor substrates in the cellular milieu. Such factors make specific *O*-glycosite and *O*-glycotype prediction difficult to generalize.[3,18−25] Overarching properties of mucin *O*-glycosylation, such as the degree of charged sialyl and sulfate groups or neutral fucose groups, enable morphological responses to environmental changes such as pH and ion transfer.[4,26] Furthermore, individual genetics and tissue types impact the baseline expression of different mucins, and anthropological factors, such as smoking and disease status, impact *O*-glycan elaboration/branching/decoration.[5,9,27,28] As a result, mucin domains demonstrate a great deal of macro- (is an *O*-glycan present at a glycosite or not), micro- (relative abundance of glycoforms found at each site including factors such as sialylation/sulfation/fucosylation), and meso-heterogeneity (as environmental factors such as pH and ion transport or disease status can impart mucin morphological changes and/or up/down-regulate certain *O*-glycoforms). Finally, protein backbones within mucin PTS domains are intrinsically disordered (Figure 1B,C),[29] but such dense modification by *O*-glycans is canonically said to imbue a "bottlebrush"-like structure. Thus, *O*-glycosylation within the mucin domain impacts biophysical and biochemical forces on its surroundings modulated by the ionic/pH microenvironment and expression levels.

Received: April 25, 2024
Revised: September 19, 2024
Accepted: September 19, 2024
Published: September 27, 2024

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

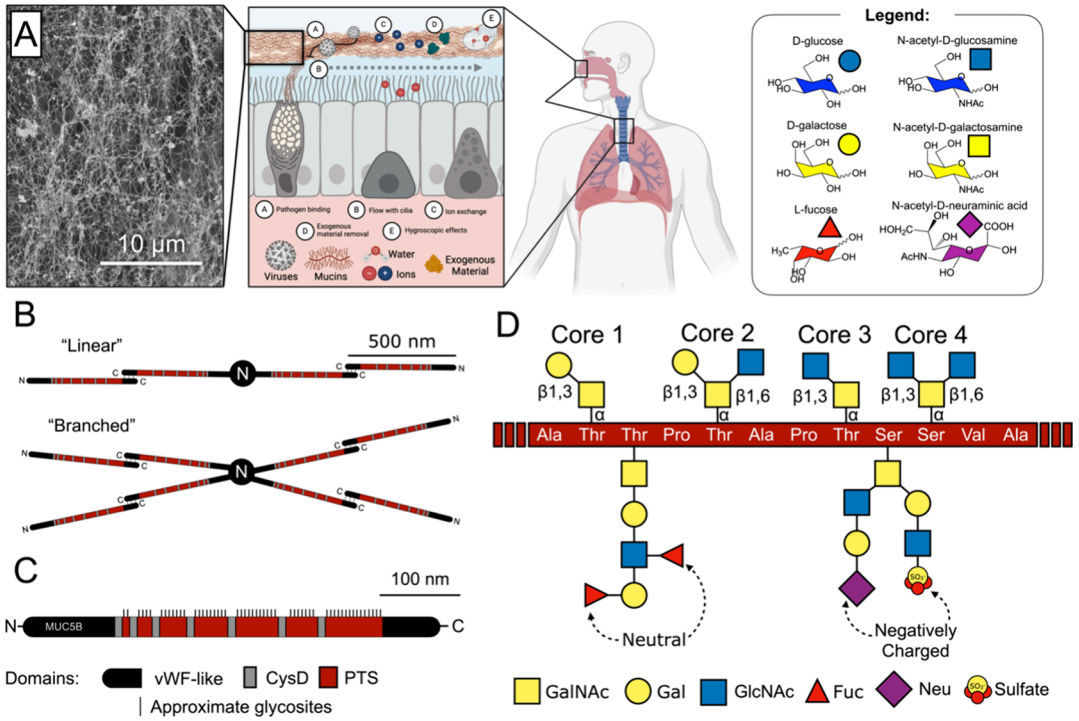


Figure 1. Legend detailing the saccharides discussed in this work and denoting their representations in the symbol nomenclature for glycans (SNFG). (A) Schematic demonstrating respiratory mucin expression, the many roles of mucins in the respiratory tract, and a scanning electron tomograph of a respiratory mucosal membrane, scanning image taken from Figure 1 panel E in Carpenter et al.[86] (B) Typical cross-linking schemes seen between mucin biopolymers, image adapted from Figure 1 in Symmes et al.[5] (C) Structural domains of MUC5B mucin, image adapted from Figure 1 in Symmes et al.[5] (D) Description of mucin structure at the atomic level, detailing common core types, connectivity, and ending motifs. All structures drawn with SNFG representation and color schemes; image adapted from Figure 1 in Symmes et al.[5] Scheme in panel (A) created with Biorender.com.

The mucosal layer and its composite mucins play many roles in cell biology and physiology, including, but not limited to (Figure 1A) signal transduction, hygroscopic effects such as moisture maintenance, ionic exchange, and removal of pathogens or exogenous material through capture and subsequently moving in conjunction with ciliary beating.[1,5,8−13,16,17,30−44] Such functions are dependent on the unique mucin structure, their resulting network-like mesh formed by long interacting strands (Figure 1A), and the ability to change morphology as a result of microenvironment ionic strength and pH.[4,26,43−45] As mentioned, disease status can often impact mucin expression levels and *O*-glycoprofiles.[1,5,27,28,46] For example, truncated aberrant mucin *O*-glycans are shown to be a hallmark of progression and metastasis in many cancers.[33,39−42,47] Relatedly, one of the few FDA-approved biomarkers for ovarian cancer is CA125, an epitope within the canonical mucin MUC16.[48−53] Furthermore, increased expression of mucins is known to inhibit immune recognition either through physical barrier formation or inhibition of T-cells via binding of mucin sialic acids to Siglecs (sialic acid immunoglobulin-like lectins).[33,36,39−42,47] Due to the many vital roles of mucins, dysregulation in expression, sequence, and structure can be detrimental to health and disease progression.

Despite their importance, because of their heterogeneity and their highly dynamic nature, structural characterization of mucins by traditional techniques such as cryo-electron microscopy and X-ray crystallography is often intractable.[29,34,54] Mass spectrometry and glycomics techniques, especially in combination with the development of mucin-specific protease (i.e., mucinases) enzymology, inform on the proportion of particular glycoforms on glycosites.[2,55−58] Further, Nason et al. have elegantly demonstrated the ability to produce highly controlled mucin fragments,[19,54] which then enabled Anggara et al. to, using scanning tunneling microscopy images, beautifully reveal images of MUC1 after gentle electrospray and surface deposition, representing some of the first work to "see" mucins in near single-molecule detail.[59] However, there remains a gap between mucin glycoproteomic sequence information and the high-resolution structure and dynamics of mucins. Molecular modeling and simulations are beginning to fill these gaps. Recently, enabled by mass spectra acquired through characterization and upcycling of the mucinase SmE, we presented the first ever all-atom molecular dynamics (MD) simulations of a mucin-domain containing proteins, T-cell immunoglobulin mucin-domain containing protein 3 and 4 (TIM-3 and TIM-4).[60] In that work, novel structural biology techniques in tandem with MD simulations revealed the roles that glycans play in the structure, stability, dynamics, and function of TIM-3 and TIM-4.[60] Additionally, nonequilibrium shear simulations by Boushehri et al. of mucin-like lubricin, reveal how *O*-glycans induce extended conformations along the protein backbone and reduce solution viscosity compared to unglycosylated systems of the same density.[61]

In this current work, we describe the construction of mucin models for simulation using respiratory mucin MUC5B as a test case. MUC5B is a massive (200 kDa to 2.5 MDa) glycoprotein that binds to and removes pathogens in the human airway, forming a thick physical barrier between our

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

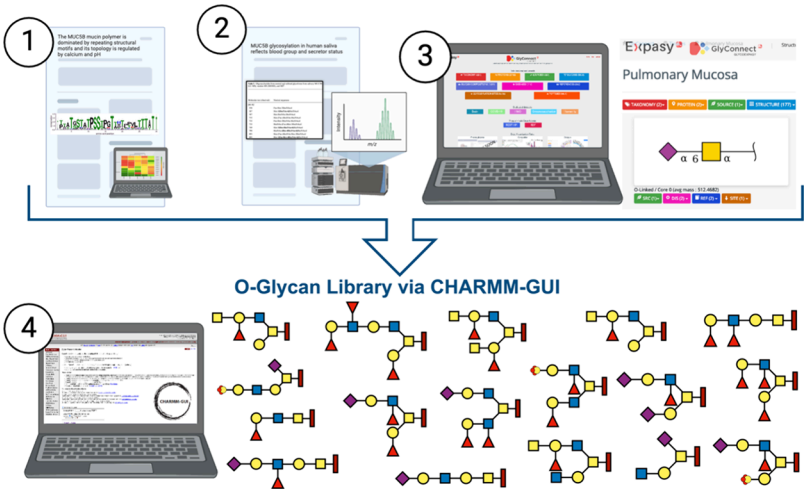


Figure 2. Workflow outlining how glycoproteomic data was collected to build miniature MUC5B mucin models simulated in this work. Steps 1 and 2 reflect collecting glycoproteomic data from the literature resources. Step 3 reflects cross-checking of those data with the GlyConnect Expasy Database and selecting of the *O*-glycan structures that are consistent with those data. Step 4 reflects constructing such *O*-glycans with the CHARMM-GUI Glycan Builder. Created with BioRender.com.

cells and the outside world.[4−6,27,29,31−34,37−39,47,62] MUC5B is expressed and cleaved from goblet cells in the lungs, binds to pathogens and exogenous materials, and removes them in conjunction with ciliary beating and aerosolization.[26,27,30,62−66] MUC5B changes phase from gel- to fluid-like in the lungs as a function of ionic concentration (particularly Ca^{2+} and Cl^{-}) and pH.[4,26] We use published mass spectrometry glycoproteomic sequencing data to construct all-atom models of two repeating MUC5B respiratory mucin PTS domains (∼30 amino acids in length). We then scaled up this model by approximately 1 order of magnitude (∼224 aa in length) and conducted further simulation to explore structural dynamics at this second scale. Our simulations allow us to characterize atomic level forces and interactions driving the experimentally observed mucin "bottlebrush" morphology. Overall, in this work we (1) demonstrate the use of a computational workflow for identifying relevant *O*-glycans and constructing mucin-domain containing glycoproteins, systems which are relevant to a vast array of biomedical and materials applications, (2) develop experimentally corroborated molecular models of MUC5B, and (3) predict atomic-scale biophysical behavior driving macro- and mesoscale mucin morphology.

METHODS

With the goal of interrogating the atomic-level structural dynamics of respiratory mucins, we employed a standard workflow for identifying relevant *O*-glycans and constructing all-atom models of such species for simulation. In doing so, we constructed and simulated two heterogeneous "mini mucin" models with slightly different degrees of charge. This workflow centers on collecting relevant protein and glycomics sequence information (and/or choosing glycans from a database known to satisfy physiochemical properties, such as charge balance) and generating a small representative library for the work in question (Figure 2). Thus, while in many cases exact glycoforms cannot be resolved, representative and relevant molecular models of mucins can still be constructed to probe atomic-scale properties.

We followed the workflow (Figure 2) to build two biochemically relevant models of short, ∼30 amino acid long segments of MUC5B. Throughout this work, we refer to these two models as "Mini1" and "Mini2," with each model reflecting a protein sequence based on a protein consensus sequence repeated within the MUC5B PTS domains. A complete step-by-step guide of our workflow can be seen in the Supporting Information (SI Section 1.0), and we provide a summary of this workflow herein. Two protein MUC5B PTS domain consensus repeat sequences were taken from Hughes et al.[5] with amino acid (aa) lengths of 26 and 30, wherein they investigated the relative conservation of sequences within PTS domains.[4] Rosetta's BuildPeptide tool[67] was used to build peptide structures of Mini1 (26 aa in length) and Mini2 (30 aa in length) sequences.[5] We then compared extensive mass spectrometry (gel electrophoresis combined with LC-MS and LC-MS/MS) data from Thomsson et al.,[6] which lists *O*-glycan structures from salivary mucosal samples but without stereochemistry or connectivity information, to *O*-glycans isolated from pulmonary mucosa from the GlyConnect Expasy database which provides stereochemistry and connectivity information.[68] *O*-Glycans from the GlyConnect Expasy database were selected that corresponded to glycans with masses comparable to molecular ions listed in Thomsson et al.'s[6] Table 1. Additionally, care was taken to select a variety of charged and neutral glycans as well as a higher proportion of Core-1 and Core-2 *O*-glycans than Core-3 and Core-4 *O*-glycans, as the former are reported to be more common. CHARMM-GUI[69] Glycan Modeler[70] was used to add selected *O*-glycan structures from steps 2 and 3 to our Mini1 and Mini2 protein sequences. All Ser and Thr residues on Mini1 and 2 sequences were modeled with *O*-glycans (except Thr26 on Mini1, which was left unglycosylated). The positions of selected *O*-glycan structures were then "pseudo-randomized" onto their assigned mucin models: random numbers were used to generate initial positions, but positions were then modified to avoid overclustering of charged glycans within mucins. The resulting *O*-glycoprofile maps for Mini1 and Mini2 can be seen in Figure 3A and proportions of core types and charged groups within Mini1 and Mini2 can be seen in Table 1.

Additionally, to investigate the role of *O*-glycans in mucin conformational dynamics, i.e., in promoting a "bottlebrush" structure, we also constructed unglycosylated mini mucin systems using only the peptide backbones for each Mini1 and

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

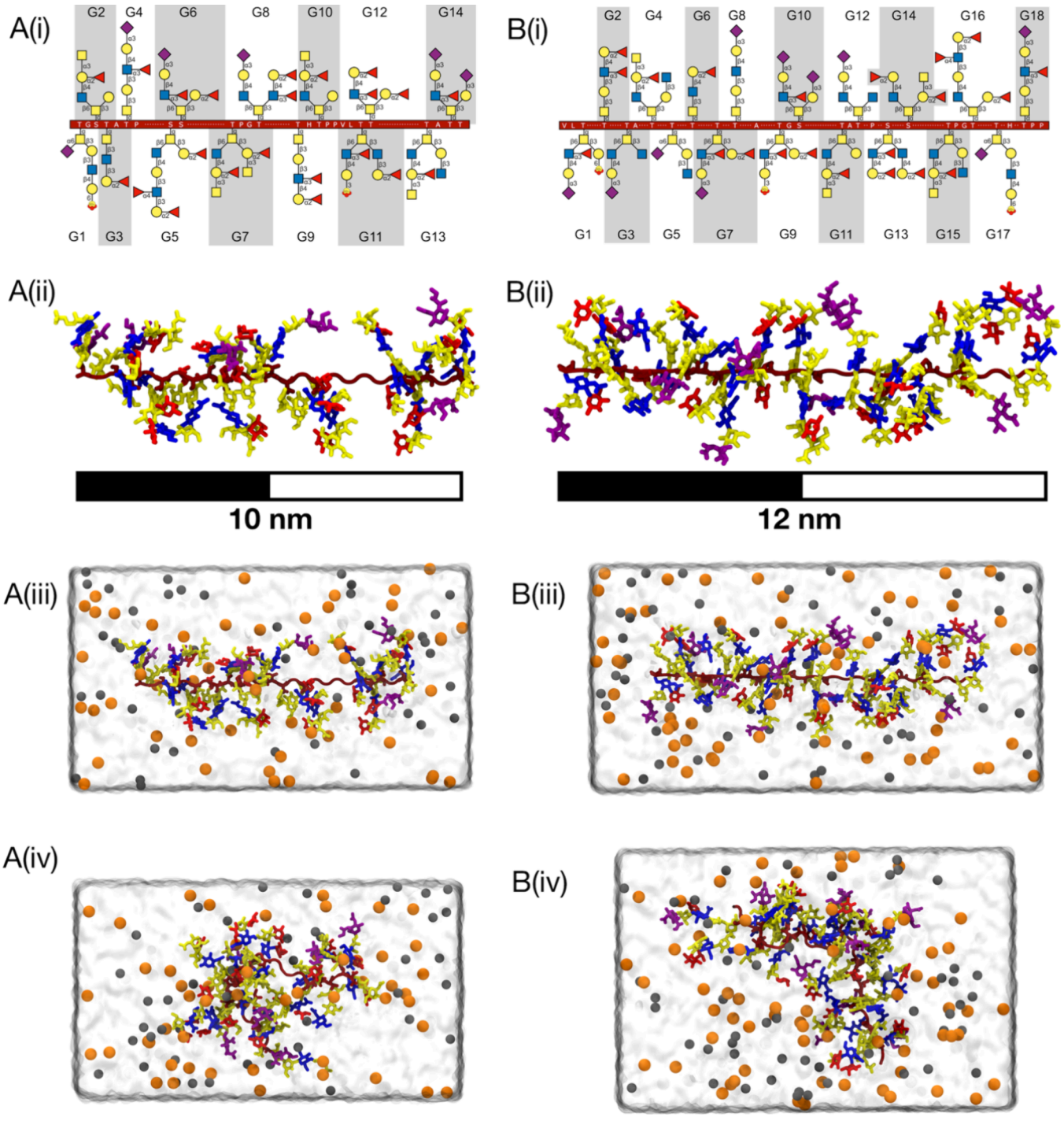


Figure 3. Construction of mini mucin systems. (A) Mini1 (i) glycoprofile and protein sequence, (ii) preminimization structure, (iii) solvated and neutralized structure, and (iv) snapshot from 550 ns conventional production simulations. (B) Mini2 (i) glycoprofile and protein sequence, (ii) preminimization structure, (iii) solvated and neutralized structure, and (iv) snapshot from 550 ns conventional production simulations. Protein backbones are colored in dark red, glycans are colored according to SNFG color scheme, and Na+ and Cl− ions are shown in orange and gray spheres, respectively.

Mini2 (Figure S1). Following the initial model construction of Mini1 and 2 glycosylated (Figure 3B) and unglycosylated (Figure S1) models, we then used VMD tools to solvate and neutralize these structures in TIP3 water boxes with 150 mM NaCl (Table 2 lists system breakdowns for all mini models in both glycosylated and unglycosylated states). All simulations were performed in three replicas with NAMD3[71,72] and parameters described by CHARMM36m[73−75] all-atom force fields using San Diego Supercomputing (SDSC) resources. For Mini1 and Mini2 simulations, periodic boundary conditions[76,77] were applied, particle mesh Ewald[78] was used to calculate long-range Coulombic potentials, Lennard−Jones potentials[79−81] were calculated according to a 10−12 Å cutoff and switching scheme, and pairwise list generation was cutoff at 15.5 Å. Langevin thermostat and piston were used to maintain the temperature at 310 K unless otherwise stated and

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

Table 1. Description of Mass Compositions in Mucin Models Mini1 and Mini2[a]

	Mini1 (#)	Mini1 (% mass)	Mini2	Mini2 (% mass)
#aminoacids	26	12.6	30	12.0
#glycans	14	87.4	18	88.0
#Core-1	3	16.8	5	21.8
#Core-2	7	46.0	8	43.1
#Core-3	2	10.3	1	3.0
#Core-4	2	14.4	4	20.1
#sialylated Glys (1,2)	3, 1	8.9	8, 1	12.3
#sulfates	2	0.8	3	1.0
#fucosylated Glys (1,2,3)	6, 5, 2	16.5	7, 3, 2	13.0
total charge	−8		−13	

[a]Sulfate %mass is calculated as the mass of the sum of SO^{3-} groups divided by total mass.

at 1 atm of pressure. Minimization: To relax all conformations and alleviate close contacts resulting from CHARMM-GUI *O*-glycosylation steps, all solvated and neutralized structures were subjected to 10,000 steps of conjugate gradient energy minimization with no atomic restraints or constraints. Heating: Systems were then heated from 10 to 310 K, increasing temperature by 25 K every 10,080 steps, with a time step of 2 fs, for a total of 240 ps of heating. Once at 310 K, an additional short equilibration of 770 ps of the simulation was conducted. Equilibration: Mini1 and Mini2 were then subjected to NpT equilibration at 310 K for 0.5 ns, with a 2 fs time step (useFlexibleCell option "on", wrapWater option "on", wrapAll option "on"). Systems were then subjected to an additional 55 ns NVT equilibration with a 2 fs time step with fixed box dimensions (useFlexibleCell option "off"). Production: 550 ns of NVT initial conventional MD simulation production runs were conducted at 310 K, with a time step of 2 fs. Following 550 ns of initial MD production runs, each simulation replica was forked to continue with (1) ∼2 μs of cMD simulations or (2) Gaussian accelerated MD (GAMD) equilibration and production for (∼600 ns for glycosylated mucins and ∼2 μs for unglycosylated mucins). See Supporting Information Schemes S1 and S2 for the graphic outline of simulation steps and see Supporting Information Methods Section 1.1 for complete simulation details as well as Table S1 for complete sampling details. All data presented in the Results and Discussion section for glycosylated and unglycosylated mini mucin systems are derived from these cMD and GAMD production runs (∼27 and ∼52 μs aggregate sampling for glycosylated and unglycosylated mini systems, respectively).

To estimate more physiologically relevant mucin conformations, we also simulated larger "Mega" mucin models. We constructed our "Mega" mucin system by patching the C-terminal of Mini1 to the N-terminal of Mini2, then quadrupling this Mini1+Mini2 unit to form a long chain (all done with VMDtools,[82] see shared files and the Supporting Information). The final resulting mega mucin was 769 Å in length along the protein backbone, comprising 224 amino acids and 128 glycans. As with the Mini mucins, we also sought to investigate the impact of *O*-glycans on mucin backbone conformations; thus, we constructed an unglycosylated mega mucin model by removing *O*-glycans and generating resultant psf/pdb starting files with psfgen. As done for the mini mucin systems, the mega mucin glycosylated and unglycosylated systems were solvated and neutralized in a TIP3 water box with 150 mM NaCl (Figure S1C,D), minimized, heated, and equilibrated according to the following simulation protocols. Due to the size of the mega mucin systems (>1 M atoms glycosylated, ∼120 K atoms unglycosylated, Table 2), and considering the potential flexibility differences between glycosylated and unglycosylated mega mucin backbones, we opted for differing simulation workflows to maximize sampling for the glycosylated mega mucin runs, which were considerably more computationally expensive than the unglycosylated mega mucins. All simulations were performed with NAMD3[71,72] using either San Diego Supercomputing or TACC Frontera Resources, depending on availability.

Mega Glycosylated Simulations. See Supporting Information Scheme S3 for a graphic outline of glycosylated mega mucin simulation steps. Minimization: To relax all conformations and alleviate close contacts resulting from CHARMM-GUI *O*-glycosylation and psfgen patching steps, one replica of the mega glycosylated mucin was subjected to 10,000 steps of conjugate gradient energy minimization with no atomic restraints or constraints. Heating: The initial replica of the minimized mega glycosylated mucin was then heated from 10 to 310 K, increasing temperature by 25 K every 10,080 steps, time step 2 fs, for a total of 240 ps of heating. Once at 310 K, an additional short equilibration of 770 ps of the simulation was conducted. Equilibration: The one replica of mega glycosylated mucin was then subjected to NpT equilibration at 310 K for 0.5 ns, 2 fs time step (useFlexibleCell option "on", wrapWater option "on", wrapAll option "on") followed by an additional 55 ns NpT equilibration with 2 fs time step with fixed box dimensions (useFlexibleCell option "on"). Initial production: 195 ns of NpT initial conventional MD simulation production run was conducted at 310 K, time step 2 fs. Following 195 ns of initial MD production runs, the single replica of mega glycosylated mucin was forked into three independent replicas; the first replica was continued from prior coordinates and velocities, but the two additional replicas were launched from randomly initialized velocities at 310 K. Replica production: Following this initial forking, each simulation replica was further simulated with conventional MD under 310 K, NpT conditions (useFlexibleCell "on"), time step 2 fs, for an additional ∼650 ns. Following this per-replica cMD production run, replica 1 was then forked again to continue with (1) ∼500 ns of cMD simulations or (2) GAMD equilibration and production for ∼100 ns. Due to resource limitations and computational expense, replicas 2 and 3 were not forked for GAMD runs but instead were allowed to continue with extended cMD runs (∼700 ns each per replica 2 and 3). See Supporting Information Methods for complete simulation details, including GAMD simulation options, and Table S2 for complete sampling details. All data presented in the Results and Discussion section for the glycosylated mega mucin systems is derived from these cMD and GAMD production runs (∼2.7 μs aggregate sampling).

Mega Unglycosylated Simulations. Considering the initial minimization steps for mega unglycosylated simulations could dramatically impact the direction of later dynamics, we opted to run ten fully independent replicas of the unglycosylated mega system. Thus, each of the following steps was conducted for a total of 10 replicas; see Supporting Information Scheme S4 for the graphic outline of unglycosylated mega mucin simulation steps. Minimization: To relax all

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

Table 2. Description of Total Solvated and Neutralized Mini Mucin Systems (Glycosylated and Unglycosylated) Simulated in This Work[a]

	Mini1-Glyco	Mini2-Glyco	Mini1-Unglyco	Mini2-Unglyco	Mega-Glyco	Mega-Unglyco
total #atoms	50,924	57,078	15,964	17,641	1,040,215	116,265
#protein atoms/residues	332/26	380/30	346/26	398/30	2827/224	2899/224
protein backbone length (Å)	88	99	88	99	769	769
#glycan atoms/#glycans	1,055/14	1,143/18	-−/−-	-−/−-	20,492/128	-−/−-
#water atoms/residues	48,231/16,077	53,841/17,947	15,588/5,196	17,211/5,737	506,343/168,781	113,154/37,718
#Na/#Cl ions	53/45	64/51	15/15	16/16	518/450	106/106
system dimensions (AxAxA)	63.9 × 68.7 × 124.99	65.7 × 68.4 × 137.1	36.2 × 39.6 × 120.8	36.7 × 40.0 × 134.6	110.6 × 112.7 × 860.2	40.0 × 39.7 × 789.9
total production sampling	Rep1:4249.79 Rep2:3811.07 Rep3:5486.26 Total: 13547.12	Rep1:4945.98 Rep2:4116.46 Rep3:5011.38 Total: 14073.82	Rep1:9172.29 Rep2:9155.46 Rep3:8655.20 Total: 26982.95	Rep1:8290.75 Rep2:8295.09 Rep3:8205.67 Total: 24791.51	Rep1:1203.72 Rep2:737.40 Rep3:728.32 Total: 2669.44	Rep1:3392.04 Rep2:3067.80 Rep3:2987.04 Rep4:2823.46 Rep5:2829.66 Rep6:2833.76 Rep7:2953.30 Rep8:3086.22 Rep9:2960.32 Rp10:2729.86 Total: 29663.46

[a]Protein backbone lengths are given for the fully outstretched state taken from the minimized frame. All sampling values are listed in nanoseconds (ns), see Tables S1 and S2 for the complete breakdown of sampling totals per model per replica and per simulation type.

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

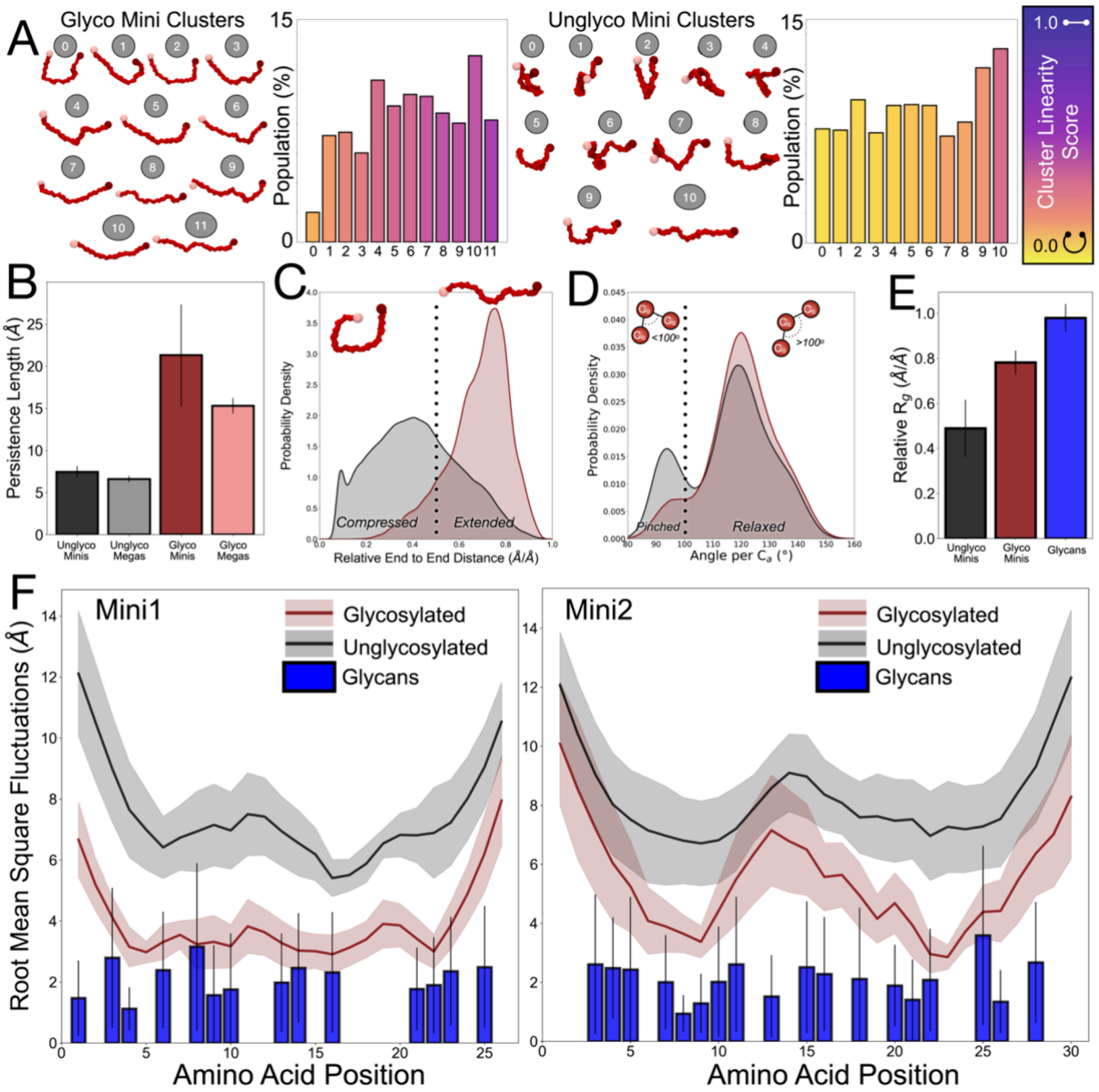


Figure 4. Describing the bottlebrush. (A) Representative structures from glycosylated and unglycosylated mini mucin clustering results alongside cluster population densities (%). Cluster population bars are colored on a scale according to the cluster linearity score. Cluster linearity score bar is accompanied by small graphics denoting near linear conformations (~1.0) and nonlinear conformations (~0.0). In (B–F) black, red, pink, and blue trends/bars indicate naked mini mucin simulations, glycosylated mini mucin simulations, glycosylated mega mucin simulations, and glycans from glycosylated mini mucins simulations, respectively. (B) Persistence lengths calculated from naked mini mucins, glycosylated mini mucins, and glycosylated mega mucins. (C) Relative end to end distances calculated for glycosylated mini mucins and naked mini mucins. See SI Figure S12A for the percent of frames unglycosylated and glycosylated mini mucins spend in compressed conformations, i.e., with relative end to end distance less than 0.5. (D) Distributions of angles per C_α for glycosylated and naked mini mucins. Angles per C_α of less than 100° correspond to rotations consistent with α-helices and are thus marked "pinched". See SI Figure S12B for the average number of pinched C_αs per frame in unglycosylated and glycosylated mini mucins. (E) Average relative radius of gyration calculated for unglycosylated mini mucins, glycosylated mini mucins, and for all glycans from glycosylated mini mucin simulations. (F) RMSFs calculated for unglycosylated and glycosylated Mini1 and Mini2.

conformations and alleviate close contacts resulting from psfgen patching steps, each of the 10 replicas of the mega unglycosylated mucin was subjected to 10,000 steps of conjugate gradient energy minimization with no atomic restraints or constraints. Heating: Each replica of the minimized mega unglycosylated mucin was then heated from 10 to 310 K, increasing temperature by 25 K every 10,080 steps, time step 2 fs, for a total of 240 ps of heating. Once at 310 K, an additional short equilibration of 770 ps of the simulation was conducted. Equilibration: Each replica of mega glycosylated mucin was then subjected to NpT equilibration at 310 K for 0.5 ns, 2 fs time step (useFlexibleCell option "on", wrapWater option "on", wrapAll option "on") followed by an additional 55 ns NpT equilibration with 2 fs time step (useFlexibleCell option "off"). Production: 600–650 ns of NpT initial conventional MD simulation production runs were conducted at 310 K, time step 2 fs. Following ~600–650 ns of initial MD production runs, each replica of mega unglycosylated mucin was forked to continue with (1) ~2 μs of cMD simulations or (2) GAMD equilibration and production for

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949–7965

200−300 ns. See Supporting Information Scheme S4 for a graphic outline of simulation steps and see Supporting Information Methods for complete simulation details, including GAMD simulation options, as well as Table S2 for complete sampling details. All data presented in the Results and Discussion section for the glycosylated mega mucin systems is derived from these cMD and GAMD production runs (∼30 μs aggregate sampling). Convergence quality metrics for all mini and mega mucin simulations (glycosylated and unglycosylated) was checked with Pymbar's timeseries module, particularly the detect_equilibration utilities, see Table S3 for sampling convergence metrics for mega mucin simulations.[83,84]

Following the above construction and simulation protocols, we then conducted extensive analyses of these simulations using MDAnalysis and in-house scripts,[85,86] see shared Supporting Information files for all analysis scripts. In several of the analyses, we refer to relative values, such as relative end-to-end distances. These relative values are calculated as ratios of, e.g., the end-to-end distance per frame by the end-to-end distance of the minimized (outstretched) structure before simulation. Thus, in this example, a relative end-to-end distance value approaching 1 would indicate that the protein backbone in that frame is nearly as extended as the minimized state, whereas a value around 0.5 indicates that the two ends of the protein backbone are half as extended as the minimized state, and less than 0.5 indicates that the two ends of the protein backbone are very close to each other, as would be required for a compact conformation.

RESULTS AND DISCUSSION

Characterizing Biophysical Properties for Glycosylated Mini Mucin Models. We sought first to compare glycosylated Mini1 versus Mini2 simulations to identify if the constructed *O*-glycan and protein sequence heterogeneity induced significant dynamical differences. No significant differences between Mini1 and Mini2 were exhibited in the following properties: persistence length, relative end-to-end distances, relative radius of gyration (R_g, ratio of instantaneous R_g to minimized R_g), relative compactness (ratio of instantaneous compactness relative to minimized compactness), curvature per C_α, angle between each C_α, and Ramachandran and Janin plots for glycosites and non-glycosites. All glycosylated Mini1 vs Mini2 data can be seen in the Supporting Information Figure S2, for complete details on how these data were calculated please see SI Methods Sections 1.3.1−1.3.11. These calculations indicate that, despite their specific protein and *O*-glycan sequence differences, on a global level the mini mucins demonstrate markedly similar conformational landscapes. Significant differences were observed for root-mean-square fluctuations (RMSFs) on a per-residue basis between Mini1 and Mini2 and we discuss those differences in detail later.

To better characterize the conformational landscapes accessible to Mini1 and Mini2 glycosylated mucin structures, we conducted principal component analysis (PCA) on all backbone conformations from all Mini1 and Mini2 simulations (all cMD and GAMD trajectories for all mini mucins concatenated and aligned by C_αs, Supporting Information Methods 1.3.1 for clustering details and Figures S3 and S4 for PCA and kmeans clustering statistics). PCA revealed that three PCs could explain greater than 60% of the variance in C_α atomic positions. We clustered all frames according to these 3 PCs with scikit-learns kmeans clustering module to identify 12 clusters. Clusters were identified via the sum of the square error inflection point and silhouette coefficient. See SI Methods Section 1.3.1 for complete details about how trajectories were merged and how PCA and kmeans clustering were conducted. Per frame, we then calculated a relative linearity score (RLS) by multiplying the framewise relative end-to-end (REE) distance to the framewise relative radius of gyration for C_α atoms (RR_gC_α). Per cluster, a cluster-specific RLS was found by taking the median RLS for all frames in the cluster. Details on how each of these terms were calculated per frame are shown in the Supporting Information Methods Section 1.3. We then ordered the original clusters from 0 to 11 according to this RLS for ease of comparison: 0 having the lowest RLS (which estimates a "low linearity", thus "high curvature"/"high compactness"), 11 having the highest RLS (which estimates a "high linearity"/"low compactness"). Protein backbone conformations for clusters 0−11 can be seen in Figure 4A, alongside the relative populations for each cluster in percentage. Each cluster's bar is colored according to that cluster's RLS. Distributions of all terms required for RLS calculation per cluster are shown in Supporting Information Figure S4. We did see that some clusters were more populated for glycosylated Mini1 or Mini2 systems, Figure S2. For example, cluster 1 was highly populated for Mini1, while clusters 4 and 10 were highly populated for Mini2. However, representative structures for these moderate linearity clusters are similar, and distinctions between cluster populations between Mini1 and 2 are likely due to sampling limitations. Thus, at this time we cannot identify glycoprotein sequence differences that would lead to conformational landscape preferences using only these mini mucin examples.

Mini1 and Mini2 were constructed with different protein backbones and *O*-glycoprofiles, and yet they have remarkably similar conformational landscapes and biophysical properties, as seen in the clustering results and persistence lengths (Figure S2). Although the two mini mucin sequences are distinct, they were both constructed from MUC5B protein consensus repeat sequences and pulmonary mucosa *O*-glycoforms found within the GlyConnect Expasy database. That the two resultant models behave similarly suggests that the key biophysical characteristics of mucins are "stored" within these consensus sequence repeats. Furthermore, we hypothesize from these results that although there is a great deal of *O*-glycan heterogeneity within mucin domains, if the overarching parameters are maintained, i.e., type similarity in protein backbones, *O*-glycan density, and charge density, then backbone conformations and overall mucin morphology could be similar.

Mucin Models Adopt Distinct Conformational States in Glycosylated and Unglycosylated States. To probe how mucin *O*-glycans promote a bottlebrush-like structure on mucin protein backbone conformations, we compared all glycosylated mini mucin data to all unglycosylated mini mucin data. Several groups have characterized the bottlebrush through electron microscopy and other techniques and have begun to piece together the key forces related to the nature of the "bottlebrush".[4] Our MD simulations reveal that the mucin *O*-glycan bottlebrush results in increased linearity of backbone conformations, increased persistence length along the peptide backbone, and decreased flexibility of backbone residues.

In addition to the clustering performed for the glycosylated mini mucins described above, we also performed PCA-guided

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

kmeans clustering for all unglycosylated mini mucin simulations (Figure 4A, Supporting Information Figures S5−S7). Here, we see a significantly different array of representative cluster shapes. Comparing clusters between unglycosylated and glycosylated mini mucin simulations, Figure 4A, unglycosylated clusters are far more compact, nonlinear, and compressed with scores ranging from 0.08 to 0.49 with a median score of 0.16. In contrast, glycosylated mini mucin clusters have linearity scores ranging from 0.21 to 0.69 with a median of 0.51. Indeed, for unglycosylated mini mucin simulations, 7 out of 11 clusters have low linearity scores (<0.2) indicating highly compressed states, while none of the glycosylated mini mucin clusters have such low linearity scores. Despite their propensity to adopt more compressed or curved clusters, unglycosylated mini mucin simulations did still populate linear conformations with unglycosylated cluster 10 having a linearity score of 0.49, comparable to that of glycosylated clusters (Supporting Information Figure S7). Indeed, while 60.0% of unglycosylated mini mucin frames did fall in clusters with low linearity scores (<0.2), 24.8% of unglycosylated frames fell in clusters with >0.3 linearity scores, and 13.0% of frames in cluster 10 with a moderate (0.48) linearity score. Our data suggest that *O*-glycans impose steric limitations that prevent compression of the mucin backbone, reducing the probability of accessing more of the compact shapes seen in unglycosylated counterparts. These results are in striking agreement with those from Bousheri et al. demonstrating that *O*-glycans prevent compression along the otherwise intrinsically disordered protein backbone of lubricin.[61]

For each simulation, glycosylated, unglycosylated, mini, and mega, we also calculated the persistence length (PL) along the mucin protein backbone and observed a drastic difference in PLs between unglycosylated and glycosylated models: unglycosylated mini simulations with a PL of 7.4 ± 0.7 Å, glycosylated mini simulations with a PL of 21.3 ± 6.0 Å, (Figure 4B and Supporting Information Figures S2 and S5). Thus, the presence of *O*-glycans along mucin backbones increases PL by a factor of 2−3 compared to unglycosylated systems. We also tracked several time correlated data (Figures S8−S11) over all mini mucin simulations, including relative end-to-end distances, root-mean-square deviations, and relative radius of gyrations for all and C_α atoms. All timeseries reveal dramatic differences in the distributions between unglycosylated and glycosylated mini mucin systems (Figure S12), particularly in the case of relative end-to-end distances. Unglycosylated mini mucins demonstrate a broad distribution of states from fully extended (relative end-to-end distance, ∼1.0) to fully compressed (relative end-to-end distance, ∼0.1) with a mean of 0.42 ± 0.19. In contrast, glycosylated mini mucins demonstrate a sharper distribution of states with a mean of 0.68 ± 0.13. For each mini mucin and per replica, we calculated the percent of frames wherein the mini mucin conformation was found to be significantly compressed, with a relative end-to-end distance < 0.5. Unglycosylated mucins spend 67.7% ± 6.6% of frames in such compressed conformations, whereas glycosylated mucins spend only 10.9% ± 9.2% of frames in such conformations (Figure S12).

Upon observing representative structures from clustering as well as trajectory frames, we also noted many unglycosylated simulation conformations adopted "curved" and "pinched" structures, with significant backbone orientation changes around individual protein residues. For example, see the representative structures of unglycosylated mini mucin clusters 0−4: in these representative structures, there are distinct "pinch points". To quantify the probability of each amino acid along each mini mucin (glycosylated and unglycosylated) adopting such a "pinched" conformation, we calculated the distributions of angles formed between each $C_{\alpha-1}$, C_α, and $C_{\alpha+1}$ (excluding terminal C_αs), Figure 4D. We see that unglycosylated mini mucin systems have a distinct global and local energetic minimum (i.e., two population maxima) for C_α angles, while glycosylated mini mucin systems have a distinct global energetic minimum, with a smaller shoulder overlapping with the local minimum seen in unglycosylated simulations. α-helical turns are characterized by ∼100° rotation per amino acid, i.e., each complete turn of the α helix is occupied by 3.6 amino acids. This ∼100° rotation angle corresponds to the local energetic maximum wherein unglycosylated mini mucin C_αs have a decreased population. As such we demarcated ∼100° as a "pinch point": C_αs with conformations below 100° were to be considered "pinched." We then calculated the average number of pinched C_αs per frame: unglycosylated mini mucins exhibit an average of 5.1 ± 2.2 pinched C_αs, while glycosylated mini mucins exhibit an average of only 2.1 ± 1.4 pinched C_αs per frame. Thus, unglycosylated mini mucins demonstrate >2× more pinched C_αs than glycosylated mini mucins and, on average, have enough "pinch points" to generate nearly circular backbone conformations. Per residue distributions of these C_α angles identify residues which are more/less likely to adopt a pinched conformation. Such pinching conformations in glycosylated mini mucin simulations are more likely for nonglycosite residues and residues outside of densely glycosylated regions (Supporting Information Figures S13 and S14), as these residues on average have broader C_α angle distributions or bimodal distributions suggesting distinct conformational states.

To isolate which components within each mini mucin structure are relatively more or less rigid, we calculated root-mean-square fluctuations (RMSFs) and relative radii of gyration (R_gs) for protein residues and glycan segments separately. To calculate relative R_gs, we calculated first the R_g for the minimized (outstretched) protein backbone, and then for each frame of the simulation, we tracked the ratio of the framewise R_g to the minimized R_g. To calculate the glycan relative R_g, we similarly calculated the R_g for each glycan in the minimized structure and then per-frame calculated the ratio of framewise R_g to the minimized R_g. From these relative R_gs, we see that each glycan is relatively rigid with a sharp distribution in relative R_gs around 0.98 ± 0.06, Figure 4E, see also Supporting Information Figure S15 for per glycan average relative R_g. The distribution of relative R_gs for the glycosylated mucins was a bit broader but still high, with an average of 0.78 ± 0.05. However, the distribution of relative R_gs for unglycosylated mucins is much lower and broader, with an average of 0.49 ± 0.13. Furthermore, from RMSF calculations, we see that glycosylated mucin protein backbones as well as *O*-glycans are far more rigid relative to their unglycosylated counterparts, Figure 4F. Additionally, some *O*-glycans are less flexible than the protein backbone itself in glycosylated mucin simulations. All our results underscoring the rigidity and linearity of mucin protein backbones are supported rigorously by Kramer et al.[87] who demonstrated the increased rigidity of mucins as a function of *O*-glycan density. Our results paint hitherto unseen details of the mucin "bottlebrush" structure: dynamics of a flexible, otherwise intrinsically disordered

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

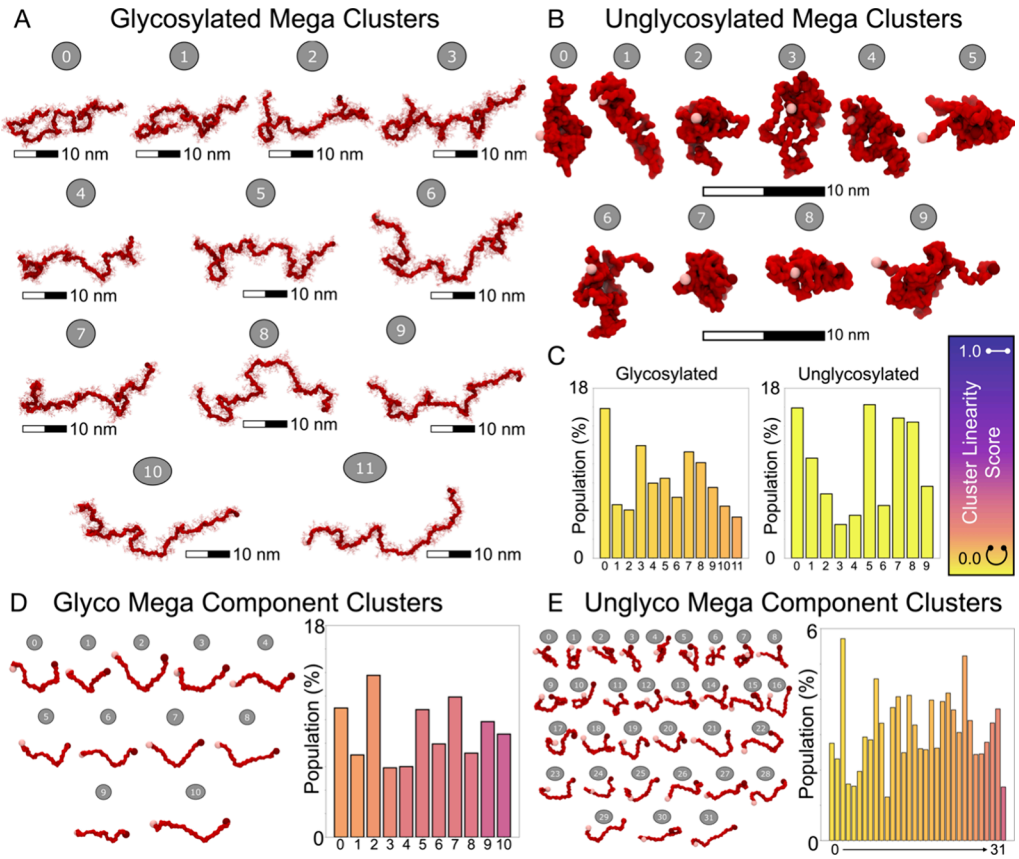


Figure 5. Relationship between size, glycosylation, and compactness of mega mucins. Representative structures from resultant clusters for (A) glycosylated and (B) unglycosylated mega mucin simulations. (C) Populations per clusters for glycosylated and unglycosylated mega mucin simulations, bars are colored according to each cluster's median linearity score. Representative structures and population distributions from clusters for component pieces of (D) glycosylated and (E) unglycosylated mega mucin simulations. Cluster linearity score bar is accompanied by small graphics denoting near linear conformations (~1.0) and nonlinear conformations (~0.0).

protein backbone are significantly dampened by the rigid/inflexible *O*-glycans.

Detailing the Physical Characteristics of a Large-Scale Bottlebrush. To understand the relationship between mucin model length and the nature of the "bottlebrush", we constructed a larger, more biologically relevant "mega" glycosylated mucin model, Figure S1C, along with a corresponding unglycosylated mega mucin counterpart, Figure S1D. The resulting glycosylated mega mucin structure was 224 amino acids in length with 128 *O*-glycans. While this model is still far from the true scale of MUC5B, which is ~6000 amino acids long, simulations of our mega mucin models will hopefully reveal conformational restraints imposed at the ~100 nm scale. In total, including conventional and GAMD sampling, we achieved ~2.7 μs of aggregate sampling for glycosylated mega mucins (Table S3) and ~30 μs of aggregate sampling for unglycosylated mega mucin simulations (Table S4). For all mega mucin simulations, we then applied similar analysis methods to those done for mini mucin simulations, PCA guided kmeans clustering (Supporting Information section 1.3.1, Table S5, Figures S16−S19), and timeseries analyses of relative end-to-end distances (Figures S20 and S21), relative R_gC_α (Figures S22 and S23), relative R_g for all atoms (Figures S24 and S25), and root-mean-square deviations (Figures S26 and S27). From clustering results, we see, once again, that glycosylated mega mucins largely maintain outstretched, albeit contorted, conformations, Figure 5A. Conversely, simulations of unglycosylated mega mucins saw dramatic and rapid compression along the protein backbone and resultant clustered unglycosylated mega mucin conformations represent globular, highly compressed, pseudofolded states, Figure 5B. We also calculated distributions of relative end-end distances and relative R_g for mega mucin simulations, Figure S28; however, from these distributions, we cannot determine a suitable "cutoff" for determining compressed mega conformations as all distributions fell below 0.5. We attribute this to the fact that the longer protein backbone presents many more opportunities for twisting and rotation that decrease relative end-end distance which then overestimates compression compared to the fully outstretched and minimized state. This artifact can be seen in the calculated RLSs shown as bar colors in Figure 5C, where the outstretched conformations (Figure 5A) of glycosylated mega mucin cluster representative structures are still calculated as having low linearity. This result demonstrates a need within the biophysics community to develop applicable metrics for defining and describing mucin conformations. We also calculated persistence lengths for our glycosylated and unglycosylated mega mucin models and saw striking agreement relative to our mini mucin simulations, Figure 4B: 21.3 ± 6.0 Å vs 15.3 ± 0.9 Å, for glycosylated minis vs mega mucin simulations, respectively. There was also marked agreement in persistence length calculations between mini and mega unglycosylated systems: 7.4 ± 0.7 Å versus 6.6 ± 0.4 Å. These results suggest that intrinsic interactions along

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

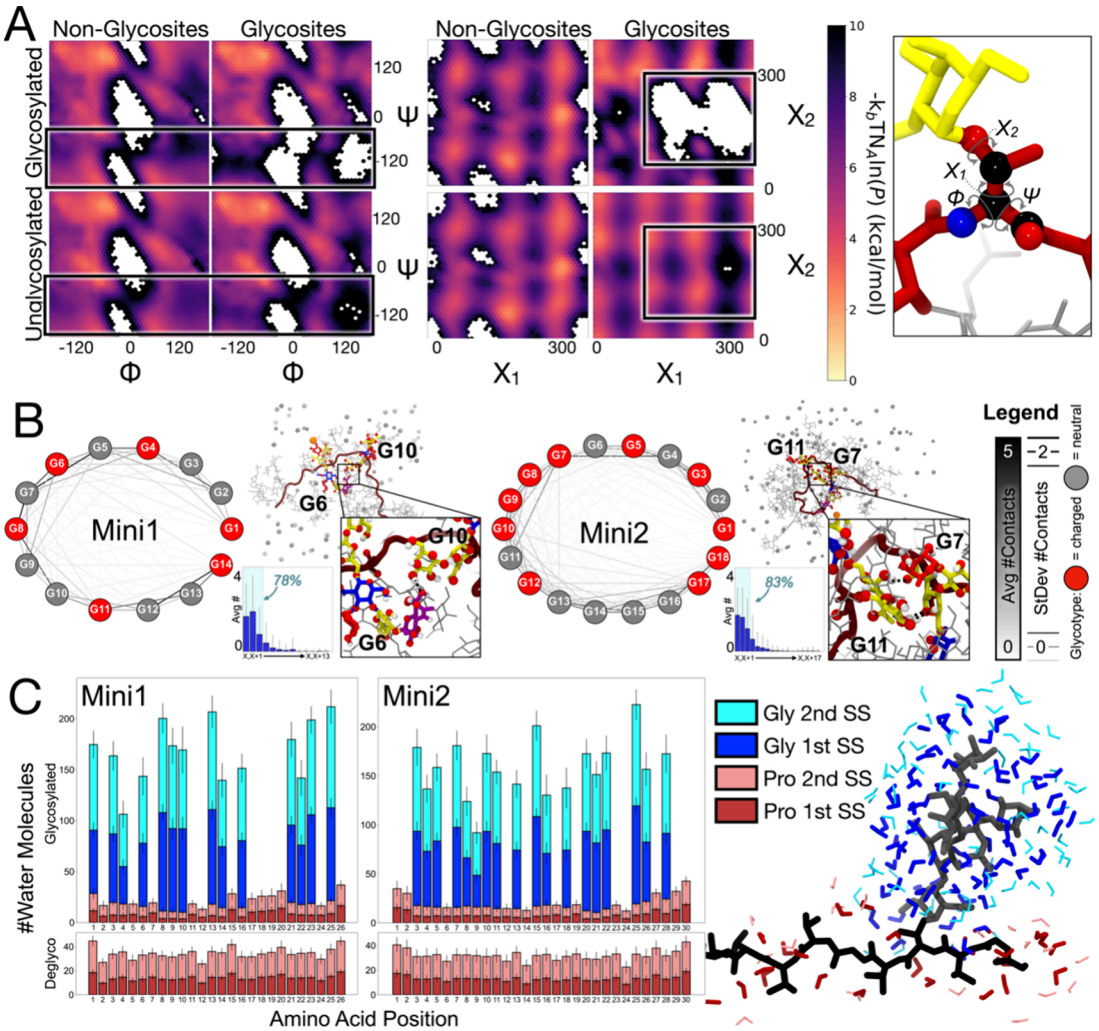


Figure 6. Elucidating the bottlebrush. (A) Two-dimensional histograms denoting Ramachandran (columns 1,2) and Janin (columns 3,4) plots for (top row) glycosylated mini mucins and (bottom row) unglycosylated mini mucins. Within the sets of Ramachandran and Janin plots, the first column is all Φ/Ψ dihedral distributions for nonglycosite residues, the second column is all Φ/Ψ dihedral distributions for glycosite residues. 2D-hexagonal histogram bins are colored according to population density on a kcal/mol scale. A molecular graphic is shown on the right to demonstrate which torsions are considered (mucin protein backbone atoms in red; carbon, nitrogen, and oxygen, atoms are highlighted in black, blue, and red spheres, respectively, *O*-glycan atoms highlighted in yellow, neighboring glycan atoms shown in light gray). (B) Network graphs highlighting the average number of glycan−glycan contacts (line color) and standard deviation in this average (line width) for Mini1 (left) and Mini2 (right). Charged glycans are represented with red circles, neutral glycans are represented with gray circles. Inset images show examples of contacts between "far" neighbor glycans, and inset bar graphs depict the average number of glycan−glycan contacts between each pair-type. (C) Left and middle panels: First and second water solvation shell counts for protein (pink and red) and glycans (cyan and blue). Right panel: Solvation shells considered in analysis: glycans (gray licorice), protein (black licorice), first and second solvation shells shown in shades of blue and red, respectively.

the protein backbone are captured and represented well at the mini mucin level and provide confidence in the mini mucins as bastions for mucin biophysics at smaller scales.

To further probe for similarities between conformations seen in mini and mega mucin simulations, we deconstructed our glycosylated and unglycosylated mega mucin simulations by selecting the 26aa segments along the backbone, saving these to independent trajectories, and then aligning and concatenating these independent component trajectories. As such, we effectively recreated glycosylated and unglycosylated mini mucin trajectories from the mega mucin trajectories and sought to determine if clusters and properties collected from the component segments of the larger mega mucins behaved similarly to the dynamics seen from free mini mucin simulations. We again performed PCA-guided kmeans clustering (see Supporting Information Methods Section 1.3.1 Table S6 and Figures S29−S32) and selected the representative structures from each cluster, (Figure 5D,E). Qualitatively, representative structures from glycosylated and unglycosylated component clusters look strikingly similar to representative structures from mini mucin clustering results, Figure 4A. Additionally, the range of calculated linearity scores per cluster is similar to those seen for the mini mucin clusters: glycosylated component clusters ranging in RLS from 0.27 to 0.53 with a median of 0.42, and unglycosylated component clusters ranging in RLS from 0.07 to 0.47 with a median of 0.17. So, while RLSs calculated for the clusters taken from whole mega mucin simulations (glycosylated megas range from 0.06 to 0.23, unglycosylated megas range from 0.003 to 0.014) are much lower than their mini mucin counterparts

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

(glycosylated minis range from 0.21 to 0.67, unglycosylated minis range from 0.08 to 0.49) this is likely simply due to end-end distances for mega mucins being far more variable: small deviations from linearity on the mini mucin scale can compound to larger contortions on the mega mucin scale that dramatically reduce the end-end distance ratio. We also see that the mega mucin component trajectories demonstrate remarkably similar distributions of relative end-to-end distances and relative R_gC_α compared to those collected for the mini mucin simulations, Figure S28G,H. Taken together, these component segment analyses from mega mucin simulations reveal that our mini mucin simulations do behave similarly to their counterparts within the larger mucin constructs. Said differently, from an MD simulations perspective and in the case of MUC5B, our mini mucin "parts" do make up the mega mucin "whole", suggesting that these mini mucin models may be more accessible alternatives to exploring mucin dynamics and interactions at the all-atom scale.

Elucidating the Forces that Imbue Structure within the Bottlebrush. We next sought to characterize the biophysical relationships between protein and *O*-glycan residues that result in pseudostructural properties within the bottlebrush domain. To do so, we tracked several degrees of freedom from glycosylated and unglycosylated mini mucin simulations. First, we calculated Ramachandaran and Janin plots for mini mucin models in glycosylated and unglycosylated states, and we plotted those degrees of freedom for nonglycosites (i.e., residues that are not glycosylated in glycosylated simulations, thus not Ser and Thr residues) and glycosites (i.e., Ser and Thr residues that would be glycosylated in glycosylated simulations, but which are not glycosylated in unglycosylated mucin simulations), Figure 6A. These results demonstrate a stark difference in torsional degrees of freedom for the protein backbone in the glycosylated and unglycosylated states. We see that glycosite residues in glycosylated mucin simulations have a loss of probability density for Ψ backbone torsions in the range of ∼−120° to ∼−50° compared to those same residues in the unglycosylated simulations and compared to nonglycosites in glycosylated simulations. We see that from the glycosylated simulations, nonglycosites occupy Ψ values in the range of −180° to −60° in ∼10% of frames. However, in these same simulations, glycosites do not occupy this state (population of ∼0% calculated through the area under the curve across Ψ kernel density function, see Supporting Information Figure S33 for 1D densities). Conversely, glycosites largely populate the 60° to 180° range of Ψ angles at 89% population versus 78% for nonglycosites. Nonglycosites and glycosites both occupy the −60° to 60° Ψ domain in approximately 9% of frames. Thus, *O*-glycosylation dramatically alters the rotatable Ψ degrees of freedom accessible to mucin protein backbone conformations.

Furthermore, with the addition of an *O*-glycan on a Ser or Thr residue, glycosylated Ser and Thr now have a nontrivial $\chi 2$ torsional potential and thus Janin plots for such glycosites in glycosylated mini mucin simulations are drastically different compared to those same residues in unglycosylated mini mucin simulations. The addition of a true $\chi 2$ results in a dramatic reduction of backbone torsional degrees of freedom. We compare these Janin plots to nonglycosite residues (i.e., not Ser or Thr) with true $\chi 2$ dihedrals from our Mini mucin simulations, specifically a His and Leu in each mini mucin. As can be seen, not only do glycosite residues demonstrate dramatically different Janin plots between glycosylated and unglycosylated mini mucin simulations (as to be expected as Ser and Thr do not have true $\chi 2$s when unglycosylated) but they also have dramatically different Janin distributions relative to His and Leu in glycosylated and unglycosylated mucin simulations. These factors demonstrate the reduced degrees of freedom accessible to mucin protein backbones due to the *O*-glycosylation. All our Ramachandran and Janin plots agree remarkably with NMR and early restrained minimization modeling data from Coltart et al.[88] who demonstrated that modification by the first *O*-GalNAc residue was sufficient to impose restraints on protein backbone conformational torsions.

We also calculated the extent of glycan−glycan contacts (defined as the number of residues from one glycan within 5 Å of another glycan) within all glycosylated mini mucin simulations (Figure 6B). Contact networks show the average number of contacts as well as the standard deviation in this average between all glycans in each mini mucin. Overall, we see that most glycans do make contact with most other glycans within each mini mucin. However, we see that the strongest degrees of glycan−glycan contacts are those within the "nearest neighbor" range, inset bar graphs in Figure 6B, i.e., >75% of all glycan−glycan contacts within each mini mucin system are within the first 3 nearest neighbors (see Supporting Information Figure S34 for full integration plots). We hypothesize that these nearest neighbor glycan−glycan contacts underlie the increasing persistence length exhibited by the glycosylated mucin systems compared with unglycosylated mucin systems.

Solvation Promotes Extended Mucin Conformations. Kramer et al.[87] hypothesized the solvation shell of bulky *O*-glycans may contribute to steric hindrance that enforces extended mucin protein backbones. Following this line of reasoning, we sought to characterize the density of solvation shells surrounding glycosylated mini mucins and unglycosylated mini mucins during simulation. To do so, for each frame in simulation, we calculated the number of water molecules within the first and second solvation shells (3.4 and 5 Å), respectively, of glycan segments or protein residues (Figure 6C, Supporting Information Figure S35). The average number of water molecules in the first and second solvation shells of protein residues in unglycosylated mini systems (protein-only) was 13 ± 4 and 32 ± 9, respectively, with variability in this average stemming largely from residue size (i.e., per residue values have much lower variability). By comparison, the number of water molecules within the first and second solvation shells of mucin residues (i.e., considering *O*-glycans) was 65 ± 40 and 123 ± 72, respectively (the average number of water molecules within the first and second solvation shells of protein residues within glycosylated mucins was 8 ± 4 and 19 ± 9, respectively). Thus, the presence of bulky *O*-glycans results in a larger solvation shell, in particular, an order of magnitude increase in the number of water molecules within the first solvation shell in the glycosylated mucin compared to the unglycosylated mucin.

We additionally sought to characterize the dynamics of the water molecules around the mini mucin models in glycosylated and unglycosylated states. Water molecules within the first solvation shell of glycosylated mucins have a residence time of 70 ± 3 ps while water molecules within the first solvation shell of unglycosylated mini mucins have a residence time of 57 ± 4 ps (Figure S36A); thus, water molecules around glycosylated

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

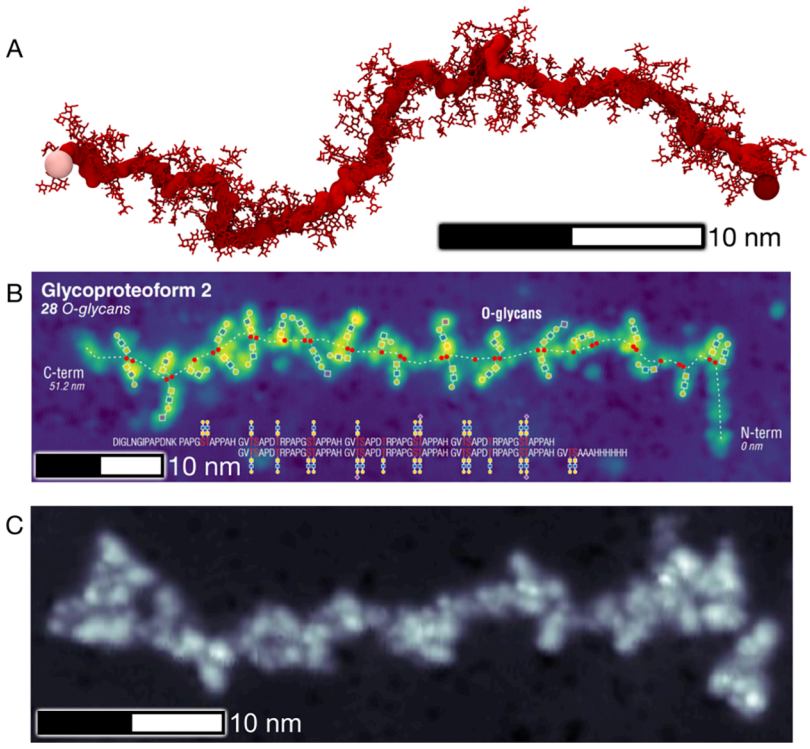


Figure 7. Direct comparisons of computational mega mucin simulation snapshot to experimentally derived 2-dimensional images of MUC1. (A) 148 amino acid section of glycosylated mega mucin taken from simulation, rendered with protein in red surface and *O*-glycans in red licorice (hydrogens not shown for clarity). The first amino acid is highlighted with a pink sphere, the last amino acid is highlighted with a maroon sphere. (B) Scanning tunneling microscopy image of soft-landed electrospray ion beam deposited MUC1, annotated by Anggara et al.[59] (C) Another representative scanning tunneling microscopy image from Anggara et al.[59] demonstrating additional qualitative similarity to (A).

mucins remain bound for on average ∼10 ps longer. Furthermore, mean square displacement (MSD) calculations reveal that water molecules within the first solvation shell of glycosylated mucins move faster on average than water molecules within the first solvation shell of unglycosylated mucin models ($1.5 \pm 0.1 \times 10^{-7}$ m^2/s and $1.1 \pm 0.2 \times 10^{-7}$ m^2/s, respectively Figure S36B). Interestingly, water molecules within the first solvation shell of glycosylated mucins move on average as fast ($1.5 \pm 0.1 \times 10^{-7}$ m^2/s) as water molecules in bulk conditions ($1.50 \pm 0.02 \times 10^{-7}$ m^2/s) while water molecules in the first solvation shell of unglycosylated mucins move more slowly ($1.1 \pm 0.2 \times 10^{-7}$ m^2/s) than waters in the bulk (Figure S36B). We hypothesize that the difference in water molecule behavior within first solvation shells of glycosylated and unglycosylated mucin and bulk conditions may be due to the number of available hydrogen bonding partners supported by *O*-glycans. Mucin *O*-glycans each have many hydroxyl and charged moieties that can each donate or accept hydrogen bonds, and thus water molecules can exchange, translate, and rotate more readily without significantly sacrificing hydrogen bonding partners. Thus, extended mucin protein backbones allow for the maximization of such hydrogen bonding partners with water molecules. However, unglycosylated mucins are relatively more hydrophobic in nature; each unglycosylated Ser and Thr can accept only one hydrogen bond and donate one hydrogen bond, but the remaining residues are largely hydrophobic (Gly, Pro, Leu, and Val). Thus, for the case of the unglycosylated mucins, we hypothesize that the hydrophobic effect contributes to protein collapse, in turn minimizing solvent-accessible surface area. Similarly, *O*-glycans may simply increase the overall solubility of, as discussed, otherwise hydrophobic mucin peptide backbones, and outstretched *O*-glycan conformations are more effective at achieving this improved solubility. The role of *O*-glycans in impacting or increasing the solubility of these potentially otherwise hydrophobic mucin peptide backbones will be explored extensively in future work.

Collectively, our results suggest that the bottlebrush phenomenon is derived from (1) reduced torsional degrees of freedom along the protein backbone due to glycan-induced steric hindrance, (2) nearest neighbor glycan contacts which amplify reduction in torsional potentials, and (3) the extension of the solvation shell which further amplifies glycan steric hindrance as well as maximization of hydrogen bonding potential to solvent.

Mega Mucin Model Captures Flexibility Demonstrated from Smaller Simulations while Recapitulating Experiment. Finally, we sought to qualitatively compare our mega mucin model simulations to experimentally characterized images of physiological mucins. Toward this aim, Anggara et al. graciously provided us with 20 scanning tunneling microscopy images of soft-landing electrospray ion beam deposited MUC1 tandem repeat fragment (6.5 tandem repeats around 130 amino acids) produced in glycoengineered cells (Figure S37).[54,59] Each STM image provides a glimpse of a single MUC1 molecule. Although MUC1 and MUC5B are different in both PTS domain tandem protein backbone repeats and specific *O*-glycoprofiles, the qualitative similarities between a snapshot from our model simulations (Figure 7A) and these experimentally acquired images, such as examples in Figure 7B,C, are too remarkable to ignore. To attempt a quantitative comparison between our simulations and their experimental

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

images, but needing to account for different protein backbone lengths, we calculated a normalized height-per-residue from our glycosylated mega mucin simulations and Anggara et al's graciously provided and annotated images. For our MUC5B simulations, we calculated a normalized height-per-residence of 1.2 ± 0.3 Å, and from Anggara et al.'s MUC1 STM images, we calculated a normalized height-per-residue of 4.0 ± 0.7 Å, Figure S38.[59] Although there is a large discrepancy in these values calculated between simulation and experiment, we believe this can be largely attributed to the differences between MUC1 and MUC5B as well as the technique of macromolecule deposition for STM imaging wherein mucins are deposited on a surface and stretched to achieve maximum clarity on *O*-glycan imaging. However, our attempts to compare experiment and modeling results in this regard underscore the need to derive mucin-specific metrics that can be more readily correlated between the two realms.

CONCLUSIONS

From our mini mucin simulations, we have delineated several key degrees of freedom that impart physical structure to otherwise intrinsically disordered mucin protein backbones: (1) decreased backbone torsional degrees of freedom, (2) strong nearest-neighbor contacts between *O*-glycans in densely glycosylated mucins, and (3) dramatically increased first and second solvation shells from the introduction of hydrophilic glycans. Each of these degrees of freedom can be traced to the additional bulk imposed by the presence of *O*-glycans along the mucin protein backbone. Furthermore, the internally consistent persistence length calculated for mini and mega mucin models, as well as qualitative comparisons between mini mucin simulations and component structures within mega mucins, demonstrates internal consistency and strengthens our confidence in using such mini mucin models to inexpensively approximate mucin biochemical properties and mechanics, for example, when analyzing overall interaction profiles between mucin and other proteins. In addition to validating our mini mucin models, our Mega mucin model provides a striking qualitative comparison to experimentally derived images of MUC1. Further, the development of small (~200 amino acid) mucin reporter constructs with tailored *O*-glycan positions, as done by Nason et al.,[19,54] allows for potential collaborative exploration via experiment and simulation with a reasonable overlap in mucin sequence and structure. Similarly, the expansion of glycomics methods, as done by the development of mucinases and LC-MS/MS techniques, will allow for increasing accuracy in *O*-glycan structures, which can then feedback into the workflow for increased predictive power in mucin modeling.[56,60] Finally, the development of novel visualization techniques for mucins will dramatically strengthen relationships between mucin simulation and experiment.[59]

We have demonstrated the construction and simulation of densely *O*-glycosylated mucin domains and how these simulations can be used to investigate the biochemical/biophysical nature of mucin domain glycoproteins. The simulations presented herein represent a starting point through which novel structure−function hypotheses can be explored in this domain. Due to their many roles, dysregulation and altered morphology of mucins are implicated in many diseases including diabetes mellitus, many cancers, and cystic fibrosis. As such, understanding the basic nature of mucins and their *O*-glycans in imparting structure to the mucosal layer is vital to understanding disease pathogenesis and progression. Future investigations into the molecular mechanisms of these diseases will likely require an in-depth, atomic-scale, structural understanding of mucin glycoproteins. Given the vast diversity within the mucinome, fully understanding the generalizable relationships between mucin structure and protein and *O*-glycan sequences will require extensive research on a broader range of mucins. MD simulations of these mucins are likely to reveal overarching principles that govern mucin structure, an area where our current understanding is just beginning to take shape.

ASSOCIATED CONTENT

Data Availability Statement

Software used in this work: The NAMD and VMD molecular modeling simulation and visualization tools are available to noncommercial users under distribution specific licenses. CHARMM-GUI and RosettaTools are free of charge for academic licenses and available for commercial use with paid licenses. MDAnalysis, scipy, pandas, scikit-learn, and matplotlib are python packages free of charge. All software used in this work are referenced throughout the methods and results sections of this work. Structures, trajectories, and analysis scripts created during this work: All trajectories, structures, simulation scripts, analysis scripts, and data files will be shared with this work through a link hosted on the Amaro Lab Web site (https://amarolab.ucsd.edu/covid19.php).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00613.

Complete methods for construction and simulation of Mini1, Mini2, and mega mucin models considered in this work, along with complete description of all analyses described; schemes graphically detailing simulation methods and equilibration times per model; time-correlated values (relative end-end distances, radii of gyration, root-mean-square deviations) plotted for all simulations demonstrating equilibration and sampling conditions; clustering statistics and details for all models; relative end-end distances; distributions from simulation of properties discussed throughout manuscript including relative end-end distances, radius of gyration, curvature, C_α angles, and $\varphi/\psi/\chi1/\chi2$ distributions; detailed contact plots for glycan−glycan contacts; detailed plots showing solvation shell size; supporting scanning-tunneling microscopy images for MUC-1 graciously provided by Anggara et al.[59] for comparison to mucin models simulated in this work (PDF)

AUTHOR INFORMATION

Corresponding Author

Rommie E. Amaro − *Department of Molecular Biology, University of California San Diego, La Jolla, California 92093-0340, United States;* orcid.org/0000-0002-9275-9553; Email: ramaro@ucsd.edu

Authors

Fiona L. Kearns − *Department of Molecular Biology, University of California San Diego, La Jolla, California 92093-0340, United States;* orcid.org/0000-0002-5469-9035

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

Mia A. Rosenfeld − *National Institute of Health, National Heart, Lung & Blood Institute, Bethesda, Maryland 20892, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c00613

Funding

M.A.R. is supported by the NHLBI Vaughan Fellowship and the NIH Undergraduate Scholarship (NIH UGSP). This work was also funded in part by NIH GM132826 to REA and NSF CHE060073N.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Kelvin Anggara and Rebecca Miller for their generosity in allowing us to use their STM images of MUC1, and Thapakorn Jaroentomeechai and Yoshiki Narimatsu who designed and produced the MUC1 mucin reporter used in that work. We also thank Elisa Fadda, Ronit Freeman, Stacy Malaker, Kelvin Anggara, Rebecca Miller, and Henrik Clausen (and their laboratories) for invaluable discussions and insights into mucin glycobiology. We acknowledge the San Diego Supercomputing Center (SDSC) for providing HPC resources that have contributed to the research results reported within this paper.

■ ABBREVIATIONS

SNFG:symbol nomenclature for glycans
MD simulations:molecular dynamics simulations
MUC5B:canonical mucin 5B
PL:persistence length
Rg:radius of gyration
RMSF:root mean square fluctuation
RMSD:root mean square deviation
RLS:relative linearity score
glycosylated Mini1:miniature mucin model #1, simulations performed with glycosylation
glycosylated Mini2:miniature mucin model #2, simulations performed with glycosylation
unglycosylated Mini1:miniature mucin model #1, simulations performed of just protein backbone
unglycosylated Mini2:miniature mucin model #2, simulations performed of just protein backbone
glycosylated mega:large mucin model, simulations performed with glycosylation
unglycosylated mega:large mucin model, simulations performed of just protein backbone
STM:scanning tunneling microscopy

■ REFERENCES

(1) Holmén, J. M.; Karlsson, N. G.; Abdullah, L. H.; Randell, S. H.; Sheehan, J. K.; Hansson, G. C.; Davis, C. W. Mucins and Their O-Glycans from Human Bronchial Epithelial Cell Cultures. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2004**, *287* (4), L824−L834.

(2) Malaker, S. A.; Riley, N. M.; Shon, D. J.; Pedram, K.; Krishnan, V.; Dorigo, O.; Bertozzi, C. R. Revealing the Human Mucinome. *Nat. Commun.* **2022**, *13* (1), 3542.

(3) Brockhausen, I.; Schachter, H.; Stanley, P. *O-GalNAc Glycans*; 2009.

(4) Hughes, G. W.; Ridley, C.; Collins, R.; Roseman, A.; Ford, R.; Thornton, D. J. The MUC5B Mucin Polymer Is Dominated by Repeating Structural Motifs and Its Topology Is Regulated by Calcium and PH. *Sci. Rep* **2019**, *9* (1), 17350.

(5) Symmes, B. A.; Stefanski, A. L.; Magin, C. M.; Evans, C. M. Role of Mucins in Lung Homeostasis: Regulated Expression and Biosynthesis in Health and Disease. *Biochem. Soc. Trans.* **2018**, *46* (3), 707−719.

(6) Thomsson, K. A.; Schulz, B. L.; Packer, N. H.; Karlsson, N. G. MUC5B Glycosylation in Human Saliva Reflects Blood Group and Secretor Status. *Glycobiology* **2005**, *15* (8), 791−804.

(7) *Essentials of Glycobiology*, 3rd ed.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).

(8) Corfield, A. P. Mucins: A Biologically Relevant Glycan Barrier in Mucosal Protection. *Biochim. Biophys. Acta* **2015**, *1850* (1), 236−252.

(9) Hansson, G. C. Mucins and the Microbiome. *Annu. Rev. Biochem.* **2020**, *89*, 769−793.

(10) Hollingsworth, M. A.; Swanson, B. J. Mucins in Cancer: Protection and Control of the Cell Surface. *Nat. Rev. Cancer* **2004**, *4* (1), 45−60.

(11) Moniaux, N.; Escande, F.; Porchet, N.; Aubert, J. P.; Batra, S. K. Structural Organization and Classification of the Human Mucin Genes. *Front. Biosci.* **2001**, *6*, D1192-206.

(12) Sheehan, J. K.; Oates, K.; Carlstedt, I. Electron Microscopy of Cervical, Gastric and Bronchial Mucus Glycoproteins. *Biochem. J.* **1986**, *239* (1), 147−153.

(13) Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide GalNAc-Transferase Gene Family. *Glycobiology* **2012**, *22* (6), 736−756.

(14) de Las Rivas, M.; Lira-Navarrete, E.; Gerken, T. A.; Hurtado-Guerrero, R. Polypeptide GalNAc-Ts: From Redundancy to Specificity. *Curr. Opin Struct Biol.* **2019**, *56*, 87−96.

(15) Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T.-B. G.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; Marcos-Silva, L.; Gupta, R.; Bennett, E. P.; Mandel, U.; Brunak, S.; Wandall, H. H.; Levery, S. B.; Clausen, H. Precision Mapping of the Human O-GalNAc Glycoproteome through SimpleCell Technology. *EMBO J.* **2013**, *32* (10), 1478−1488.

(16) Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. Vertebrate Protein Glycosylation: Diversity, Synthesis and Function. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (7), 448−462.

(17) Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H. Global View of Human Protein Glycosylation Pathways and Functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (12), 729−749.

(18) Duca, M.; Malagolini, N.; Dall'Olio, F. The Mutual Relationship between Glycosylation and Non-Coding RNAs in Cancer and Other Physio-Pathological Conditions. *Int. J. Mol. Sci.* **2022**, *23* (24), 15804.

(19) Konstantinidi, A.; Nason, R.; Čaval, T.; Sun, L.; Sørensen, D. M.; Furukawa, S.; Ye, Z.; Vincentelli, R.; Narimatsu, Y.; Vakhrushev, S. Y.; Clausen, H. Exploring the Glycosylation of Mucins by Use of O-Glycodomain Reporters Recombinantly Expressed in Glycoengineered HEK293 Cells. *J. Biol. Chem.* **2022**, *298* (4), No. 101784.

(20) Hill, H. D.; Reynolds, J. A.; Hill, R. L. Purification, Composition, Molecular Weight, and Subunit Structure of Ovine Submaxillary Mucin. *J. Biol. Chem.* **1977**, *252* (11), 3791−3798.

(21) Hill, H. D.; Schwyzer, M.; Steinman, H. M.; Hill, R. L. Ovine Submaxillary Mucin. Primary Structure and Peptide Substrates of UDP-N-Acetylgalactosamine:Mucin Transferase. *J. Biol. Chem.* **1977**, *252* (11), 3799−3804.

(22) Gerken, T. A.; Owens, C. L.; Pasumarthy, M. Site-Specific Core 1 O-Glycosylation Pattern of the Porcine Submaxillary Gland Mucin Tandem Repeat. *J. Biol. Chem.* **1998**, *273* (41), 26580−26588.

(23) Gerken, T. A.; Gilmore, M.; Zhang, J. Determination of the Site-Specific Oligosaccharide Distribution of the O-Glycans Attached to the Porcine Submaxillary Mucin Tandem Repeat. *J. Biol. Chem.* **2002**, *277* (10), 7736−7751.

(24) Gerken, T. A.; Tep, C.; Rarick, J. Role of Peptide Sequence and Neighboring Residue Glycosylation on the Substrate Specificity of the Uridine 5′-Diphosphate−α- *N* -Acetylgalactosamine:Polypeptide *N* -Acetylgalactosaminyl Transferases T1 and T2: Kinetic Modeling of the Porcine and Canine Submaxillary Gland Mucin Tandem Repeats. *Biochemistry* **2004**, *43* (30), 9888−9900.
(25) Gerken, T. A.; Owens, C. L.; Pasumarthy, M. Determination of the Site-Specific O-Glycosylation Pattern of the Porcine Submaxillary Mucin Tandem Repeat Glycopeptide. *J. Biol. Chem.* **1997**, *272* (15), 9709−9719.
(26) Paz, H. B.; Tisdale, A. S.; Danjo, Y.; Spurr-Michaud, S. J.; Argüeso, P.; Gipson, I. K. The Role of Calcium in Mucin Packaging within Goblet Cells. *Exp. Eye Res.* **2003**, *77* (1), 69−75.
(27) Rose, M. C.; Voynow, J. A. Respiratory Tract Mucin Genes and Mucin Glycoproteins in Health and Disease. *Physiol Rev.* **2006**, *86* (1), 245−278.
(28) Zhang, N.; Van Crombruggen, K.; Gevaert, E.; Bachert, C. Barrier Function of the Nasal Mucosa in Health and Type-2 Biased Airway Diseases. *Allergy* **2016**, *71* (3), 295−307.
(29) Carmicheal, J.; Atri, P.; Sharma, S.; Kumar, S.; Chirravuri Venkata, R.; Kulkarni, P.; Salgia, R.; Ghersi, D.; Kaur, S.; Batra, S. K. Presence and Structure-activity Relationship of Intrinsically Disordered Regions across Mucins. *FASEB J.* **2020**, *34* (2), 1939−1957.
(30) Hansson, G. C. Mucins and the Microbiome. *Annu. Rev. Biochem.* **2020**, *89* (1), 769−793.
(31) Linden, S. K.; Sutton, P.; Karlsson, N. G.; Korolik, V.; McGuckin, M. A. Mucins in the Mucosal Barrier to Infection. *Mucosal Immunol* **2008**, *1* (3), 183−197.
(32) Wardzala, C. L.; Wood, A. M.; Belnap, D. M.; Kramer, J. R. Mucins Inhibit Coronavirus Infection in a Glycan-Dependent Manner. *ACS Cent Sci.* **2022**, *8* (3), 351−360.
(33) Hollingsworth, M. A.; Swanson, B. J. Mucins in Cancer: Protection and Control of the Cell Surface. *Nat. Rev. Cancer* **2004**, *4* (1), 45−60.
(34) Ostedgaard, L. S.; Moninger, T. O.; McMenimen, J. D.; Sawin, N. M.; Parker, C. P.; Thornell, I. M.; Powers, L. S.; Gansemer, N. D.; Bouzek, D. C.; Cook, D. P.; Meyerholz, D. K.; Abou Alaiwa, M. H.; Stoltz, D. A.; Welsh, M. J. Gel-Forming Mucins Form Distinct Morphologic Structures in Airways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (26), 6842−6847.
(35) Wagner, C. E.; Wheeler, K. M.; Ribbeck, K. Mucins and Their Role in Shaping the Functions of Mucus Barriers. *Annu. Rev. Cell Dev Biol.* **2018**, *34* (1), 189−215.
(36) Kiwamoto, T.; Katoh, T.; Evans, C. M.; Janssen, W. J.; Brummet, M. E.; Hudson, S. A.; Zhu, Z.; Tiemeyer, M.; Bochner, B. S. Endogenous Airway Mucins Carry Glycans That Bind Siglec-F and Induce Eosinophil Apoptosis. *J. Allergy Clin. Immunol.* **2015**, *135* (5), 1329−1340.e9.
(37) Habte, H. H.; de Beer, C.; Lotz, Z. E.; Roux, P.; Mall, A. S. Anti-HIV-1 Activity of Salivary MUC5B and MUC7Mucins from HIV Patients with Different CD4 Counts. *Virol J.* **2010**, *7* (1), 269.
(38) Singh, P. K.; Hollingsworth, M. A. Cell Surface-Associated Mucins in Signal Transduction. *Trends Cell Biol.* **2006**, *16* (9), 467−476.
(39) Kufe, D. W. Mucins in Cancer: Function, Prognosis and Therapy. *Nat. Rev. Cancer* **2009**, *9* (12), 874−885.
(40) Jonckheere, N.; Van Seuningen, I. The Membrane-Bound Mucins: From Cell Signalling to Transcriptional Regulation and Expression in Epithelial Cancers. *Biochimie* **2010**, *92* (1), 1−11.
(41) van Putten, J. P. M.; Strijbis, K. Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. *J. Innate Immun* **2017**, *9* (3), 281−299.
(42) Bhatia, R.; Gautam, S. K.; Cannon, A.; Thompson, C.; Hall, B. R.; Aithal, A.; Banerjee, K.; Jain, M.; Solheim, J. C.; Kumar, S.; Batra, S. K. Cancer-Associated Mucins: Role in Immune Modulation and Metastasis. *Cancer and Metastasis Reviews* **2019**, *38* (1−2), 223−236.
(43) Morrison, C. B.; Markovetz, M. R.; Ehre, C. Mucins, and Cystic Fibrosis. *Pediatr. Pulmonol.* **2019**, *54* (S3), S84−S96.
(44) Kreda, S. M.; Davis, C. W.; Rose, M. C. CFTR, Mucins, and Mucus Obstruction in Cystic Fibrosis. *Cold Spring Harb Perspect Med.* **2012**, *2* (9), a009589−a009589.
(45) Ehre, C.; Ridley, C.; Thornton, D. J. Cystic Fibrosis: An Inherited Disease Affecting Mucin-Producing Organs. *Int. J. Biochem. Cell Biol.* **2014**, *52*, 136−145.
(46) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol* **2019**, *15* (6), 346−366.
(47) Chugh, S.; Gnanapragassam, V. S.; Jain, M.; Rachagani, S.; Ponnusamy, M. P.; Batra, S. K. Pathobiological Implications of Mucin Glycans in Cancer: Sweet Poison and Novel Targets. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2015**, *1856* (2), 211−225.
(48) Aithal, A.; Rauth, S.; Kshirsagar, P.; Shah, A.; Lakshmanan, I.; Junker, W. M.; Jain, M.; Ponnusamy, M. P.; Batra, S. K. MUC16 as a Novel Target for Cancer Therapy. *Expert Opin Ther Targets* **2018**, *22* (8), 675−686.
(49) Felder, M.; Kapur, A.; Gonzalez-Bosquet, J.; Horibata, S.; Heintz, J.; Albrecht, R.; Fass, L.; Kaur, J.; Hu, K.; Shojaei, H.; Whelan, R. J.; Patankar, M. S. MUC16 (CA125): Tumor Biomarker to Cancer Therapy, a Work in Progress. *Mol. Cancer* **2014**, *13* (1), 129.
(50) Das, S.; Batra, S. K. Understanding the Unique Attributes of MUC16 (CA125): Potential Implications in Targeted Therapy. *Cancer Res.* **2015**, *75* (22), 4669−4674.
(51) Gubbels, J. A.; Felder, M.; Horibata, S.; Belisle, J. A.; Kapur, A.; Holden, H.; Petrie, S.; Migneault, M.; Rancourt, C.; Connor, J. P.; Patankar, M. S. MUC16 Provides Immune Protection by Inhibiting Synapse Formation between NK and Ovarian Tumor Cells. *Mol. Cancer* **2010**, *9* (1), 11.
(52) White, B.; Patterson, M.; Karnwal, S.; Brooks, C. L. Crystal Structure of a Human MUC16 SEA Domain Reveals Insight into the Nature of the CA125 Tumor Marker. *Proteins: Struct., Funct., Bioinf.* **2022**, *90* (5), 1210−1218.
(53) Burchell, J. M.; Beatson, R.; Graham, R.; Taylor-Papadimitriou, J.; Tajadura-Ortega, V. O-Linked Mucin-Type Glycosylation in Breast Cancer. *Biochem. Soc. Trans.* **2018**, *46* (4), 779−788.
(54) Nason, R.; Büll, C.; Konstantinidi, A.; Sun, L.; Ye, Z.; Halim, A.; Du, W.; Sørensen, D. M.; Durbesson, F.; Furukawa, S.; Mandel, U.; Joshi, H. J.; Dworkin, L. A.; Hansen, L.; David, L.; Iverson, T. M.; Bensing, B. A.; Sullam, P. M.; Varki, A.; de Vries, E.; de Haan, C. A. M.; Vincentelli, R.; Henrissat, B.; Vakhrushev, S. Y.; Clausen, H.; Narimatsu, Y. Display of the Human Mucinome with Defined O-Glycans by Gene Engineered Cells. *Nat. Commun.* **2021**, *12* (1), 4070.
(55) Lu, L.; Riley, N. M.; Shortreed, M. R.; Bertozzi, C. R.; Smith, L. M. O-Pair Search with MetaMorpheus for O-Glycopeptide Characterization. *Nat. Methods* **2020**, *17* (11), 1133−1138.
(56) Malaker, S. A.; Pedram, K.; Ferracane, M. J.; Bensing, B. A.; Krishnan, V.; Pett, C.; Yu, J.; Woods, E. C.; Kramer, J. R.; Westerlind, U.; Dorigo, O.; Bertozzi, C. R. The Mucin-Selective Protease StcE Enables Molecular and Functional Analysis of Human Cancer-Associated Mucins. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (15), 7278−7287.
(57) Bagdonaite, I.; Malaker, S. A.; Polasky, D. A.; Riley, N. M.; Schjoldager, K.; Vakhrushev, S. Y.; Halim, A.; Aoki-Kinoshita, K. F.; Nesvizhskii, A. I.; Bertozzi, C. R.; Wandall, H. H.; Parker, B. L.; Thaysen-Andersen, M.; Scott, N. E. Glycoproteomics. *Nature Reviews Methods Primers* **2022**, *2* (1), 48.
(58) Oliveira, T.; Thaysen-Andersen, M.; Packer, N. H.; Kolarich, D. The Hitchhiker's Guide to Glycoproteomics. *Biochem. Soc. Trans.* **2021**, *49* (4), 1643−1662.
(59) Anggara, K.; Sršan, L.; Jaroentomeechai, T.; Wu, X.; Rauschenbach, S.; Narimatsu, Y.; Clausen, H.; Ziegler, T.; Miller, R. L.; Kern, K. Direct Observation of Glycans Bonded to Proteins and Lipids at the Single-Molecule Level. *Science* **1979**, *382* (6667), 219−223.
(60) Chongsaritsinsuk, J.; Steigmeyer, A. D.; Mahoney, K. E.; Rosenfeld, M. A.; Lucas, T. M.; Smith, C. M.; Li, A.; Ince, D.; Kearns, F. L.; Battison, A. S.; Hollenhorst, M. A.; Judy Shon, D.; Tiemeyer, K. H.; Attah, V.; Kwon, C.; Bertozzi, C. R.; Ferracane, M. J.; Lemmon,

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965

M. A.; Amaro, R. E.; Malaker, S. A. Glycoproteomic Landscape and Structural Dynamics of TIM Family Immune Checkpoints Enabled by Mucinase SmE. *Nat. Commun.* **2023**, *14* (1), 6169.
(61) Boushehri, S.; Holey, H.; Brosz, M.; Gumbsch, P.; Pastewka, L.; Aponte-Santamaría, C.; Gräter, F. O-Glycans Expand Lubricin and Attenuate Its Viscosity and Shear Thinning. *Biomacromolecules* **2024**, *25* (7), 3893−3908.
(62) Chatterjee, M.; van Putten, J. P. M.; Strijbis, K.; Garsin, D. A. Defensive Properties of Mucin Glycoproteins during Respiratory Infections—Relevance for SARS-CoV-2. *mBio* **2020**, *11* (6). .
(63) Ermund, A.; Meiss, L. N.; Rodriguez-Pineiro, A. M.; Bähr, A.; Nilsson, H. E.; Trillo-Muyo, S.; Ridley, C.; Thornton, D. J.; Wine, J. J.; Hebert, H.; Klymiuk, N.; Hansson, G. C. The Normal Trachea Is Cleaned by MUC5B Mucin Bundles from the Submucosal Glands Coated with the MUC5AC Mucin. *Biochem. Biophys. Res. Commun.* **2017**, *492* (3), 331−337.
(64) Bos, M. F.; Ermund, A.; Hansson, G. C.; de Graaf, J.; Reis, R. Goblet Cell Interactions Reorient Bundled Mucus Strands for Efficient Airway Clearance. *PNAS Nexus* **2023**, *2* (11). .
(65) Ermund, A.; Trillo-Muyo, S.; Hansson, G. C. Assembly, Release, and Transport of Airway Mucins in Pigs and Humans. *Ann. Am. Thorac. Soc.* **2018**, *15*, S159−S163.
(66) Trillo-Muyo, S.; Nilsson, H. E.; Recktenwald, C. V.; Ermund, A.; Ridley, C.; Meiss, L. N.; Bähr, A.; Klymiuk, N.; Wine, J. J.; Koeck, P. J. B.; Thornton, D. J.; Hebert, H.; Hansson, G. C. Granule-Stored MUC5B Mucins Are Packed by the Non-Covalent Formation of N-Terminal Head-to-Head Tetramers. *J. Biol. Chem.* **2018**, *293* (15), 5746−5754.
(67) Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O. Rosetta FlexPepDock Ab-Initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLoS One* **2011**, *6* (4), No. e18934.
(68) Alocci, D.; Mariethoz, J.; Gastaldello, A.; Gasteiger, E.; Karlsson, N. G.; Kolarich, D.; Packer, N. H.; Lisacek, F. GlyConnect: Glycoproteomics Goes Visual, Interactive, and Analytical. *J. Proteome Res.* **2019**, *18* (2), 664−677.
(69) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859−1865.
(70) Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**, *29* (4), 320−331.
(71) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(72) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 44130.
(73) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14* (1), 71−73.
(74) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135−2145.
(75) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput* **2009**, *5* (9), 2353−2370.
(76) Brooks, C. L. Computer Simulation of Liquids. *J. Solution Chem.* **1989**, *18* (1), 99−99.
(77) Adams, D. J. Computer Simulation Study of Polar Liquids: Static and Dynamic Properties. *Proc. R. Soc. London, Ser. A* **1984**, *394* (1806), 137−160.
(78) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(79) Jones, J. E. On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. London, Ser. A* **1924**, *106* (738), 441−462.
(80) Jones, J. E. On the determination of molecular fields. —II. From the equation of state of a gas. *Proc. R. Soc. London, Ser. A* **1924**, *106* (738), 463−477.
(81) Lennard-Jones, J. E. Cohesion. *Proceedings of the Physical Society* **1931**, *43* (5), 461−482.
(82) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph* **1996**, *14* (1), 33−38.
(83) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12). .
(84) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput* **2007**, *3* (1), 26−41.
(85) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319−2327.
(86) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. Python Sci. Conf.* **2016**, 98−105.
(87) Kramer, J. R.; Onoa, B.; Bustamante, C.; Bertozzi, C. R. Chemically Tunable Mucin Chimeras Assembled on Living Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12574−12579.
(88) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. Principles of Mucin Architecture: Structural Studies on Synthetic Glycopeptides Bearing Clustered Mono-, Di-, Tri-, and Hexasaccharide Glycodomains. *J. Am. Chem. Soc.* **2002**, *124* (33), 9833−9844.

https://doi.org/10.1021/acs.jcim.4c00613
J. Chem. Inf. Model. 2024, 64, 7949−7965