

Glycan Conformation in the Heavily Glycosylated Protein, CEACAM1

Monique J. Rogals, Alexander Eletsky, Chin Huang, Laura C. Morris, Kelley W. Moremen, and James H. Prestegard*



Cite This: *ACS Chem. Biol.* 2022, 17, 3527–3534



Read Online

ACCESS |



Metrics & More

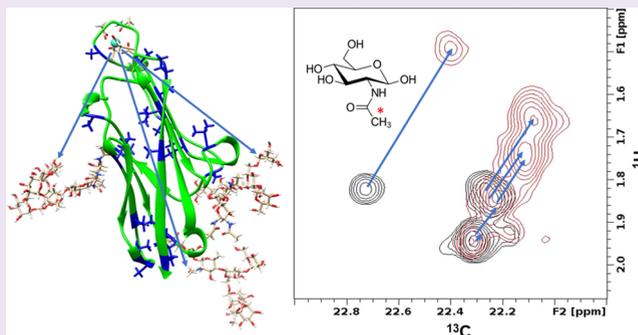


Article Recommendations



Supporting Information

ABSTRACT: Glycans attached to glycoproteins can contribute to stability, mediate interactions with other proteins, and initiate signal transduction. Glycan conformation, which is critical to these processes, is highly variable and often depicted as sampling a multitude of conformers. These conformers can be generated by molecular dynamics simulations, and more inclusively by accelerated molecular dynamics, as well as other extended sampling methods. However, experimental assessments of the contribution that various conformers make to a native ensemble are rare. Here, we use long-range pseudo-contact shifts (PCSs) of NMR resonances from an isotopically labeled glycoprotein to identify preferred conformations of its glycans. The *N*-terminal domain from human Carcinoembryonic Antigen Cell Adhesion Molecule 1, hCEACAM1-Ig1, was used as the model glycoprotein in this study. It has been engineered to include a lanthanide-ion-binding loop that generates PCSs, as well as a homogeneous set of three ^{13}C -labeled *N*-glycans. Analysis of the PCSs indicates that preferred glycan conformers have extensive contacts with the protein surface. Factors leading to this preference appear to include interactions between *N*-acetyl methyls of GlcNAc residues and hydrophobic surface pockets on the protein surface.



INTRODUCTION

Glycans adorn the surface of nearly all cells in nature, and most secreted and membrane-anchored proteins in humans carry covalently attached glycans.¹ They play roles that include control of the protein folding in the endoplasmic reticulum (ER), regulation of serum protein lifetimes, modulation of interactions between cells, and initiation of signals that alter cell function.² Pathogens have also adopted glycosylation as a means of evading our immune defenses either by mimicking host glycosylation or simply shielding antigenic surfaces.^{3,4} Much research has been directed at determining the composition and primary structure of free glycans.^{5,6} While conformations of free glycans in solution are also well studied,⁷ defining their dynamics and the multiple conformations they sample remains a challenge. Much less research has been directed at glycan conformation in the context of covalent attachment to a protein surface. What does exist has come from computer simulations, including long molecular dynamics (MD) trajectories.^{8–10} The prevailing view is that, here too, the glycans are quite flexible, sampling significant ranges of the phi and psi torsion angles that link individual residues of the glycans. In larger glycans (7 to more than 12 residues), the accumulated variations of phi and psi angles make accessible conformational space very large. Some recent depictions of this space can be seen in superpositions of sampled conformations meant to depict the glycan shield of viruses.^{11,12} But, are all conformations sampled equally, or are there preferred

conformations that may govern the functional properties of glycans? It is our intent to experimentally explore conformational preferences as they exist in the context of covalent attachment to a glycoprotein surface.

The methods we choose for this exploration begin with the generation of a probable set of conformers using a special type of molecular dynamics simulation, accelerated MD (aMD).^{13,14} This is followed by conformer evaluation using paramagnetic effects on nuclear magnetic resonances (NMRs).^{15,16} Conventional MD (cMD) simulations running for a few microseconds may provide effective exploration of glycan conformations when they have just a few glycosidic torsion angles, but without access to very specialized computer resources,¹⁷ cMD simulations long enough to sample conformational space for larger glycans, and those attached to protein surfaces, become impractical. Accelerated MD overcomes this limitation by boosting the minima of potential energy wells and allowing more frequent transitions to adjoining conformational states. A microsecond simulation

Received: September 16, 2022

Accepted: November 7, 2022

Published: November 23, 2022



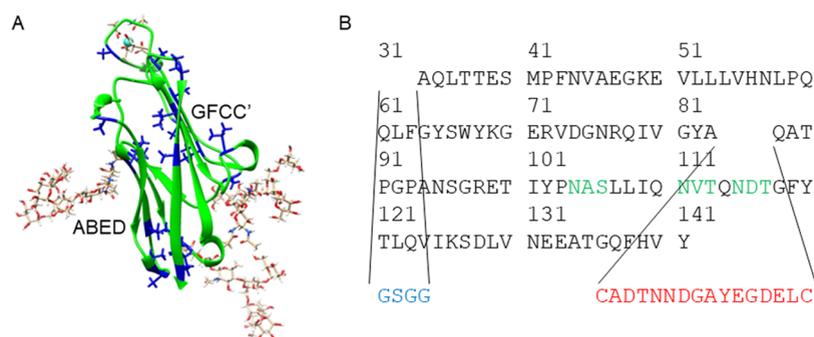


Figure 1. hCEACAM1-Ig1-LBP4 structure (A) and sequence (B). (A) Isotopically labeled sites (Ala and Val) are colored blue, the bound Tb^{3+} ion is in cyan, and glycans are colored by element. (B) Inserted LBP is in red, glycosylation sites are in green, and the scar left by TEV cleavage is in blue.

appears to sample conformations similar to those found in a millisecond simulation using cMD.¹³ This comes with the expense of having an unrealistic energy landscape and an inability to directly assign probabilities to representative conformers. Procedures to remove boost energies exist, and refinement of procedures is progressing. A recent advance imposes a Gaussian restriction on boosted energies, allowing a more reliable return to normal energies and estimation of conformer populations. The resulting package, called Gaussian accelerated MD (GaMD), will be used in our application.¹⁴ It allows us to select a few highly probable conformations for screening with experimental data.

We chose solution NMR as a means of experimental validation because glycans are not conformationally restricted in solution as they might be in crystal or low-temperature cryo-EM work. Normally, solution NMR structural data on glycans would come from a combination of nuclear Overhauser effects (NOEs) and scalar couplings.⁷ However, both are short-range observations (<5 Å or a few bonds), with ambiguities that are compounded by propagation across several glycosidic linkages. Also, the accurate measurement of trans-glycosidic scalar couplings is limited by the additional broadening of resonances when glycans are attached to a protein. We instead opt for the use of pseudo-contact shifts (PCSs) and paramagnetic relaxation enhancements (PREs), which are longer range (5–40 Å) and more appropriate for positioning glycans relative to a protein surface.¹⁵ Both are distance-dependent, as shown in eqs 1 and 2, where R is the length of a vector connecting a paramagnetic entity and the NMR nucleus observed. PCSs have additional angular dependencies through the polar angles, ϕ and φ , which relate the R vector to the principal frame of the anisotropic part of the entity's magnetic susceptibility tensor, " $\Delta\chi''$ ". Both have parameters (ΔX_{ax} , ΔX_{rh} , and R_0) that must be determined from effects on known structures (in our case, the structure of the protein as opposed to glycans). The effects are easily measured from changes in two-dimensional heteronuclear single quantum coherence (HSQC) spectra, loss of cross-peak intensity (I) compared to that of a reference spectrum carrying a diamagnetic entity (I_0) for PREs, and change in chemical shift compared to that seen in the same reference spectrum for PCSs

$$\delta^{PCS} = \frac{1}{12R^3} \left\{ \Delta X_{ax} (3 \cos 2\theta - 1) + \Delta X_{rh} \frac{3}{2} \sin 2\theta \cos(2\varphi) \right\} \quad (1)$$

$$PRE: I/I_0 = \exp(-2.303 \times (R_0/R)^6) \quad (2)$$

Most proteins do not carry a native paramagnetic entity; so, in these cases, a suitable entity must be introduced. Options include reacting a surface cysteine with a functionalized nitroxide (for PREs) or a lanthanide-binding chelate for PREs and PCSs.¹⁵ Inserting a lanthanide-ion-binding peptide loop in the protein sequence is also an option that provides PREs and PCSs.^{18,19} We choose the latter option, which has some advantages. Proteins frequently have native cysteines that are part of active sites or participate in structurally important disulfide bonds, and selectively reacting a newly introduced cysteine is often problematic. Peptide loop insertion is straightforward as expression constructs are often synthesized by commercial laboratories from provided sequences, and addition of a peptide loop to these sequences adds little to the process. However, choosing sites for the addition that neither distorts native protein structure nor reduces binding affinities of the loop must be done carefully. Our approach for the target discussed here has recently been described, along with a resonance assignment strategy for proteins sparsely labeled with NMR-active isotopes.²⁰

The protein target we select for glycan conformational analysis is the N-terminal domain of human Carcinoembryonic Antigen Cell Adhesion Molecule 1, hCEACAM1. hCEACAM1 is a highly glycosylated extracellular protein receptor, which has been implicated in gastrointestinal autoimmune disorders, cancer, and host–pathogen interactions.²¹ The N-terminal domain (hCEACAM1-Ig1) is an immunoglobulin (Ig-like) domain made up of two β sheets composed of strands ABED and GFCC', respectively (see Figure 1A). It has three glycosylation sites where the reducing end of a glycan can be linked to the sidechain nitrogen of asparagine residues (N-glycosylation). Our construct, illustrated in Figure 1B, carries a lanthanide-binding peptide inserted between native residues 83 and 87, raising the effective molecular weight to ~18 kDa. For this study, expression is carried out in mammalian cells that lack the MGAT1 enzyme. This stops N-glycan processing at the Man₅GlcNAc₂ stage, yielding a homogeneous complement of glycoforms.^{22,23} This falls short of the longer glycans natively present, but it would be much more difficult to achieve homogeneity of these complex glycans, and the shorter glycans will provide some insight into the nature of glycan–protein interactions. The disulfide bond included in the lanthanide-binding peptide both stabilizes the structure and raises occupancy of the three N-glycan sites to ~90%.²⁰

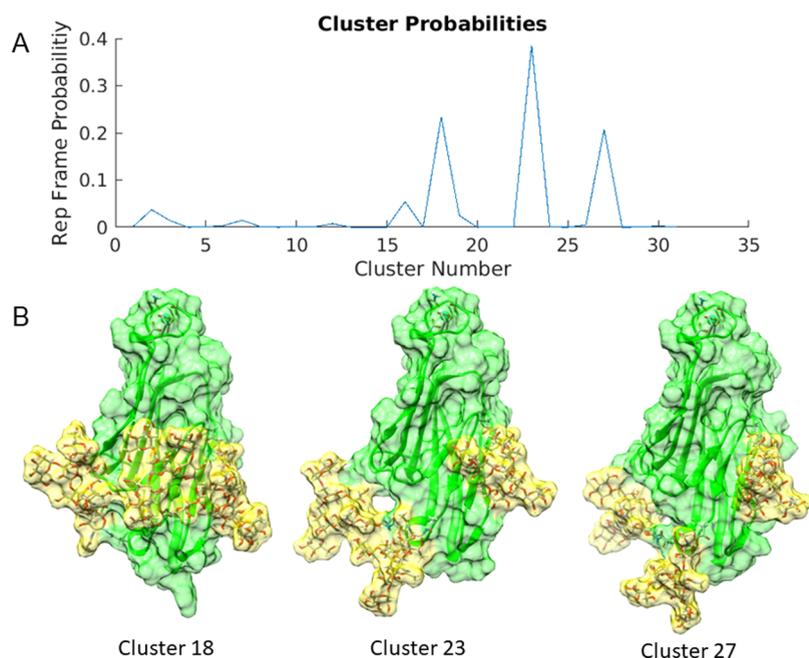


Figure 2. Glycan structures from a GaMD simulation are clustered based on pairwise rmsd's. (A) Probabilities for various clusters. (B) Structures for representative frames. Surfaces for the protein and glycans are shown in green and yellow, respectively.

The native hCEACAM1-Ig1 domain has a strong tendency to homodimerize, and the conformations of glycans may be of some interest because of their possible effect on dimerization. It is now well-established that the dominant mode of homodimerization is through the GFCC' face and glycans do not affect this dimerization.^{24,25} However, homodimer interactions through the ABED face of nonglycosylated forms have been reported, both in the original crystal structure²⁶ and subsequent electron tomography studies.²⁷ There also remain good arguments for continued interest in possible heterodimerization and the effect of glycans because of the range of other molecules with which CEACAM1 interacts.²⁸ We will not investigate dimerization directly in the following studies as interpretation of PREs and PCSs is complex when effects come from two lanthanide ions. The position of the lanthanide-binding loop depicted in Figure 1, in fact, inhibits dimerization while minimizing direct contact with the glycans. However, we anticipate that the conformational information generated will show positional preferences for glycans useful in assessing interactions through the ABED face.

RESULTS

Generation of Glycan Conformers. The initial model for the protein carrying the lanthanide-binding loop (hCEACAM1_Ig1_LBP4) was produced using the computational tool, AlphaFold.²⁹ A lanthanide ion (Tb^{3+}) was added and the coordinating ligands adjusted to mimic the geometry seen in the crystal structure of the initially designed loop (1TJB).³⁰ $\text{Man}_3\text{GlcNAc}_2$ glycans were added using tools available in the GLYCAM website (<https://glycam.org/>). This initial structure was solvated in TIP5P water and subjected to a 1 μs GaMD trajectory as described generally in the AMBER20 manual³¹ and more specifically in our previously submitted publication.²⁰ The TIP5P water model, as opposed to the more computationally efficient TIP3P model, was chosen to avoid an overemphasis of protein–glycan interactions that had been noted previously.^{8,10}

To extract a set of glycan conformers small enough to test against experimental observation, we clustered the GaMD frames based on a root mean square deviation (rmsd) of the position of central ring atoms (C1, C2, C4, C5) of all of the glycan residues on proteins aligned pairwise frame by frame for every tenth frame. In total, 31 clusters containing more than 500 members and separated by an inconsistency measure of 2.2 were produced and a structure with the minimum energy for each cluster was selected for examination. Because boost energies are added to facilitate a broad exploration of conformational space, populations of clusters do not reflect populations expected in solution. The effects of boost energies can be removed by a reweighting process.³² The results of this process (described more fully in our Methods section) are presented in Figure 2A. There are three clusters identified as highly probable and two more with significantly lower probability. Structures for the representative frames of the three most probable clusters are shown in Figure 2B. The surface representations in Figure 2B make it clear that the glycans in these structures are not extended but fold near the protein surface. The representative frames of the lower probability clusters, in particular, that of cluster 16, do have more extended glycans.

Experimental Evaluation of Glycan Conformers.

Pseudo-contact shifts prove to be the best probe of conformational preferences. Their $1/r^3$ dependence on the distance between the paramagnetic ion (Tb^{3+}) and an isotopically labeled site has effects at longer ranges than the $1/r^6$ dependent paramagnetic relaxation enhancements (PREs), and they are easily measured as displacements of cross-peaks for isotopically labeled sites in two-dimensional HSQC or HETCOR spectra. The first step is to use PCSs of sites in the protein to calculate the anisotropic part of the ion's susceptibility tensor ($\Delta\chi$). Supplementation of the expression media with a combination of valine (^{13}C -labeled at both methyl groups) and glucose (^{13}C -labeled at the C1 position) provides the necessary labeled sites. Valines are incorporated

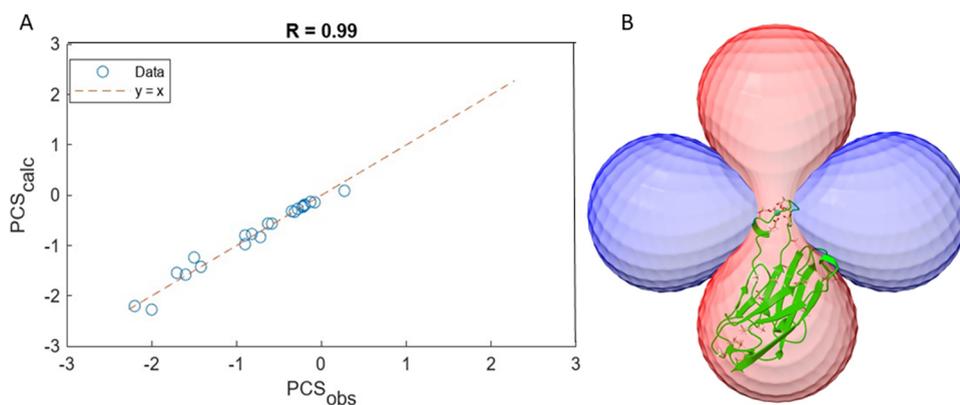


Figure 3. (A) Pseudo-Contact shifts (PCSs) can be fit to extract susceptibility tensors ($\Delta\chi$). (B) Surfaces for a shift of 0.13 ppm; red, negative and blue, positive.

directly, and glucose undergoes glycolysis to pyruvate, some of which is converted to alanine ^{13}C -labeled at its methyl group. The resulting cross-peaks were assigned to specific sites in our previously submitted publication using a combination of predicted chemical shifts, NOEs, PREs, and PCSs, along with a structure for hCEACAM_Ig1_LBP4 generated using a combination of AlphaFold structure prediction and a GaMD conformational search.²⁰ PCSs were measured for sites spanning distances of 5–44 Å between a Tb^{3+} ion in the lanthanide-binding loop and the isotopically labeled carbons of alanines or valines. In Figure 3A, these are plotted versus back-calculated values using an optimized $\Delta\chi$ tensor. Agreement between calculated and experimental values is excellent. The principal elements of this tensor, derived using REDCAT software,³³ are $\Delta X_{xx} = 0.3 \times 10^{-32}$, $\Delta X_{yy} = 75.2 \times 10^{-32}$, and $\Delta X_{zz} = -75.5 \times 10^{-32}$. Surfaces at points having a PCS of ± 0.13 ppm are shown in Figure 3B (red -0.13 ppm; blue $+0.13$ ppm). Concentric surfaces at larger and smaller distances (r) from the ion would have PCSs decreasing and increasing with $1/r^3$, respectively.

Figure 4 shows PCS surface depictions identical to those in Figure 3B, but now with glycans shown. On the left is the

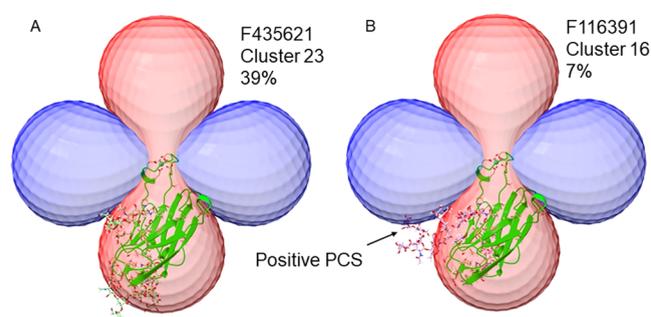


Figure 4. PCS surfaces overlaid on representative frames from clusters 23 (A) and 16 (B). (A) All glycans are predicted to have negative shifts. (B) Many glycans are predicted to have smaller PCSs and some to have positive shifts.

representative frame from the cluster predicted to be most probable (cluster 23) and on the right is the representative frame from a cluster predicted to have low probability, cluster 16. The glycans of the cluster 23 representative fall largely within the red surface, indicating that large negative PCSs would be observed if conformers from this cluster dominated.

Many of the glycans of the cluster 16 representative fall outside the red surface, suggesting smaller PCSs would be observed if conformers from this cluster dominated. Some glycan residues also fall within the blue surface, indicating that large positive PCSs would be observed along with smaller negative PCSs if conformers from this cluster dominated.

The $\text{C1-}^{13}\text{C}$ -labeled glucose supplied in the expression medium results in isotopic labeling of the anomeric carbons of all glycan residues.^{34,35} In addition to labeling alanine, the portion degraded to pyruvate goes to acetyl-CoA, and the ^{13}C ends up in the methyls of acetyl groups of the GlcNAc residues. Regions of a ^{13}C - ^1H HETCOR spectrum showing cross-peaks from anomeric and acetyl methyls are shown in Figure 5A,B, respectively. The black contours are for a sample prepared with a diamagnetic Lu^{3+} ion. The red contours are for a sample prepared with a paramagnetic Tb^{3+} ion. Black contours for residues on the three hCEACAM-Ig1-LBP4 glycans often overlap (there are 21 sites and just 11 resolved cross-peaks). Fortunately, chemical shifts for various residues and linkages are sufficiently different to allow grouping of cross-peaks by residue and linkage type. The annotations in Figure 5B come from a database/prediction tool on the CASPER site.³⁶ Cross-peaks defined by red contours are better resolved leading to discrete cross-peaks and a measurable number of PCSs approximating the total number of labeled glycan sites. Breaking the degeneracies of cross-peaks from similar or identical glycans is an added benefit of PCSs that has been previously noted.³⁷ The diagonal arrows connecting black and red contours have lengths representing PCSs. These are also reported in Table S1. Strikingly, all shifts are upfield (negative ppm). This suggests that, indeed, conformers contained in cluster 23 are better descriptors for conformers populated in solution than those contained in cluster 16. On examining additional clusters, we also find that all three highly probable clusters (18, 23, and 27) have representative frames with glycans primarily within the red surface of Figure 4, strengthening the argument for conformers with close protein contacts being favored.

DISCUSSION

An obvious question arises as to why close contacts may be favored. A close look at the structure of the representative frames from the most probable clusters provides some answers. All three show very similar contact to that depicted in Figure 6A for the glycan attached to asparagine 104 of hCEACAM-Ig1-LBP4. There are close contacts between the *N*-acetyl

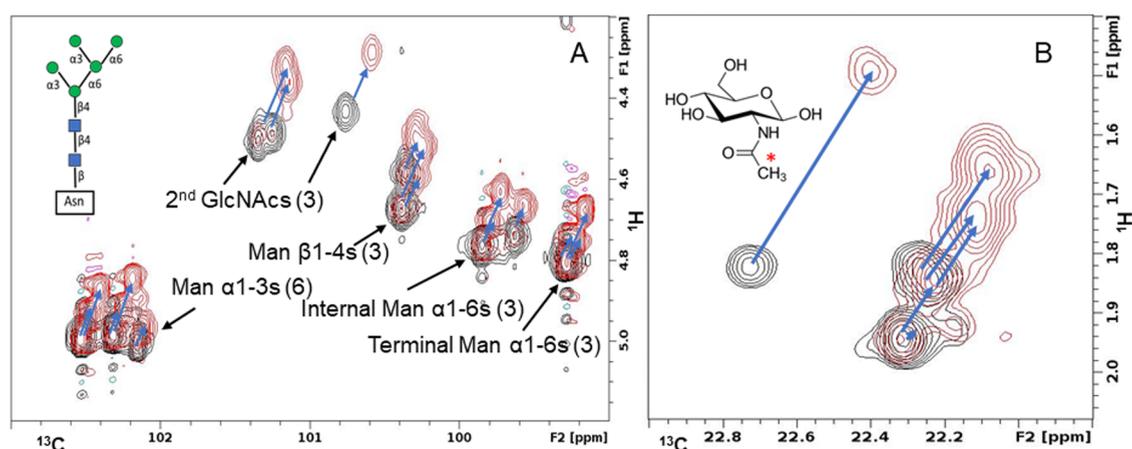


Figure 5. Overlaid HETCOR spectra of a Lu^{3+} (black contours) and a Tb^{3+} (red contours) containing samples. (A) Anomeric region. (B) *N*-acetyl region. Blue arrows connecting contours show magnitudes of PCSs. All shifts are negative.

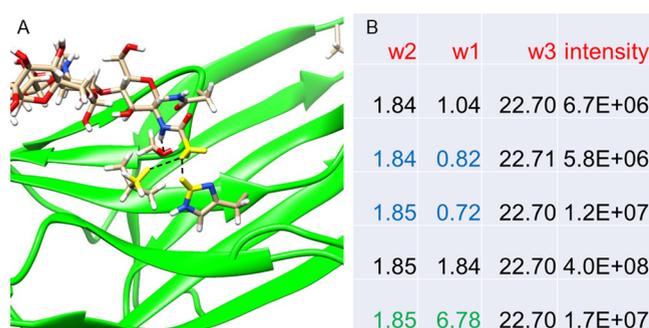


Figure 6. Hydrophobic interactions between *N*-acetyl methyl and L25 and H27 at the protein surface. (A) Structure showing close contacts. (B) NOE data supporting interaction.

methyl of the first GlcNAc and methyl protons of leucine 25 and a ring proton of histidine 27, which form a hydrophobic pocket on the surface of the protein. This interaction is supported by a hydrogen bond between the sidechain oxygen of serine 89 and the amide proton of the *N*-acetyl group. The existence of the hydrophobic interaction is not without additional support. NOEs were collected in the process of making the previously reported protein resonance assignments.²⁰ Along with protein-specific NOEs came NOEs involving the *N*-acetyl methyl of glycans. The peak list for the glycan showing the largest *N*-acetyl PCS (consistent with it being that attached to N104) is shown in Figure 6B (a complete set of strip plots for acetyl groups is included in Figure S1). Among the chemical shifts for donating protons (w1 column) are two that are consistent with leucine methyls (0.82 and 0.72 ppm) and a shift that is consistent with either a dramatically shifted amide resonance (6.78 ppm) or the HD2 proton of a histidine. Other hydrophobic and hydrogen-bonding interactions can be found for other glycans, often involving mannose residues that are farther from the point of attachment. Hence, we have the beginning of an understanding of forces that can influence conformations of glycans attached to the surface of glycoproteins. With respect to potential protein–protein interactions through the ABED interface, four of the five most probable clusters position the glycan attached to N104 in a manner similar to that depicted in Figure 6A and one that would protect this interface. The representative structure for cluster 16 has the N104 glycan more extended but in a direction that would still interfere with interactions

through the ABED face. Glycans attached to N111 and N115 are in all cases positioned well away from the ABED face and the dimerization surfaces of the CFCC' face.

Of course, a more quantitative description of glycan conformations would be desirable. It is possible to quantitatively predict PCSs using the ΔX tensor we have determined and the representative frames of the various clusters. We have done this for a selection of glycan sites and the three most probable clusters (Figure S2). Comparison to measured PCSs requires assignment to specific sites. We find that we can do this with reasonable agreement for some, but not all, sites. There is also a tendency for predicted PCSs to be larger than observed PCSs. This lack of complete agreement indicates that, while conformers with close protein contacts are favored, considerable conformational averaging with more extended forms still occurs. Also, the accuracy of probabilities predicted for various clusters is still in question. Those shown are based on boosting and correcting total energies, including those for the protein, solvent, and all three glycans. This was dictated by the energy boost and correction routine available at the start of our research. It would be desirable to treat glycans separately from the protein and separately from one another. GaMD techniques that allow boosting of energies for separate parts of a system are now available,³⁸ and application to selective boosting of glycan energies may well be possible in the future. Once we have an improved selection of probable clusters, we could undertake fitting of observed PCSs to averages over predicted PCSs for representative frames. Water models are also still in question. TIPSP does produce less glycan–glycan clustering than TIP3P, but these models continue to evolve,³⁹ and experimental testing of theoretical predictions will be important. Nevertheless, we have the beginning of an understanding of forces that may influence glycan conformations as they exist on the surface of glycoproteins, and we have the experimental procedures to test this understanding.

METHODS

Sample Preparation. The amino acid sequence of the hCEACAM1-Ig1_LBP4 construct is presented in Figure 1. The lanthanide-binding tag sequence is derived from the initial work of the Imperiali Laboratory,¹⁹ and the rationale behind the introduction of the disulfide bond and its placement in the native hCEACAM1 sequence is described in our prior publication.²⁰ Procedures for expression and purification of protein are also described in that

publication. But briefly, the synthesized gene was inserted into a pGen2 expression vector. HEK293S (MGAT1 knockout) cells, which produce primarily $\text{Man}_5\text{GlcNAc}_2\text{N-glycans}$, were transfected and expression occurred in 1 L of a custom version of FreeStyle 293 media (Gibco, ThermoFisher Scientific). This medium, which lacked both glucose and amino acids, was supplemented with 5 g of ^{13}C -glucose, and 150 mg of ^{13}C -dimethyl-valine (Cambridge Isotope Labs, Tewksbury, MA) along with other amino acids normally present in the medium. The glycoprotein that was secreted into the medium was purified by metal affinity chromatography to remove media components and treated with TEV protease to remove the His tag and GFP fusion sequences. Further purification by metal affinity chromatography and Superose 75 gel filtration led to a final yield of ~ 10 mg. The protein sequence and glycosylation level (90%) were verified by mass spectrometry. For NMR spectroscopy, the protein was exchanged into a buffer composed of 25 mM Tris, 100 mM NaCl, pH 7.4, 0.02% NaN_3 , 10 μM DSS, 90/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$ at a final protein concentration of 300 μM . Near-molar equivalents (~ 0.9) of TbCl_3 and LuCl_3 were added to PCS and control samples, respectively.

Generation of Structural Models. An initial structural model for hCEACAM1-Ig1-LBP4 was prepared using AlphaFold software²⁹ installed on the high-performance cluster at the University of Georgia Advanced Computing Resource Center. As ion-binding motifs are not well reproduced, a lanthanide-binding loop with ion coordination observed in prior structures was manually added and minimized using the UCSF Chimera molecular modeling system.⁴⁰ In addition, three $\text{Man}_5\text{GlcNAc}_2$ glycans were added to the three asparagines found in N-glycosylation sequons, N104, N111, and N115, using the tools available at the Glycam website (<https://glycam.org/>). The model was subjected to a further structural search using a 1 μs Gaussian accelerated molecular dynamics simulation.⁴¹ This structure was solvated in TIP5P water in a rectangular box with a minimum distance between the solute and edge of the box of 8 Å following the protocols in AMBER20. The ff14SB forcefield was used for amino acids, and the Glycam06j forcefield was used for glycans. After minimization and a short (100ns) conventional MD run, the system was subjected to 1 μs of Gaussian-accelerated MD (GaMD) using a dual boost protocol as implemented in AMBER20. Steps of 2 fs were used in the simulation and frames were saved every 2 ps.

Clustering, as well as energy adjustment, was accomplished using the Matlab scripts described in our previously submitted publication.²⁰ Clustering was based on an rmsd pairwise comparison of the spatial positions of C1, C2, C4, and C5 atoms of all glycan residues in every tenth saved frame. The resulting agglomerative hierarchical clusters were produced using the Ward-Linkage Algorithm.⁴² Clusters having more than 500 members and an inconsistency measure of 2.0 were selected. Probabilities were assigned to these clusters based on adjusted dihedral energies,³² and the frame having the minimum energy within each cluster was designated as the representative frame.

Collection of Experimental Data. NMR spectra were acquired at 25 °C on a Bruker AVANCE NEO 900 MHz spectrometer using a triple-resonance 5 mm TXO cryogenic probe optimized for ^{13}C and ^{15}N observations. ^{13}C -detected 2D HETCOR spectra were recorded for hCEACAM1-Ig1-LBP4 samples containing Tb^{3+} and Lu^{3+} with 1024×64 complex points and 48 scans per increment using a version of standard Bruker hxnepph pulse sequence modified for semi-constant time evolution in the indirect (^1H) dimension. Sweep widths of 61.4 and 4.0 ppm were used for the ^{13}C and ^1H dimensions, respectively. The forward INEPT delay was set to $1/(2J)$ (4.0 ms), while the refocusing INEPT delays were set to $1/(6J)$ (1.3 ms), where J was assumed to be 125 Hz. Processing was done in TOPSPIN 4.1 using Gaussian weighting in both dimensions. Cross-peaks in the Tb^{3+} and Lu^{3+} samples were paired using diagonal lines and PCSs measured from the length of these lines.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscchembio.2c00714>.

Information includes strip plots of N-acetyl NOE data, an example of a strategy for site-specific assignment of glycan resonances, tables of PCS and NOE data, and Matlab scripts used in the analysis of GaMD trajectories (PDF)

■ Accession Codes

Assignments of alanine and valine methyl resonances used in the determination of susceptibility tensors have been deposited with the Biological Magnetic Resonance Data Bank (BMRB), accession code 51599.

■ AUTHOR INFORMATION

■ Corresponding Author

James H. Prestegard – Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States; orcid.org/0000-0002-1602-4790; Email: jpresteg@ccrc.uga.edu

■ Authors

Monique J. Rogals – Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

Alexander Eletsky – Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

Chin Huang – Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

Laura C. Morris – Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

Kelley W. Moremen – Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acscchembio.2c00714>

■ Author Contributions

M.J.R. and A.E. collected NMR data. C.H. expressed and purified the protein. L.C.M. ran the accelerated MD trajectories. J.H.P. analyzed the trajectories and drafted the manuscript. J.H.P. and K.W.M. conceived and directed the project. All authors edited the final manuscript.

■ Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the US National Institutes of Health (NIH), R01 GM033225 and R01 GM134335, to J.H.P. It also benefitted from instrumentation made available through NIH grant S10 OD021623 to J.H.P. The authors thank Y. Miao of the University of Kansas for advice on accelerated MD simulations.

■ REFERENCES

- (1) Varki A, K. S. Historical background and overview. In *Essentials of Glycobiology*, 4th ed.; Varki A, C. R.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2022; pp 1–19.
- (2) Gagneux, P.; Varki, T. H.; ABiological functions of glycans. In *Essentials of Glycobiology*, 4th ed.; Varki A, C. R.; Esko, J. D.; Stanley,

- P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2022; pp 79–91.
- (3) Lee, S.; Inzerillo, S.; Lee, G. Y.; Bosire, E. M.; Mahato, S. K.; Song, J. Glycan-mediated molecular interactions in bacterial pathogenesis. *Trends Microbiol.* **2022**, *30*, 254–267.
- (4) Banahene, N.; Kavunja, H. W.; Swarts, B. M. Chemical Reporters for Bacterial Glycans: Development and Applications. *Chem. Rev.* **2022**, *122*, 3336–3413.
- (5) Grabarics, M.; Lettow, M.; Kirschbaum, C.; Greis, K.; Manz, C.; Pagel, K. Mass Spectrometry-Based Techniques to Elucidate the Sugar Code. *Chem. Rev.* **2022**, *122*, 7840–7908.
- (6) Cañada, F. J.; Canales, A.; Valverde, P.; de Toro, B. F.; Martinez-Orts, M.; Phillips, P. O.; Pereda, A. Conformational and Structural Characterization of Carbohydrates and their Interactions Studied by NMR. *Curr. Med. Chem.* **2022**, *29*, 1147–1172.
- (7) Battistel, M. D.; Azurmendi, H. F.; Yu, B.; Freedberg, D. I. NMR of glycans: Shedding new light on old problems. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *79*, 48–68.
- (8) Woods, R. J. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem. Rev.* **2018**, *118*, 8005–8024.
- (9) Perez, S.; Makshakova, O. Multifaceted Computational Modeling in Glycoscience. *Chem. Rev.* **2022**, *122*, 15914–15970.
- (10) Fadda, E. Molecular simulations of complex carbohydrates and glycoconjugates. *Curr. Opin. Chem. Biol.* **2022**, *69*, No. 102175.
- (11) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **2020**, *10*, No. 14991.
- (12) Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Jeong, P. U.; Dommer, A. C.; Amaro, R. E. Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent. Sci.* **2020**, *6*, 189–196.
- (13) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
- (14) Miao, Y. L.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595.
- (15) Miao, Q.; Nitsche, C.; Orton, H.; Overhand, M.; Otting, G.; Ubbink, M. Paramagnetic Chemical Probes for Studying Biological Macromolecules. *Chem. Rev.* **2022**, *122*, 9571–9642.
- (16) Otting, G. Protein NMR Using Paramagnetic Ions. In *Annual Review of Biophysics*; Rees, D. C.; Dill, K. A.; Williamson, J. R., Eds.; 2010; Vol. 39, pp 387–405.
- (17) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.
- (18) Barthelmes, D.; Barthelmes, K.; Schnorr, K.; Jonker, H. R. A.; Bodmer, B.; Allen, K. N.; Imperiali, B.; Schwalbe, H. Conformational dynamics and alignment properties of loop lanthanide-binding-tags (LBTs) studied in interleukin-1 beta. *J. Biomol. NMR* **2017**, *68*, 187–194.
- (19) Allen, K. N.; Imperiali, B. Lanthanide-tagged proteins - an illuminating partnership. *Curr. Opin. Chem. Biol.* **2010**, *14*, 247–254.
- (20) Williams, R. V.; Rogals, M. J.; Eletsky, A.; Huang, C.; Morris, L. C.; Moremen, K. W.; Prestegard, J. H. AssignSLP_GUI, a Software Tool Exploiting AI for NMR Resonance Assignment of Sparsely Labeled Proteins. *J. Magn. Reson.* **2022**, DOI: 10.1016/j.jmr.2022.107336.
- (21) Kim, W. M.; Huang, Y. H.; Gandhi, A.; Blumberg, R. S. CEACAM1 structure and function in immunity and its therapeutic implications. *Semin. Immunol.* **2019**, *42*, No. 101296.
- (22) Moure, M. J.; Eletsky, A.; Gao, Q.; Morris, L. C.; Yang, J. Y.; Chapla, D.; Zhao, Y. J.; Zong, C. L.; Amster, I. J.; Moremen, K. W.; Boons, G. J.; Prestegard, J. H. Paramagnetic Tag for Glycosylation Sites in Glycoproteins: Structural Constraints on Heparan Sulfate Binding to Robo1. *ACS Chem. Biol.* **2018**, *13*, 2560–2567.
- (23) Moremen, K. W.; Ramiah, A.; Stuart, M.; Steel, J.; Meng, L.; Forouhar, F.; Moniz, H. A.; Gahlay, G.; Gao, Z. W.; Chapla, D.; Wang, S.; Yang, J. Y.; Prabhakar, P. K.; Johnson, R.; dela Rosa, M.; Geisler, C.; Nairn, A. V.; Seetharaman, J.; Wu, S. C.; Tong, L.; Gilbert, H. J.; LaBaer, J.; Jarvis, D. L. Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* **2018**, *14*, 156–162.
- (24) Belcher Dufresne, M.; Swope, N.; Kieber, M.; Yang, J. Y.; Han, J.; Li, J.; Moremen, K. W.; Prestegard, J. H.; Columbus, L. Human CEACAM1 N-domain dimerization is independent from glycan modifications. *Structure* **2022**, *30*, 658–670.
- (25) Zhuo, Y.; Yang, J. Y.; Moremen, K. W.; Prestegard, J. H. Glycosylation alters dimerization properties of a cell-surface signaling protein, carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) (vol 291, pg 20085, 2016). *J. Biol. Chem.* **2020**, *291*, 20085–20095.
- (26) Fedarovich, A.; Tomberg, J.; Nicholas, R. A.; Davies, C. Structure of the N-terminal domain of human CEACAM1: binding target of the opacity proteins during invasion of *Neisseria meningitidis* and *N-gonorrhoeae*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 971–979.
- (27) Klaile, E.; Vorontsova, O.; Sigmundsson, K.; Muller, M. M.; Singer, B. B.; Ofverstedt, L. G.; Svensson, S.; Skoglund, U.; Obrink, B. The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters. *J. Cell Biol.* **2009**, *187*, 553–567.
- (28) Gandhi, A. K.; Sun, Z. Y. J.; Kim, W. M.; Huang, Y. H.; Kondo, Y.; Bonsor, D. A.; Sundberg, E. J.; Wagner, G.; Kuchroo, V. K.; Petsko, G. A.; Blumberg, R. S. Structural basis of the dynamic human CEACAM1 monomer-dimer equilibrium. *Commun Biol.* **2021**, *4*, No. 360.
- (29) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohli, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (30) Nitz, M.; Sherawat, M.; Franz, K. J.; Peisach, E.; Allen, K. N.; Imperiali, B. Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. *Angew. Chem., Int. Ed.* **2004**, *43*, 3682–3685.
- (31) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T. E. Cheatham, I.; Cruzeiro, V.W.D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Iz-mailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; Kollman, D. M. Y. AMBER 2020.2020.
- (32) Miao, Y. L.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10*, 2677–2689.
- (33) Valafar, H.; Prestegard, J. H. REDCAT: a residual dipolar coupling analysis tool. *J. Magn. Reson.* **2004**, *167*, 228–241.
- (34) Subedi, G. P.; Falconer, D. J.; Barb, A. W. Carbohydrate-Polypeptide Contacts in the Antibody Receptor CD16A Identified through Solution NMR Spectroscopy. *Biochemistry* **2017**, *56*, 3174–3177.
- (35) Yamaguchi, Y.; Takizawa, T.; Kato, K.; Arata, Y.; Shimada, I. H-1 and C-13 NMR assignments for the glycans in glycoproteins by

using H-2/C-13-labeled glucose as a metabolic precursor. *J. Biomol. NMR* **2000**, *18*, 357–360.

(36) Stahle, J.; Widmalm, G. NMR Chemical Shift Predictions and Structural Elucidation of Oligo- and Polysaccharides by the Computer Program CASPER. In *NMR in Glycoscience and Glycotechnology*; Kato, K.; Peters, T., Eds.; 2017; Vol. 10, pp 335–352.

(37) Canales, A.; Mallagaray, A.; Perez-Castells, J.; Boos, I.; Unverzagt, C.; Andre, S.; Gabius, H. J.; Canada, F. J.; Jimenez-Barbero, J. Breaking Pseudo-Symmetry in Multiantennary Complex N-Glycans Using Lanthanide-Binding Tags and NMR Pseudo-Contact Shifts. *Angew. Chem., Int. Ed.* **2013**, *52*, 13789–13793.

(38) Wang, J. N.; Miao, Y. L. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Chem. Phys.* **2020**, *153*, No. 154109.

(39) Morozova, T. I.; Garcia, N. A.; Barrat, J. L. Temperature dependence of thermodynamic, dynamical, and dielectric properties of water models. *J. Chem. Phys.* **2022**, *156*, No. 126101.

(40) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(41) Miao, Y. L.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. In *Annual Reports in Computational Chemistry*; Dixon, D. A., Ed.; 2017; Vol. 13, pp 231–278.

(42) Peng, J. H.; Wang, W.; Yu, Y. Q.; Gu, H. L.; Huang, X. H. Clustering Algorithms to Analyze Molecular Dynamics Simulation Trajectories for Complex Chemical and Biological Systems. *Chin. J. Chem. Phys.* **2018**, *31*, 404–420.