ENERGETIC CHARACTERIZATION OF INTERACTIONS BETWEEN N-LINKED GLYCANS AND THE UNDERLYING PROTEIN SURFACE

by

ELIZABETH (ELIZA) GAZAWAY

(Under the Direction of Robert J. Woods)

ABSTRACT

Molecular dynamics (MD) simulations are widely used to model the 3D structure of glycoproteins, including the dynamics of the attached N-linked glycans, which play a critical role, for example, in shielding pathogens from immune recognition and in modulating intercellular adhesion. While MD is a valuable tool to model the conformational ensemble of N-linked glycans, achieving convergence of the distribution of glycan orientations relative to the protein surface remains a challenge. To alleviate this obstacle, tools have been developed to rapidly model conformational ensembles of glycans onto protein surfaces. These approaches neglect interactions between N-linked glycans and neighboring protein residues that can influence glycan orientation in solution. Here, we surveyed glycoprotein structures in the Protein Data Bank and employed *in silico* interaction energy calculations to quantify pairwise interactions between N-linked glycans and neighboring protein residues. We also examined the impact of core fucosylation on modulating glycan—protein interactions. Identifying stabilizing interactions observed across glycoprotein systems can inform the development of modeling strategies that aim to reflect the preferred orientations of N-linked glycans in solution.

INDEX WORDS: N-linked glycosylation, Glycoprotein, Glycoprotein Modeling, Molecular dynamics, Carbohydrate modeling, MM/GBSA

ENERGETIC CHARACTERIZATION OF INTERACTIONS BETWEEN N-LINKED GLYCANS AND THE UNDERLYING PROTEIN SURFACE

by

ELIZABETH (ELIZA) GAZAWAY

B.S., The University of Georgia, 2023

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2025

© 2025

Eliza Gazaway

All Rights Reserved

ENERGETIC CHARACTERIZATION OF INTERACTIONS BETWEEN N-LINKED GLYCANS AND THE UNDERLYING PROTEIN SURFACE

by

ELIZABETH (ELIZA) GAZAWAY

Major Professor:	Robert J. Woods
Committee:	Ana S. Ramírez
	Nathan E. Lewis

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2025

ACKNOWLEDGEMENTS

I would like to thank my family for their constant support and encouragement throughout my life and in my academic career. I would also like to thank Dr. Woods for his mentorship and for providing me the opportunity to work in his lab for the past 4 years. I would also like to acknowledge and thank my other mentors in the Woods' Group, including Drs. Oliver Grant, Lachele Foley, and Yao Xiao, who all have passed their knowledge and skills down to me.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

CHAPTER

1 INTRODUCTION 1

2 MATERIALS AND METHODS 11

3 RESULTS 17

4 DISCUSSION 32

5 CONCLUSIONS 37

6 REFERENCES 38

7 APPENDIX 49

Supplemental Information 49

CHAPTER 1

INTRODUCTION

General structure and biological processing of N-linked glycans

N-glycosylation is the most prevalent protein modification in eukaryotic cells, occurring in most secreted and membrane-bound proteins. N-glycosylation occurs co-translationally when the lipid-linked glycan $Glucose_3Mannose_9N\text{-}acetylglucosamine_2$ ($Glc_3Man_9GlcNAc_2$) is conjugated onto to asparagine (N) residues present in N-X(≠P)-S/T sequons by oligosaccharyltransferase (OST) (Aebi, 2013; Bai et al., 2018; Kornfeld & Kornfeld, 1985) (Figure 1a). In the ER, N-linked glycans function as ligands for protein folding machinery. The terminal glucose residues are removed by resident glucosidases, yielding $Glc_1Man_9GlcNAc_2$, which is recognized by chaperones calnexin and calreticulin (Hammond et al., 1994; Jakob et al., 1998). Moreover, N-linked glycans act as "molecular clocks" to monitor folding status. Proteins that remain persistently misfolded are retained in the ER, where the mannose residues of their N-linked glycans are gradually trimmed by ER degradation-enhancing mannosidase I, which targets them to be degraded by cytosolic proteosomes (Shenkman et al., 2018; Zhou et al., 2015).

Properly folded glycoproteins are trafficked through the secretory pathway where they are exposed to a series of glycosidases and glycosyltransferases that can further modify the glycan structure before they are transported to their final destinations. Due to the branched structure of N-glycans, they are characterized as having two "arms" that originate from the core mannose residue, Man-3 (Figure 1a). The first arm extends from the α1,3-linked mannose, and the other extends from the α1,6-linked mannose. ER α-mannosidase I removes a terminal α-linked mannose residue

before transfer to the Golgi complex (Figure 1b) (Xiang et al., 2016). In mammals, the terminal α1,2-linked mannose residues of the "high-mannose" glycan can be trimmed in the *cis*-Golgi to yield $Man_5GlcNAc_2$ (Figure 1c, left). This structure is the substrate of MGAT1, which catalyzes the transfer of an N-acetylglucosamine (GlcNAc) to the α1,3-linked mannose residue through an α1,2-linkage (Figure 1c) (Gordon et al., 2006). The addition of this residue converts the classification of the glycan from high mannose to "hybrid," as the α1,6 arm contains mannose residues from the original structure, whereas the α1,3 arm has been modified by the addition of a non-mannose residue. In the *trans*-Golgi, hybrid glycans may be further processed into "complex" glycans through the removal of the terminal mannose residues of the α1,6 arm, followed by the addition of an α1,2-linked GlcNAc residue to the remaining mannose (Figure 1d, left) (Kadirvelraj et al., 2018). This complex glycan may be core-fucosylated by FUT8, which catalyzes the transfer of an α1,6-linked fucose to the core GlcNAc residue (Figure 1d) (Boruah et al., 2020). Core-fucosylation has a variety of biological roles, for example, in regulating antibody-dependent cellular cytotoxicity (Falconer et al., 2018; Shields et al., 2002), and can serve as a cancer biomarker (Schneider et al., 2017). The complex glycan structure may be further modified, for example, by the addition of Galβ1-4GlcNAc ("LacNAc") repeats or capping with sialic acid, which plays a key role in cell signaling, for example in influenza infection (Stencel-Baerenwald et al., 2014; Weis et al., 1988) and in trafficking leukocytes to sites of infection (Zhu et al., 2024).

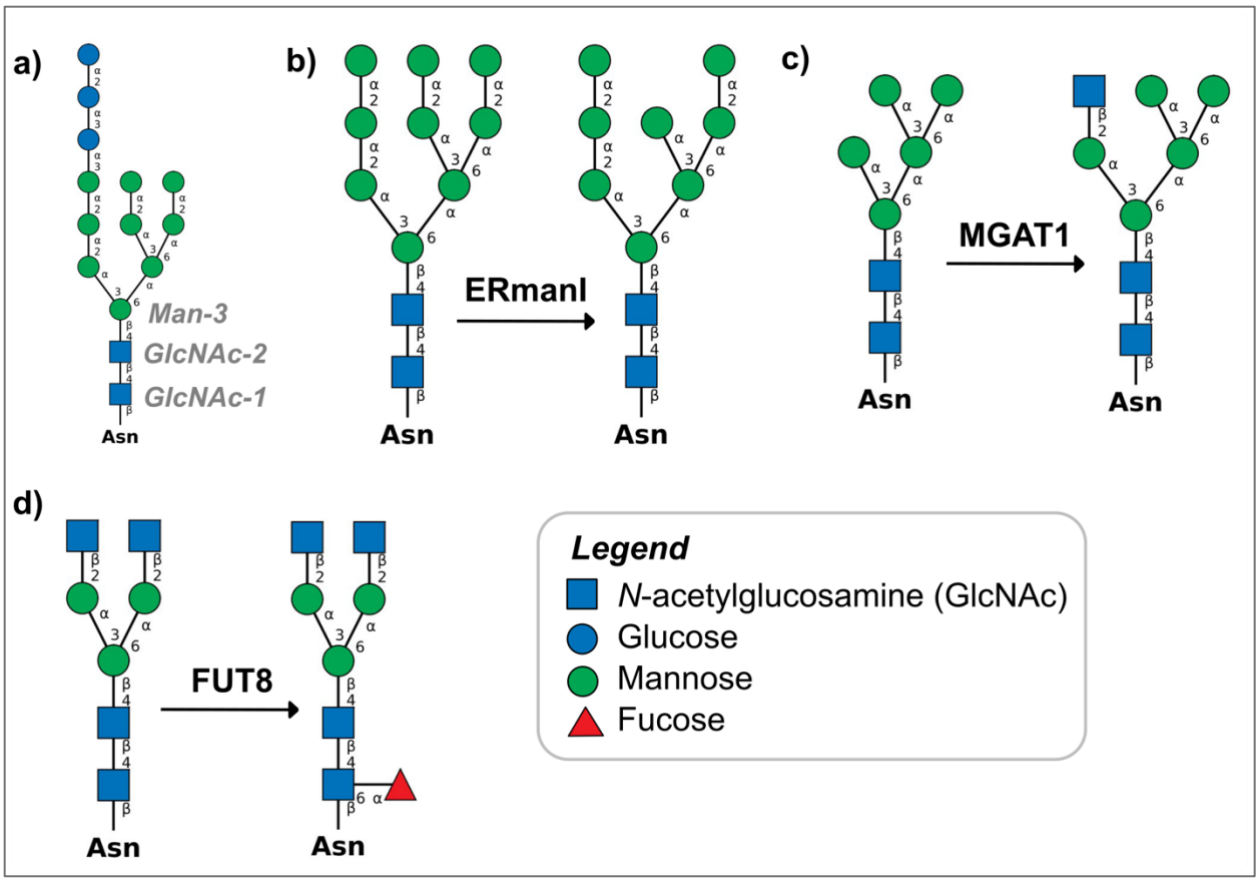


Figure 1: **N-Glycan structures at key stages in the maturation pathway. a)** SNFG depiction of the initial $Glc_3Man_9GlcNAc_2$ structure that is co-translationally transferred *en bloc* onto asparagine residues in the ER lumen. The glycosidic linkage between two residues is labeled in black, and the names the glycan residues in the trisaccharide core are labeled in gray. **b)** ER α-mannosidase I (ERmanI) catalyzes the hydrolysis of a terminal α1,2-linked mannose residue on the N-linked glycans of folded proteins before transfer to the Golgi complex. The N-linked glycans that initially enter the *cis*-Golgi are classified as "high-mannose" because the glycan consists of exclusively mannose residues attached to the chitobiose core ($GlcNAc_2$). **c)** MGAT1, the first enzyme in the maturation of hybrid and complex N-glycans, catalyzes the addition of a GlcNAc to the α1,3 arm. **d)** FUT8 catalyzes the addition of a fucose residue to the core GlcNAc through an α1,6 linkage. SNFG images were generated using DrawGlycan (Cheng et al., 2017).

Because glycoproteins are limited in the amount of time that they are in flux through the Golgi complex (Hirschberg et al., 1998), various factors impact the degree of processing that occurs to a given glycan, including enzyme expression levels (Kizuka, 2024) and glycosylation site accessibility (Adams et al., 2022; An et al., 2009; de Haan et al., 2023; Hang et al., 2015; Zhao et al., 2020). As a result, there is often variation in the glycan composition observed at a particular glycosylation site (microheterogeneity) as well as in the glycan composition observed between glycosylation sites on the same protein (macroheterogeneity).

Conformational properties of N-linked glycans

The internal glycosidic linkages between two glycan residues can consist of two or three bonds, depending of the position of the hydroxyl group on the non-reducing sugar that the anomeric carbon of the reducing sugar is linked to (Fontana & Widmalm, 2023). Two-bond linkages are defined by torsion angles φ and ψ, and the ω angle defines the third bond in three-bond linkages (Figure 2a). NMR characterization of the high-mannose glycan $Man_9GlcNAc_2$ (Figure 1b, left) revealed that the φ and ψ torsion angles of its α1,2 and α1,3 glycosidic linkages adopted a single dominant conformation, with oscillation of ± 10-15° (Woods et al., 1998). The restricted torsional flexibility was attributed to stereoelectronic effects, including the gauche and exoanomeric effects (Lemieux et al., 1979; Wiberg et al., 1990; Wolfe, 1972; Wolfe et al., 1979), and to electrostatic and steric interactions between the carbohydrate hydroxyl groups and solvent molecules that prevented major conformational changes in the glycan (Kirschner & Woods, 2001). A recent survey of the glycosidic linkages in the conserved pentasaccharide core, also referred to as paucimannose, of N-linked glycans (Figure 2a) was consistent with the limited conformational flexibility of glycosidic linkages, showing that the φ and ψ torsion angles occupy to a single

predominant conformation (Gazaway et al., 2025). The bond defined by the ω torsion angle introduces an additional degree of freedom in three-bond glycosidic linkages. Analysis of the ω angle in the α1,6 linkage in the paucimannose core showed two predominant conformations at approximately 60° and 180° with respect to the C4-C5-C6-O6 atoms, which are commonly referred as the *gauche/gauche* and *gauche/trans* conformations, respectively (Figure 2a). Structurally, variation in the ω torsion alters the presentation of the outer arm that extends from the α1,6 linkage (Woods, 2018) (Figure 2b).

Because they are covalently linked to the protein surface, the presentation (orientation relative to the protein) of N-linked glycans is influenced by the bond rotation of the linking asparagine side chain, which is described by the χ_1 and χ_2 torsion angles (Figure 2b). Surveys of these torsion angles in glycosylated asparagine side chains from the RCSB Protein Data Bank (PDB) identified three predominant conformations of the χ_1 angle at approximately 60°, 180°, and 300°, whereas greater torsional variability was observed in the χ_2 angle, which had a value between 120° and 240° (Gazaway et al., 2025; Petrescu et al., 2004). Variation in these angles greatly increases the conformational space accessible to N-linked glycans in solution. This feature can be illustrated by modeling the N-glycan pentasaccharide core (paucimannose) at the three glycosylations sites in erythropoietin (Syed et al., 1998). Considerable changes in the orientation of the paucimannose glycans are observed across distinct χ_1 and χ_2 conformers, even without altering the internal glycosidic torsion angles (Figure 2c). NMR analysis and computational modeling of N-glycoproteins have further indicated that the glycans exist as an ensemble of orientations on the protein surface in solution—a characteristic of intrinsically disordered molecules (Figure 2c) (Fogarty & Fadda, 2021; Slynko et al., 2009; Sticht et al., 1998; Woods et al., 1994).

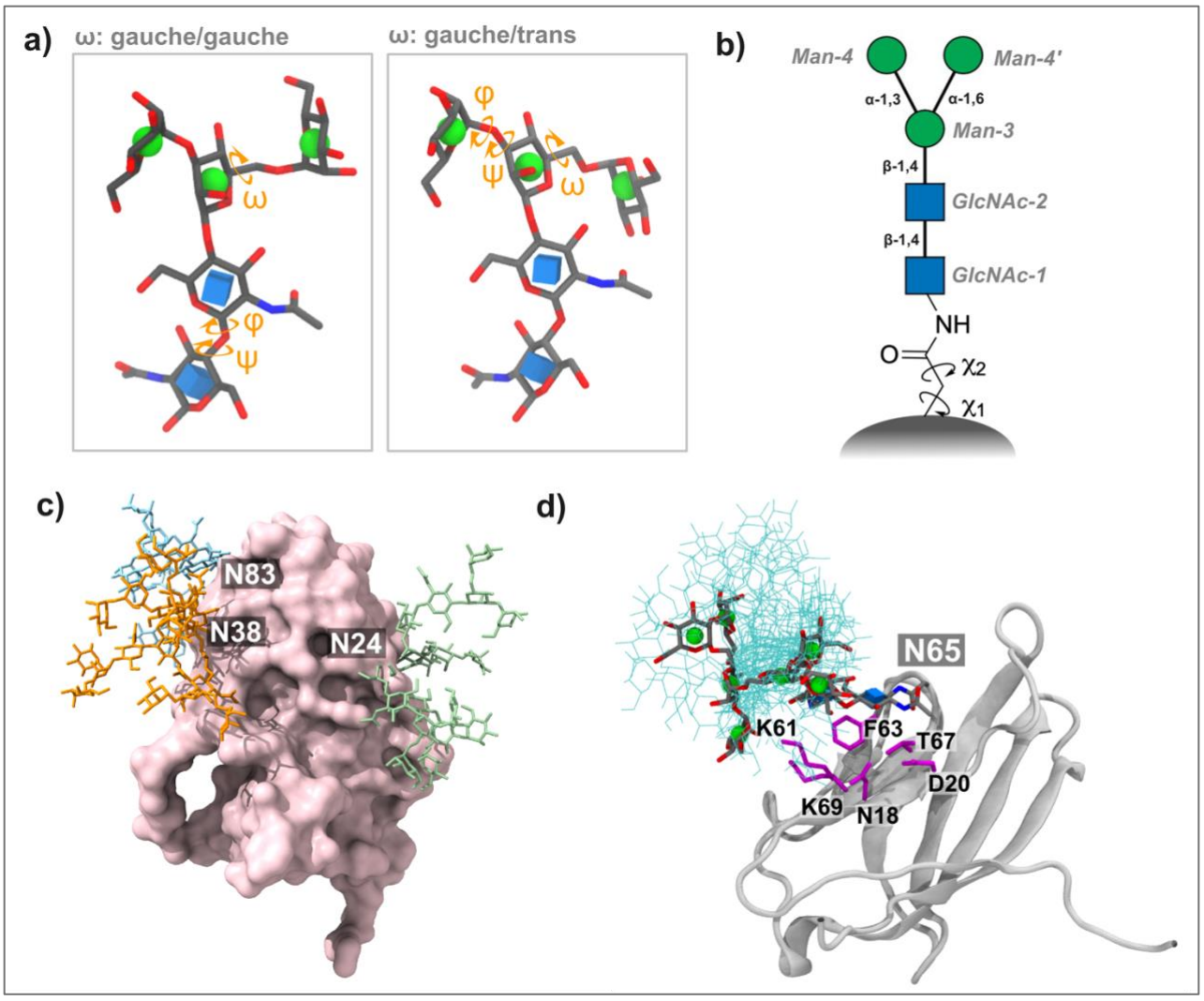


Figure 2: **Conformational characteristics of N-linked glycans**, partially adapted from (Gazaway et al., 2025). **a)** Two representative structures of paucimannose, the conserved pentasaccharide core that is present in all human N-linked glycans, are shown as 3D-SNFG icons (Thieker et al., 2016). The ψ, φ, and ω bonds of the glycosidic linkages are identified by orange arrows. In the left image, the ω torsion angle is in the *gauche/gauche* conformation, which is defined as ~60° with respect to the C4-C5-C6-O6 atoms. In the right image, ω torsion angle is depicted in the *gauche/trans* conformation, which is defined by ~180°. In this conformation, the terminal mannose residue is oriented towards the glycan core. **b)** 2D-SNFG depiction of paucimannose linked to an asparagine side chain. The χ_1 and χ_2 torsion angles of the asparagine are defined by the N-Cα-Cβ-

Cγ and Cα-Cβ-Cγ-N_{Asn} atoms, respectively. The glycosidic linkage between two residues is labeled in black, and the names the glycan residues are labeled in gray. **c)** Surface representation of erythropoietin (PDB: 1EER) glycosylated with paucimannose at its three glycosylation sites (Syed et al., 1998). The glycosylated model was generated with the Glycoprotein Builder available on GLYCAM-Web. The model is shown with four distinct χ_1/χ_2 conformers that were observed in glycoprotein crystal structures (Petrescu et al., 2004). The separate conformers are shown simultaneously to illustrate the conformational space accessible to the N-linked glycans. **d)** Ribbon representation of the N-terminal domain of CD2, which contains a high-mannose glycan at N65 (PDB: 1GYA) (Wyss et al., 1995). Side chains of the surface protein residues that had exhibited nuclear Overhauser enhancements with the glycan are magenta. The glycan of the primary model is shown in 3D-SNFG icons, and the glycan configurations in the other 17 models of the solution structure are shown as cyan lines.

Although the asparagine—glycan linkage increases the conformational space accessible to N-linked glycans, favorable contact between the glycans and the underlying protein has been shown to influence glycan presentation. For example, in solution NMR analysis of the N-terminal domain of CD2, nuclear Overhauser enhancements (Gronenborn & Clore, 1985) between the high-mannose glycan and neighboring protein residues indicated that the glycan was closely associated with the underlying protein surface, which was critical for stabilizing the protein structure (Figure 2d) (Wyss et al., 1995). Additionally, pseudocontact shift NMR was used to show that the N-linked glycans on the IgG domain of CEACAM1 was proximal to the protein surface, presumably to shield hydrophobic pockets on the protein from solution (Rogals et al., 2022). Constraining N-linked glycans to the protein surface through favorable glycan—protein contact may have

biological implications. For example, the glycans in transferrin Tf-2 glycoform shielded hydrophobic surfaces on the protein—a function that was missing in the Tf-1 glycoform, which is associated with normal pressure hydrocephalus (Nagae et al., 2014).

Approaches and challenges in computational modeling of glycoproteins

Understanding the three-dimensional structure of glycoproteins is of considerable interest for elucidating structure-function relationships, for example, in characterizing their biological functions and in the development of biosimilar therapeutics (Borza et al., 2020; Liu et al., 2020). However, due to their intrinsic disorder and microheterogeneity, full N-linked glycans are very rarely observed in experimental structures of glycoproteins. Recent quantification of the size of N-linked glycans observed in the PDB revealed that less than 3% were resolved beyond the trisaccharide core (Gazaway et al., 2025). To overcome this limitation, complete glycoprotein models can be generated *in silico* by attaching N-glycan structures to surface asparagine residues, an approach that is accessible through web-based platforms, including GLYCAM-Web, CHARMM GUI (Park et al., 2019), and Glycoshape (Ives et al., 2024). The glycoprotein structures generated by these tools are often subjected to Molecular Dynamics (MD) simulation to model the conformational flexibility N-linked glycans on the protein surface. These simulations are widely employed to study the biological roles of glycoproteins, for example, in shielding viral surface proteins from immune recognition, (Stewart-Jones et al., 2016; Zhao et al., 2020), mediating antibody accessibility (Grant et al., 2020), and elucidating the sources of microheterogeneity during N-glycan maturation (de Haan et al., 2023; Hang et al., 2015; Quintana et al., 2024).

The development of parallelized simulation code has enabled macromolecular systems, including glycoproteins, to be modelled at the μs timescale (Lee et al., 2018). However, achieving

convergence of the distribution of glycan conformers even at this prolonged timescale remains a challenge. For example, a 1 μs simulation of the conserved N-glycan pentasaccharide core was not sufficient to achieve convergence of the conformational states of the free glycan (Jo et al., 2016). Further, independent 1 μs simulations of CD2, initiated with different glycan orientations, did not converge to the same distribution of conformations of the glycosylated asparagine side chain (Gazaway et al., 2025). Moreover, because MD requires expertise in simulation theory and programming, as well as access to high-performance computing resources, more accessible modeling tools have been developed to model glycan ensembles without the need for prolonged glycoprotein simulations. These tools employ conformational sampling (Lemmin & Soto, 2019) or the addition of pre-simulated glycan structures onto glycosylation sites (Ives et al., 2024; Tsai et al., 2024) to model an ensemble of glycan conformers that do not sterically clash with the protein surface. However, whether these ensembles reflect the distributions of glycan orientations that exist in solution has not been validated experimentally. Moreover, these tools neglect the influence of protein–glycan contacts, which can stabilize glycan orientations associated with the protein surface, as observed in CD2 and in the CEACAM-IgG domain (Rogals et al., 2022; Wyss et al., 1995). Therefore, characterizing the protein environment around N-glycosylation sites is key to understanding how protein—glycan interactions influence the presentation of the glycan.

Previous analyses have characterized the protein environment surrounding N-glycosylation sites by surveying experimental glycoprotein structures available in the PDB. For example, a comprehensive analysis of the protein residues within 5 Å of 626 glycosylation sites was conducted to identify common features of the protein environment that may enhance OST efficiency (Petrescu et al., 2004). This study found that aromatic amino acids were enriched directly upstream and downstream of glycosylation sites, whereas basic amino acids were largely

absent from the protein environment. Additionally, similar studies have been conducted to elucidate features of the protein region near N-glycosylation sites that modulate N-glycan maturation (Suga et al., 2018; Thaysen-Andersen & Packer, 2012). The study by Suga et al. found that asparagine, aspartate, and leucine were upregulated near immature N-linked glycans, leading the authors to suggest that interactions between these residues and N-linked glycans reduced the accessibility of processing enzymes (Suga et al., 2018).

While the studies to date provide a valuable insight into the amino acids that are proximal to N-glycosylation sites, the extent to which these residues form energetically favorable (or unfavorable) contacts with N-linked glycans has not been assessed. *In silico* interaction energy calculations, such as Molecular Mechanics Generalized Born Surface Area (MM/GBSA) are frequently employed to measure the binding free energy between a ligand and receptor (Genheden & Ryde, 2015; Gohlke & Case, 2004; Sood et al., 2018; Still et al., 1990). MM/GBSA quantifies the enthalpic component of the binding interaction energy between two molecules as the sum of gas-phase molecular mechanics energies—including electrostatics and van der Waals—and energy of desolvation, which is estimated using the Generalized Born implicit solvent model (Hadden et al., 2015). Here, we employed MM/GBSA calculations to quantify the interaction energies between N-linked glycans and neighboring protein residues in glycoprotein structures from the PDB, allowing us to assess the relative strength of the interactions. Identifying commonly occurring interactions between N-linked glycans and nearby protein residues can provide insight into features of the protein environment influence the distribution of glycan orientations that exist in a conformational ensemble modeled in solution.

CHAPTER 2

MATERIALS AND METHODS

Dataset generation

The GlyFinder web tool (available on GLYCAM-Web) was used to query the RCSB Protein Data Bank (RCSB.org) (Berman et al., 2000) to identify glycoprotein structures that contained at least one N-linked glycan with the core trisaccharide Manβ1-4GlcNAcβ1-4GlcNAcβ-ASN or a larger structure. This trisaccharide core is shared across all N-linked glycan structures, independent of processing. A separate dataset of core-fucosylated glycans was generated by querying GlyFinder for glycoproteins containing at least a Fucα1-6-GlcNAcβ-ASN structure. Each dataset was refined to include only structures with a resolution of <=2.5 Å and a maximum oligosaccharide B-factor of 30. Redundant glycoprotein systems with >95% amino acid sequence similarity were identified using the *Seq* object available in the Biopython package (Cock et al., 2009). N-linked glycans that contained two core fucose residues linked to the same GlcNAc (i.e. α1,6-linked and α1,3-linked fucose) residue were excluded from the dataset. A complete list of the glycoprotein systems modeled in this study is included in Table S1 and Table S2.

Preparation of glycoprotein systems for MD

System preparation for MD, MD simulations, and MM/GBSA calculations of each glycoprotein system were performed using programs available on Amber24 (Case et al., 2025). Molecules that are not compatible with AMBER, including crystalizing agents, ligands, and co-factors, were removed from each glycoprotein system using *pdb4amber*. The *Reduce* program was employed to

determine the preferred orientation of glutamine, non-glycosylated asparagine, and histidine residues, and to assign protonation to Nε or Nδ based on hydrogen bonding analysis (Word et al., 1999). The N- and C-termini of unresolved protein segments, which were present in flexible regions such as loops, were capped at the with an N-methyl and acetyl group, respectively, to prevent the artificial exposure of charged backbone atoms. Hydrogen atoms were then added to the modified glycoprotein structures using tLeap. The protonation of ionizable amino acid residues (ASP, ARG, GLU, LYS, HIS) were assigned according to their standard forms at neutral pH. Each system was neutralized with Na^+ and Cl^- counter ions and explicitly solvated in an octahedral periodic box of OPC water molecules (Izadi et al., 2014), with at least 10Å of water between the solute and box boundary. Water molecules that were originally present in the crystal structures were retained.

Energy minimization and molecular dynamics (MD) simulations

Each glycoprotein system was simulated with the ff19SB protein (Tian et al., 2020) and GLYCAM-06j carbohydrate (Kirschner et al., 2008) force fields, with parameters for monovalent and divalent ions optimized for use with the OPC water model (Li et al., 2020; Sengupta et al., 2021). The experimental structure of each glycoprotein system was energy minimized and equilibrated following the protocol established by Roe & Brooks (Roe & Brooks, 2020). The structure generated by the final step of equilibration was subsequently used as the initial coordinates for MD simulation. Each solvated glycoprotein system was simulated independently for 1000 picoseconds (1 nanosecond) under constant pressure (1 atm) and temperature (300K) conditions, employing the *pmemd.cuda* simulation code available in AMBER. A Cartesian restraint with a weight of 1.5 $kcal \cdot mol^{-1} \cdot Å^{-2}$ was applied to Cα atoms of the protein backbone to prevent

the major fluctuations in the protein structure. Non-bonded interactions were calculated for each atom using a distance cutoff of 8Å. An integration timestep of 0.002 picoseconds was used, and coordinates were saved every 10 picoseconds.

N-linked glycan—protein interaction energy calculations

For each glycoprotein system, interaction energies were calculated for 100 snapshots that were extracted at 10 ps intervals from the 1 ns simulation. Explicit solvent and ions were removed from the system topology prior to energy calculations using *cpptraj* (Roe & Cheatham, 2013). The interactions energies between the N-linked glycans and the protein were estimated using Molecular Mechanics Generalized Born Surface Area (MM/GBSA) method (Genheden & Ryde, 2015). In this approach, the binding energy between two pairs of molecules is the sum of gas-phase molecular mechanics interactions and solvation energy, which is estimated through an implicit solvent model (Still et al., 1990). The final interaction energy between a pair of residues was the average value calculated between the residues from 100 snapshots. The MM/GBSA calculations were performed using the MMGBSA.MPI.py code in AMBER (Hawkins et al., 1996). The *idecomp=3* option was employed to calculate pairwise, per-residue energies (Lees et al., 2017). An internal solute dielectric constant of 1 was applied to all systems.

Analyzing the interaction energy between each glycan residue and the protein surface

The total interaction energy calculated by MM/GBSA between a pair of solute residues is the combination of gas-phase molecular mechanics (MM) energies and energy of solvation approximated by the Generalized Born implicit solvent model (Onufriev & Case, 2019; Still et al., 1990). The MM component is the sum van der Waals (E_{ij}^{vdW}) and electrostatic ($E_{ij}^{electrostatic}$) contacts

measured in a vacuum. The energy of solvation approximates the impact of solvation on the binding interaction, which consists of a polar and non-polar component. Therefore, the total interaction energy between residues i and j is the sum of each energy term:

$$E_{ij}^{total} = E_{ij}^{vdW} + E_{ij}^{electrostatic} + E_{ij}^{polar\ solvation} + E_{ij}^{non-polar\ solvation} \quad (1)$$

For each glycoprotein system, to determine the interaction energy of each glycan residue with the protein surface, pairwise interaction energies between each glycan residue and non-glycosylated protein residues were extracted. The total interaction energy for each glycan residue was then calculated as the sum of its interaction energies with all non-glycosylated protein residues (Equation 2).

$$E_i^{glycan\ residue-protein} = \sum_{j=1}^{N_{protein\ residues}} E_{ij}^{total} \quad (2)$$

Amino acid enrichment analysis

Interaction energies between each GlcNAc-1 residue in the dataset and individual protein residues from the corresponding glycoprotein structures were analyzed to identify protein residues that contributed most favorably to glycan binding. Surface residues of each glycoprotein were identified using a 1.4 Å-radius probe to calculate the solvent accessibility of each protein residue using the *surf* command available within *cpptraj* (Roe & Cheatham, 2013). N-linked glycans were stripped from the glycoprotein structure prior to measuring solvent accessibility. Protein residues with positive solvent accessibility were identified as surface residues. To define surface protein residues that formed favorable contact with the GlcNAc-1, the most energetically favorable protein residue—GlcNAc-1 pair was identified. This interaction corresponded to the minimum E_{ij}^{total} value across all protein residue—GlcNAc-1 interactions in the dataset. The other surface protein residues in the dataset were considered to form favorable interactions with GlcNAc-1 if their E_{ij}^{total} was at

least 10% of the most favorable protein residue—GlcNAc-1 interaction. The frequency of each amino acid among surface protein residues forming favorable contact with GlcNAc-1 was calculated (Equation 3). The background distribution of each amino acid type was determined by calculating the percentage of the amino acid among surface protein residues not forming favorable contact with GlcNAc-1 (Equation 4).

To assess whether each amino acid type was significantly enriched participating in favorable contact with GlcNAc-1 compared to its background distribution, a 2x2 contingency table was constructed for each amino acid type (Table 1) (Adelin, 2017). A χ^2 test of independence was then performed on the table using the *chi2_contingency* function within the *scipy.stats* module available in Python (Virtanen et al., 2020).

$$\%AA_{contact} = \frac{N_{AA,surface,contact}}{N_{total,surface,contact}} \; x \; 100 \tag{3}$$

$$\%AA_{background} = \frac{N_{AA,surface,non-contact}}{N_{total,surface,non-contact}} \; x \; 100 \tag{4}$$

This analysis protocol was applied the evaluate the amino acid enrichment of protein residues forming favorable interactions with the following glycans residues: (i) GlcNAc-1 in non-core-fucosylated glycans, (ii) GlcNAc-1 in core-fucosylated glycans, and (iii) core fucose in core-fucosylated glycans.

	Amino acid X	All other amino acids
Formed favorable contact with GlcNAc-1 (≤ *threshold* kcal·mol^{-})	*a*	*c*
Did not form favorable contact with GlcNAc-1 (> *threshold* kcal·mol^{-})	*b*	*d*

Table 1: **Example 2x2 contingency table for generic amino acid X**, where ***a*** = number of amino acid X residues on the protein surface that formed favorable contact with GlcNAc-1, **b** = number

of amino acid X residues on the protein surface that did not form favorable contact with GlcNAc-1, **c** = number of all other surface protein residues that formed favorable contact with GlcNAc-1, and **d** = number of all other surface protein residues that did not form favorable contact with GlcNAc-1

CHAPTER 3

RESULTS

The chitobiose core of non-core-fucosylated glycans formed the strongest interactions with the protein surface.

In the solution structure of the N-terminal domain of CD2, the interactions observed between the high-mannose glycan and the protein surface included CH-π interactions between F63 and GlcNAc-1 and electrostatic interactions between K69/K61 and a distal mannose residue of the α1,6 arm (Figure 2d) (Wyss et al., 1995). Moreover, analysis of glycoprotein crystal structures indicated that the chitobiose core, the first two GlcNAc residues of the glycan, was most frequently positioned in close proximity to the protein surface (Jo et al., 2013; Petrescu et al., 2004). To evaluate whether proximal or distal residues of N-linked glycans contribute more to energetically favorable interactions with the protein surface, we employed MM/GBSA calculations to quantify N-glycan—protein interaction energies across experimental glycoprotein structures. A dataset of glycoproteins with N-linked glycans that contained at least the trisaccharide core (Manβ1-4GlcNAcβ1-4GlcNAc) (Figure 3a) was generated using the GlyFinder web tool available on GLYCAM-Web. N-glycosylation sites that contained a single GlcNAc were excluded to avoid structures treated with endo-β-N-acetylglucosaminidase H (Endo H) (Chang et al., 2007; Davis, 2010). To ensure that the dataset included only high-quality structures, the initial list was filtered to select structures with a resolution of less than or equal to 2.5 Å and an average oligosaccharide B-factor of less than or equal to 30. The refined dataset employed in this study contained 199 predominantly high-mannose glycans across 78 non-redundant glycoprotein structures (Table S1).

Prior to performing interaction energy calculations, the 78 glycoprotein structures were energy minimized and subjected to a short 1 ns MD simulation. Energy minimization ensured that atomic clashes present in the experimental coordinates were resolved to prevent artificially unfavorable interaction energies between atoms. Moreover, the purpose of the brief MD simulation of each glycoprotein system was not to sample conformational changes in the glycan structure, but to model short time-scale vibrational motion while retaining the experimental conformation of the glycan. The interaction energy of each glycan residue was calculated with all protein residues of the glycoprotein, excluding glycosylated asparagine residues.

To compare the interaction strength of proximal versus distal residues of N-linked glycans, the interaction energy between each glycan residue and the underlying protein was calculated by summing all pairwise interaction energies calculated between the glycan residue and non-glycosylated protein residues (Equation 2; see Methods). Comparison of the range of interaction energies calculated between glycan residue and the protein surface shows that the core GlcNAc residues, GlcANc-1 and GlcNAc-2, exhibited the greatest variation, with interaction energies between -31.2 and -0.8 kcal·mol^{-1} for GlcNAc-1 and between -32.4 and -0.07 kcal·mol^{-1} for GlcNAc-2 (Figure b). Despite this variability, comparison of the median interaction energies of core and distal glycan residues indicates that, in general, the strongest interactions with the protein surface were measured with the core GlcNAc residues (Figure 3b).

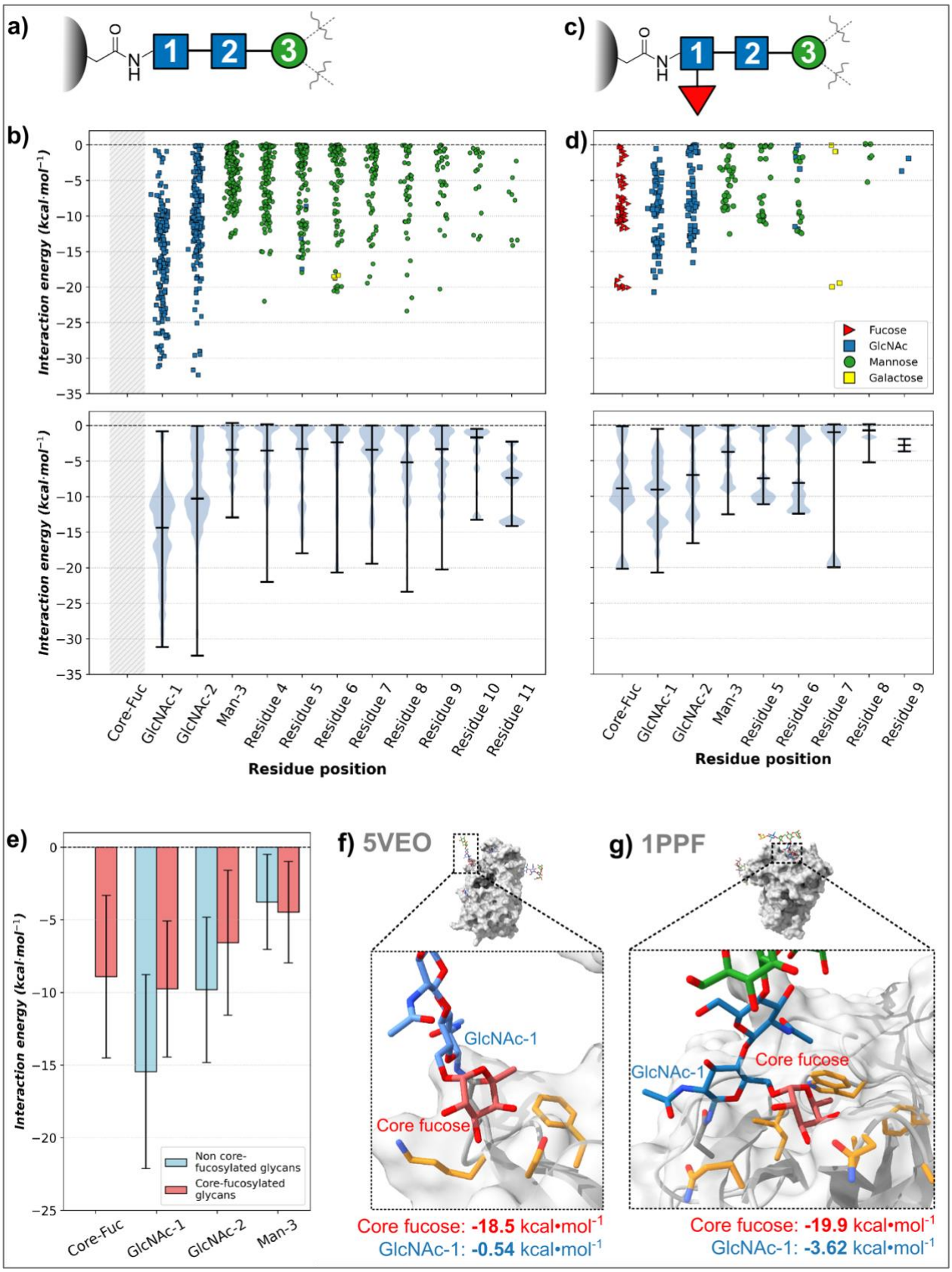

a)
c)
b)
d)
Interaction energy (kcal·mol⁻¹)
Fucose
GlcNAc
Mannose
Galactose
Core-Fuc
GlcNAc-1
GlcNAc-2
Man-3
Residue 4
Residue 5
Residue 6
Residue 7
Residue 8
Residue 9
Residue 10
Residue 11
Residue position
e)
Non core-fucosylated glycans
Core-fucosylated glycans
f) 5VEO
g) 1PPF
GlcNAc-1
Core fucose
Core fucose: -18.5 kcal•mol-1
GlcNAc-1: -0.54 kcal•mol-1
Core fucose: -19.9 kcal•mol-1
GlcNAc-1: -3.62 kcal•mol-1

Figure 3: **Comparison of the interaction energy between each glycan residue and the protein surface in non-core-fucosylated and core-fucosylated glycans. a)** SNFG depiction of the N-linked trisaccharide core that was the minimal structure present in the glycans of the non-core-fucosylated glycan dataset mined from the PDB. Core residues are labeled GlcNAc-1, GlcNAc-2, and Man-3. **b)** *Top*. The interaction energy between each glycan residue in the **dataset of non-core-fucosylated glycans** and the protein surface is plotted according to the residue's position within the glycan. The color and shape of the datapoints correspond to the monosaccaride's SNFG depiction. *Bottom.* Violin plot showing the distribution of interaction energies plotted above. The black bars indicate the minimum, median, and maximum interaction energies calculated at each glycan position. The width of the violin represents the density of data points at a given interaction energy. **c)** SNFG depiction of the core-fucosylated N-linked tetrasaccharide core. The minimal structure present in the glycans of the core-fucosylated glycan dataset mined from the PDB was GlcNAc-1 with an α1,6-linked fucose. **d)** *Top*. The interaction energy between each glycan residue in the **dataset of core-fucosylated glycans** and the protein surface is plotted according to the residue's position within the glycan. *Bottom.* Violin plot showing the distribution of plotted interaction energies. **e)** The average interaction energies of the core residues in non-core-fucosylated (light blue) and core-fucosylated glycans (red) are shown. The error bars represent the standard deviation of the interaction energies at each glycan position. **f-g)** Representative N-linked glycans with strong protein—core-fucose interactions and comparatively weaker protein—GlcNAc-1 interactions. Zoomed image of the N101 glycan of murine ectonucleotide pyrophosphatase (PDB: 5VEO) (Gorelik et al., 2017) (f) and the N159 glycan of human leukocyte elastase (PDB: 1PPF) (Bode et al., 1986) (g) illustrate contacts between the core fucose and the protein surface (gray, translucent surface). The side chains of neighboring protein residues within

5 Å of the core fucose are shown in orange. The interaction energies between the core fucose and GlcNAc-1 residues and the neighboring protein surface are listed below the zoomed image.

Core fucose mediates contact between N-linked glycans and the protein surface

During N-glycan maturation in the Golgi complex, FUT8 may catalyze the addition of an α1,6-linked fucose onto GlcNAc-1 (Figure 1d) (Boruah et al., 2020). This modification plays a critical role in several biological processes, including IgG Fc recognition during antibody-dependent cell-mediated cytotoxicity (Yamaguchi & Barb, 2020). An NMR study of the IgG Fc structure suggested that core fucosylation restricted the conformational freedom on the N162 glycan, reducing its affinity for the Fcγ receptor (Falconer et al., 2018). Because core fucose has been observed to influence the dynamics of N-linked glycans, we evaluated its influence on protein—glycan contact. GlyFinder was used to generate a separate dataset of glycoproteins with N-linked glycans containing an α1,6-linked core fucose. Because complex, core-fucosylated N-linked glycans are not susceptible to treatment with Endo H (Freeze & Kranz, 2010), this dataset included glycans which contained at least a GlcNAc-1 and a core fucose (Fucα1-6-βGlcNAc). N-linked glycans containing an α1,3-linked fucose and both an α1,6 and α1,3-linked fucose on GlcNAc-1 were excluded from this dataset. The initial results were also refined to include structures with at least a 2.5 Å resolution and a mean oligosaccharide B-factor less than or equal to 30. The final dataset of core-fucosylated N-linked glycans included 62 glycosylation sites across 31 non-redundant glycoprotein structures. These structures were energy minimized and subjected to a 1 ns MD simulation prior to MM/GBSA calculations. The parameters employed for the simulations and interaction energy calculations were identical to those applied to the glycoprotein structures containing non-core-fucosylated glycans.

To assess the involvement of core fucose in the contact between N-linked glycans and the underlying protein surface, the interaction energy between each residue in the core-fucosylated glycans and the underlying protein was calculated. As was observed in the dataset of non-core-fucosylated glycans, the median interaction calculated between the protein and the chitobiose core was stronger than that of Man-3 (Figure 3d). In comparison to the interaction energies observed between the protein surface and the chitobiose core of non-core-fucosylated glycans (Figure 3b), interaction between the chitobiocore of core-fucosylated glycans and the protein surface were generally weaker (Figure 3d-e). Moreover, the most favorable interaction between non-core-fucosylated GlcNAc-1 residues and the protein surface was -31.2 kcal·mol^{-1} (Figure 3b), whereas the interaction energies with core-fucosylated GlcNAc-1 residues ranged from -0.54 and -20.7 kcal·mol^{-1} (Figure 3d). This difference in the minimum interaction energy suggests that the addition of core fucose attenuates the strength of interactions that GlcNAc-1 can form with the protein surface. Despite the attenuation of interaction strength, the range of interactions between core fucose and the proteins surface (-0.16 and -20.2 kcal·mol^{-1}) was similar to that of core-fucosylated GlcNAc-1 residues (Figure 3d), suggesting that favorable interactions with core fucose compensated for the reduced favorability of protein—GlcNAc-1 interactions.

While the overall range of interaction energies was similar for core fucose and core-fucosylated GlcNAc-1 residues, the distribution of protein—core-fucose interaction energies revealed a subset of residues with markedly more favorable interactions, between -18.5 and -20.2 kcal·mol^{-1} compared to the other core fucose residues (-11.9 to -0.16 kcal·mol^{-1}) (Figure 3d). Analysis of the glycosylation sites that exhibited the strongest interactions between core fucose and the protein surface revealed structures in which the corresponding GlcNAc-1 interaction energy was more than five times weaker than that of core fucose. For example, in the N101 glycan

of the murine ectonucleotide pyrophosphatase (Gorelik et al., 2017), the fucose C6 methyl group interacted with a phenylalanine residue on the protein surface, which was reflected in favorable van der Waals energy calculated between the two residues (Figure 3f). Hydrophobic contact between core fucose and the protein surface was also observed in the N159 glycan of human leukocyte elastase (Bode et al., 1986), where the fucose C6 interacted with a hydrophobic patch formed by tryptophan, phenylalanine, valine, and leucine residues (Figure 3g). These observations suggest that core fucose can form strong interactions with the protein surface independently of the core GlcNAc residue.

Charged and aromatic residues predominate in energetically favorable interactions with GlcNAc-1, but electrostatic contributions are offset by energy of desolvation

Because favorable glycan—protein interactions primarily involved residues in the glycan core, we examined the protein environment to determine which amino acid types were enriched in forming favorable contact with GlcNAc-1 in non-fucosylated glycans. Because nearly all GlcNAc-1–protein residue pairs had negative (favorable) interaction energies, a threshold was established to identify the most favorable interactions in the dataset. Among all interactions between GlcNAc-1 residues and protein residues, the strongest total interaction energy (E_{ij}^{total}, see Equation 1) calculated by MM/GBSA was observed between K391 and GlcNAc-1 of the N67 glycan in neuraminidase N6, with an E_{ij}^{total} of -8.97 kcal·mol^{-1} (Sun et al., 2014). 10% of this minimum interaction energy (-0.897 kcal·mol^{-1}) was used as the threshold to identify surface protein residues of each protein system that made appreciable interactions with GlcNAc-1. For example, in the N-terminal adhesion domain of CD2, surface protein residues that met this criterion included D20, P22, F63, and T67 (Figure 4a, left). The frequencies of amino acid types forming appreciable

interactions with GlcNAc-1 were compared to the background frequencies among surface residues with E_{ij}^{total} above the threshold. The differences in the frequency of each amino acid among these populations are shown in Figure 4a (right) to assess whether each amino acid was enriched relative to its background distribution. This analysis revealed that among hydrophilic amino acids, aspartate and threonine were significantly enriched in interactions with GlcNAc-1 compared their background presence on the protein surface. Additionally, aromatic amino acids (tryptophan and tyrosine) and isoleucine were also significantly enriched.

The total interaction energy between two residues is the sum of contributions from gas-phase molecular mechanics energies and solvation energy, which is estimated by the Generalized Born model. Therefore, the E_{ij}^{total} is affected by approximations of the GB model. As a result, favorable electrostatic interactions observed in the gas-phase are often offset by the energetic cost of desolvation. To identify the favorable protein—GlcNAc-1 interactions independently of solvation effects, the electrostatic component of the interaction energies ($E_{ij}^{electrostatic}$) was analyzed separately. The most favorable gas-phase electrostatic interaction was observed between D560 and the N590 GlcNAc-1 in the human toll-like receptor 8 (Tanji et al., 2015), with an $E_{ij}^{electrostatic}$ of -14.7 kcal·mol^{-1}. 10% of this energy (-1.47 kcal·mol^{-}) was used as the threshold to identify surface amino acid residues from the dataset with the most favorable gas-phase electrostatic interactions. For example, in CD2, the $E_{ij}^{electrostatic}$ calculated between D20 and N65 GlcNAc-1 was -3.52 kcal·mol^{-1} (Figure 4b, left).

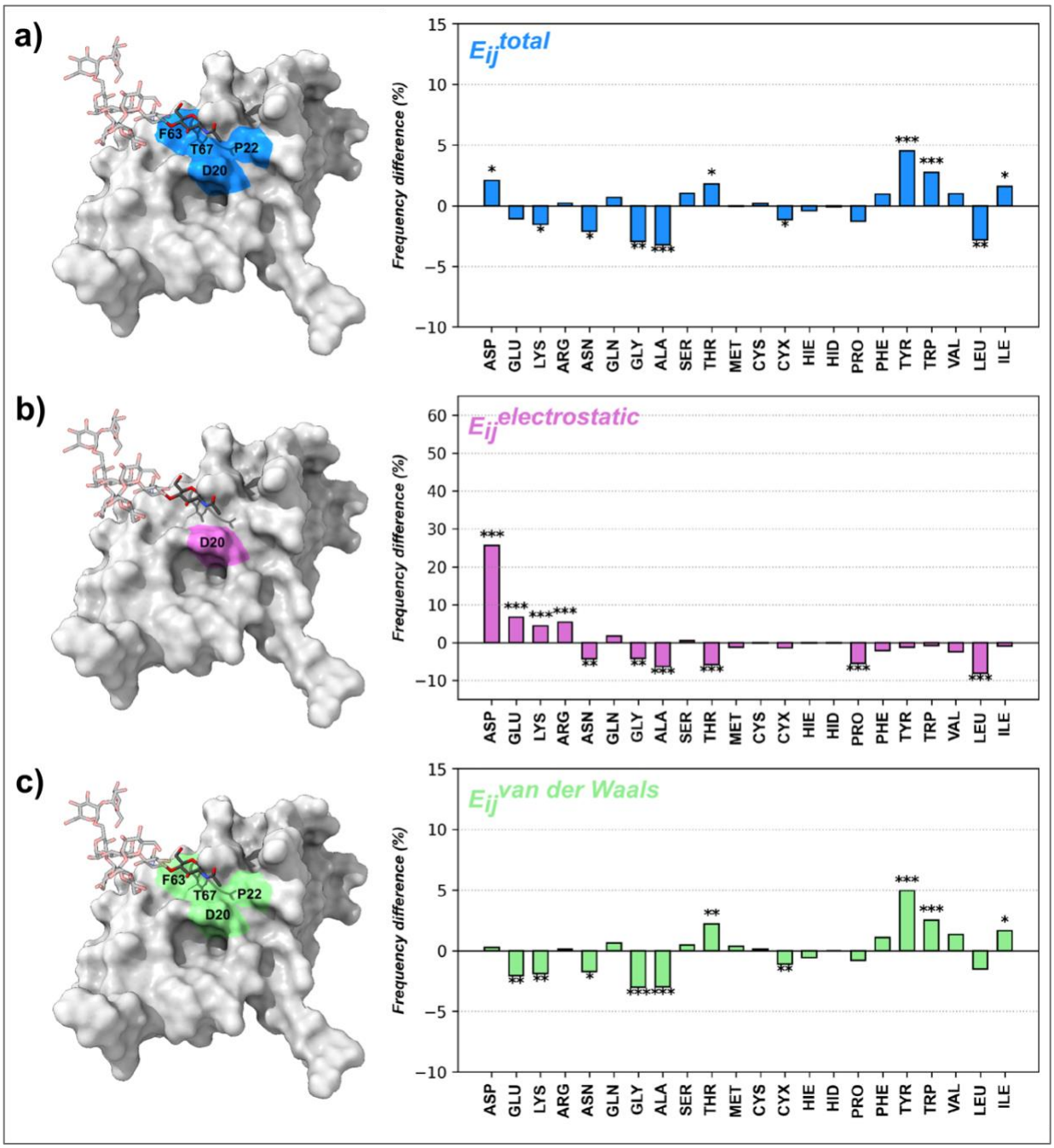


Figure 4: **Amino acid enrichment among surface protein residues involved in energetically favorable interactions with non-core-fucosylated GlcNAc-1 residues, decomposed by energy term. a)** *Left*. Surface representation of the N-terminal domain of CD2 with N-linked glycans shown in stick representation (PDB: 1GYA) (Wyss et al., 1995). Surface protein residues with a total interaction energy (E_{ij}^{total}) with GlcNAc-1 less than or equal to the threshold value of -0.897

kcal·mol^{-1} are colored blue. *Right*. Amino acid enrichment among surface protein residues involved in favorable interactions with GlcNAc-1, where interaction favorability was determined based on E_{ij}^{total}. The differences in amino acid frequencies were calculated by comparing the distribution of each amino acid among surface residues with $E_{ij}^{total} \leq -0.897$ kcal·mol^{-1} for GlcNAc-1 to the distribution among residues with $E_{ij}^{total} > -0.897$ kcal·mol^{-1}. CYX and CYS refer to cysteine residues in a disulfide bond and not in a disulfide bond, respectively. HIE and HID refer to histidine residues protonated at Nε and Nδ, respectively. **b)** *Left*. The surface protein residue of CD2 in which the electrostatic component of the interaction energy ($E_{ij}^{electrostatic}$) with GlcNAc-1 was less than or equal to the threshold value of -1.47 kcal·mol^{-} is colored pink. *Right*. Amino acid enrichment among surface protein residues that formed favorable gas-phase electrostatic interactions with GlcNAc-1. The differences in amino acid frequencies were calculated by comparing the distribution of each amino acid among surface residues with $E_{ij}^{electrostatic} \leq -1.47$ kcal·mol^{-} for GlcNAc-1 the distribution among residues with $E_{ij}^{electrostatic} > -1.47$ kcal·mol^{-}. **c)** *Left*. Surface protein residues of CD2 in which the van der Waals component of the interaction energy ($E_{ij}^{van\ der\ Waals}$) with GlcNAc-1 was less than or equal to the threshold value of –0.326 kcal·mol^{-} are colored green. *Right*. Amino acid enrichment among surface protein residues that formed favorable gas-phase van der Waals interactions with GlcNAc-1. The differences in amino acid frequencies were calculated by comparing the distribution of each amino acid among surface residues with $E_{ij}^{van\ der\ Waals} \leq -0.326$ kcal·mol^{-1} with GlcNAc-1 and the distribution among residues with $E_{ij}^{van\ der\ Waals} > -0.326$ kcal·mol^{-1}. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Comparison of the frequency of amino acids among residues involved in favorable electrostatic interactions with GlcNAc-1 and those in background distribution shows that charged residues at neutral pH were significantly enriched (4b, right). The lack of significant enrichment of these residues in the analysis of E_{ij}^{total} (Figure 4b) indicates that the strong electrostatic interactions measured in the gas phase were largely offset by energetic penalty of desolvating the protein to form the interaction with GlcNAc-1. Moreover, amino acid enrichment analysis of protein residues that formed favorable van der Waals interactions with GlcNAc-1 in the gas phase ($E_{ij}^{van\ der\ Waals}$) more closely resembled the enrichment observed for the total interaction energy (Figure 4c). This indicates that the total binding energy calculated by MM/GBSA between GlcNAc-1 of non-core-fucosylated glycans and neighboring protein residues is dominated by van der Waals contributions because the strong electrostatic interactions, although present in the gas phase, are largely cancelled out by the energy penalty of desolvation. Counts of each amino acid in the GlcNAc-1-interacting and background populations used for the enrichment analyses are reported in Table S3.

Distinct amino acids were enriched in interactions with GlcNAc-1 and the fucose of core fucosylated N-linked glycans

We then evaluated the amino acid enrichment of surface protein residues involved in energetically favorable interactions with GlcNAc-1 in 62 core-fucosylated glycans in the dataset to characterize their interactions with the protein surface. To identify protein residues that interacted favorably with GlcNAc-1, the same analysis protocol used to characterize non-core-fucosylated glycans was applied. The most favorable E_{ij}^{total} calculated by MM/GBSA was observed between Y2 and GlcNAc-1 of the anti-HIV V3 Fab N60 glycan, with an E_{ij}^{total} of -8.17 kcal·mol^{-1} (Gorny et al., 2011). Protein residues with total interaction energies with the GlcNAc-1 residues in the dataset

more favorable than −0.817 kcal·mol^{-1} (10% of the most favorable interaction) were classified as forming favorable interactions. Comparison of the amino acid frequencies among surface protein residues involved in favorable interactions with GlcNAc-1 verses those not classified as forming favorable interactions revealed that arginine, glycine, serine, and valine were significantly enriched (Figure 5a). This enrichment profile was distinct from that of GlcNAc-1 of non-core-fucosylated glycans (Figure 4a, right), which included aspartate, threonine, tyrosine, tryptophan, and isoleucine. In fact, most of these amino acids were underrepresented among protein residues forming favorable interactions with the GlcNAc-1 of core fucosylated glycans. Tyrosine and tryptophan were enriched but not significantly (Figure 5a).

We next analyzed the protein residues involved in energetically favorable interactions with the core fucose residue to evaluate whether the same amino acids enriched in GlcNAc-1 of core-fucosylated glycans were also enriched interacting with the core fucose. Favorable protein—core fucose interactions were identified using the same energy threshold applied to protein—GlcNAc-1 interactions (-0.817 kcal·mol^{-1}). Amino acids significantly enriched among surface protein residues involved in favorable interactions with core fucose included arginine, phenylalanine, and tryptophan (Figure 5b). Arginine was the only residue significantly enriched in interactions with core fucose and GlcNAc-1. Phenylalanine and tryptophan were sightly enriched in protein residues that formed favorable interactions with GlcNAc-1 in core-fucosylated glycans, but not significantly (Figure 5a). Moreover, tryptophan was significantly enrichment in interactions with GlcNAc-1 in non-core-fucosylated glycans (Figure 4a, right) and with core fucose (Figure 5b), but not with GlcNAc-1 in core-fucosylated glycans (Figure 5a). The distinct amino acid enrichment between GlcNAc-1 and core fucose indicates that, despite being linked to GlcNAc-1, core fucose does not form energetically favorable interactions with the same protein residues.

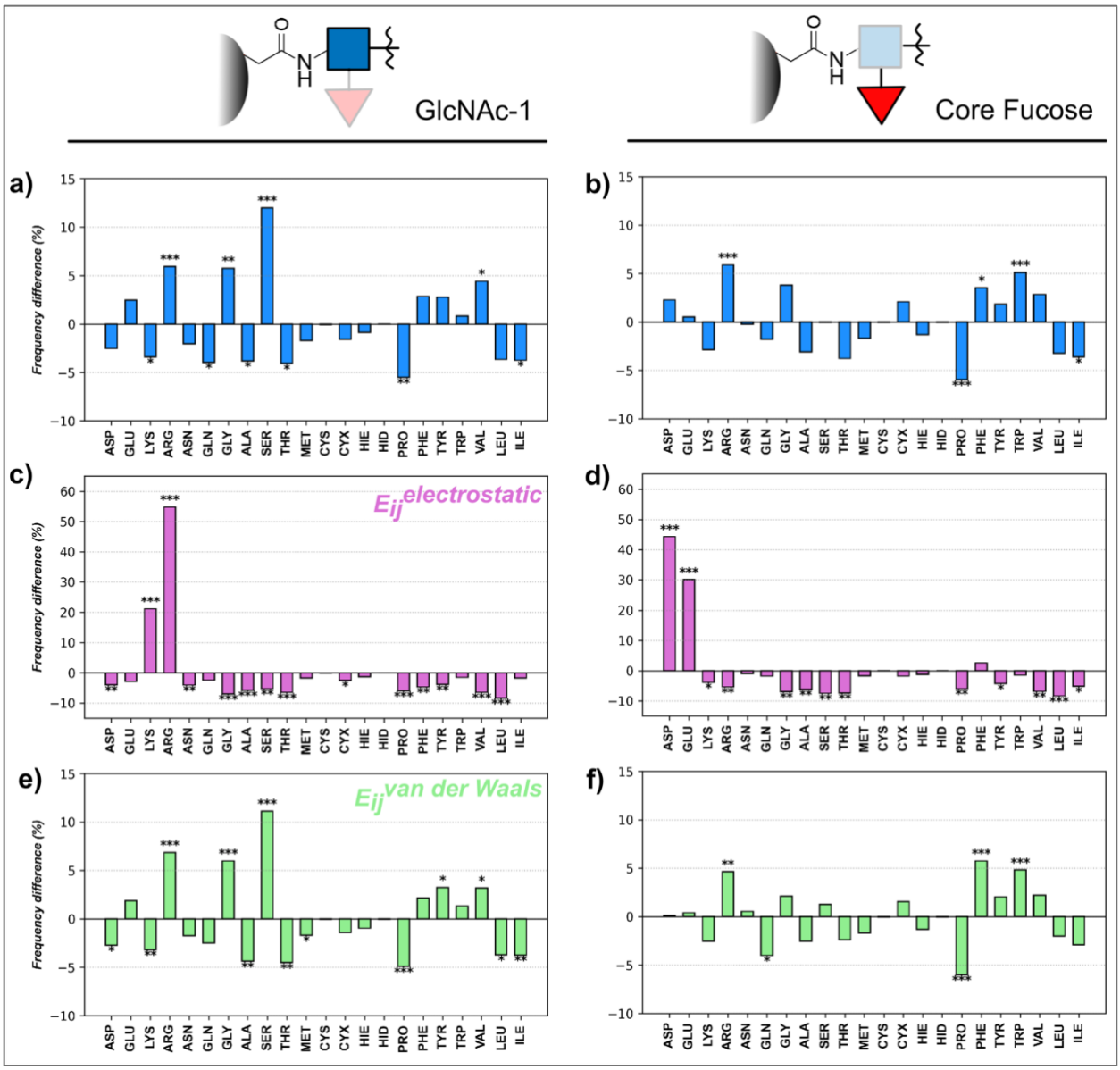


Figure 5: **Amino acid enrichment of the protein residues involved in energetically favorable interactions with GlcNAc-1 and fucose of core-fucosylated glycans, decomposed by energy term. a-b)** Amino acid enrichment among surface protein residues in the dataset of glycoproteins with core-fucosylated glycans involved in favorable interactions with GlcNAc-1 (a) or core fucose (b), where interaction favorability was determined based on E_{ij}^{total}. The differences in amino acid frequencies were calculated by comparing the distribution of each amino acid among surface residues with $E_{ij}^{total} \leq -0.817$ kcal·mol^{-1} to the distribution among residues with $E_{ij}^{total} > -0.817$

kcal·mol^{-1} with GlcNAc-1 or core fucose. **c-d)** Amino acid enrichment among surface protein residues involved in favorable gas-phase electrostatic interactions with GlcNAc-1 (c) or core fucose (d). The differences in amino acid frequencies were calculated by comparing the distribution of each amino acid among surface residues with $E_{ij}^{electrostatic} \leq -1.67$ kcal·mol^{-1} to the distribution among residues with $E_{ij}^{electrostatic} > -1.67$ kcal·mol^{-1} with GlcNAc-1 or core fucose. **d-e)** Amino acid enrichment among surface protein residues involved in favorable gas-phase van der Waals interactions with GlcNAc-1 (c) or core fucose (d). The differences in amino acid frequencies were calculated by comparing the distribution of each amino acid among surface residues with $E_{ij}^{van\ der\ Waals} \leq -0.206$ kcal·mol^{-1} to the distribution among residues with $E_{ij}^{van\ der\ Waals} > -.206$ kcal·mol^{-1} with GlcNAc-1 or core fucose. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In the dataset of non-core fucosylated glycans, van der Waals interactions dominated the total interaction energy with GlcNAc-1, as favorable gas phase electrostatic interactions were offset by the energetic penalty of desolvating the residues to form the glycan—protein contacts. To assess whether this occurred in the interaction energy calculations of core-fucosylated glycans, we analyzed the amino acid enrichment of protein—GlcNAc-1 and protein—core fucose interactions based on the gas-phase electrostatic and van der Waals components of the MM/GBSA calculations (Figures 6c-f). Lysine and arginine were significantly enriched among residues that formed favorable electrostatic interactions with GlcNAc-1, while aspartate and glutamate were significantly enriched in electrostatic interactions with core fucose residue (Figures 5c-d). However, these charged residues, with the exception of arginine, were not enriched when binding favorability was assessed using the total interaction energy calculated by MM/GBSA (Figure 5a-

b), indicating that the favorable electrostatic interactions observed in the gas phase were offset by desolvation. Moreover, arginine was enriched among protein residues with favorable van der Waals interactions with GlcNAc-1 (Figure 5e) and core fucose (Figure 5f) in the gas phase, indicating that the enrichment of arginine observed in the total interaction energy likely arose from favorable nonpolar interactions, not electrostatic interactions. Counts of each amino acid used in the enrichment analyses of protein residues involved in favorable interactions with core-fucosylated GlcNAc-1 and core fucose are reported in Table S4 and in Table S5, respectively.

CHAPTER 4

DISCUSSION

MM/GBSA calculations were employed to quantify the intramolecular N-glycan—protein interactions for 199 non-core-fucosylated and 62 core-fucosylated glycans found in glycoprotein structures in the RCSB PDB. MM/GBSA measures the enthalpic component (ΔH) of the binding energy between two molecules and is traditionally combined with an estimate of the entropic contribution (ΔS) to calculate the Gibbs free energy of binding (ΔG) between a ligand and receptor (Genheden & Ryde, 2015; Gohlke & Case, 2004; Hadden et al., 2015; Wang et al., 2017). In this study, the entropy of the N-linked glycans in the bound and unbound states was not included in the interaction energy calculations; therefore, the impact of entropy on the favorability of the interactions between the N-linked glycans and the protein surface was not evaluated. For this reason, the magnitude of the reported interaction energies does not reflect the binding free energy between N-linked glycans and their underlying protein surfaces, because the entropic penalty associated with restricting glycan flexibility was not accounted for (Gao et al., 2010). Instead, the interaction energies calculated by MM/GBSA were used to compare the relative strength of the protein—glycan interactions.

Analysis of the interaction energies between non-core-fucosylated glycans and their corresponding protein surface revealed that the strongest interactions with the protein involved the chitobiose core (Figure 3b). This observation builds on previous analyses of glycoprotein crystal structures, which inferred these favorable interactions based on the distances between chitobiose core residues and the adjacent protein (Jo et al., 2013; Petrescu et al., 2004). In a handful of

glycans, distal residues beyond Man-3 formed comparably strong interactions with the protein surface up to -23.4 kcal·mol^{-1} (Figure 3b), suggesting that these glycans were in close proximity to the protein surface. These structures would be interesting cases for future characterization of the protein—glycan interactions that occur with distal glycan residues.

Analysis of the interaction energies between core-fucosylated glycans and the protein surface revealed that the strength of the protein—chitobiose core interactions were attenuated compared to that of non-core-fucosylated glycans (Figure 3d-e). However, comparison of the range of interaction energies observed across GlcNAc-1 and core fucose residues indicated that core fucose can form interactions with the protein surface of similar magnitude to GlcNAc-1 (Figure 3d). Analysis of glycosylation sites in which core fucose formed strong interactions with the protein surface revealed that the C6 methyl group formed strong van der Waals contacts with hydrophobic protein residues, while GlcNAc-1 did not form these extensive contacts, resulting in an interaction energy more than five times less favorable (Figure 3f-g). This observation suggests that core fucosylation increases the potential for favorable contacts between N-linked glycans and the protein surface, which may contribute to the reduced conformational dynamics observed in core-fucosylated glycans compared to non-core-fucosylated glycans (Falconer et al., 2018).

We then characterized the protein residues that formed appreciable interactions with the GlcNAc-1 residues in non-core-fucosylated glycans. Aspartate, threonine, tyrosine, tryptophan, and isoleucine were enriched among these residues relative to their background frequencies on the surface of the glycoproteins in the dataset (Figure 4a). Previous analysis of the amino acids proximal to high-mannose glycans in glycoproteins structures identified aspartate, asparagine, and leucine as being enriched in those regions (Suga et al., 2018). In this study, the authors proposed that interactions with these residues reduced the accessibility of N-glycan processing enzymes.

Although asparagine and leucine were not enriched in this energetic analysis (Figure 4a), the enrichment of aspartate among protein residues forming favorable interactions with GlcNAc-1 is consistent with their observation. Moreover, it has been reported that aromatic residues are often proximal to N-glycosylation sites (Petrescu et al., 2004). This may be explained by the observation that aromatic residues at the N+2 position relative to asparagine increased the efficiency of OST (Murray et al., 2015). In addition to increasing glycosylation occupancy, contact between phenylalanine residues and GlcNAc-1 has been shown to increase protein stability in CD2 (Wyss et al., 1995) and in Pin WW (Culyba et al., 2011), presumably through CH-π interactions between the aromatic rings and the ring hydrogens of GlcNAc-1. In this study, tyrosine and tryptophan, rather than phenylalanine, were enriched forming favorable van der Waals interactions with non-fucosylated GlcNAc-1 residues (Figure 4a,c). Visualization of these tyrosine and tryptophan residues relative to GlcNAc-1 did not reveal a clear preference for interacting with a particular portion of GlcNAc-1 (Figure 6b).

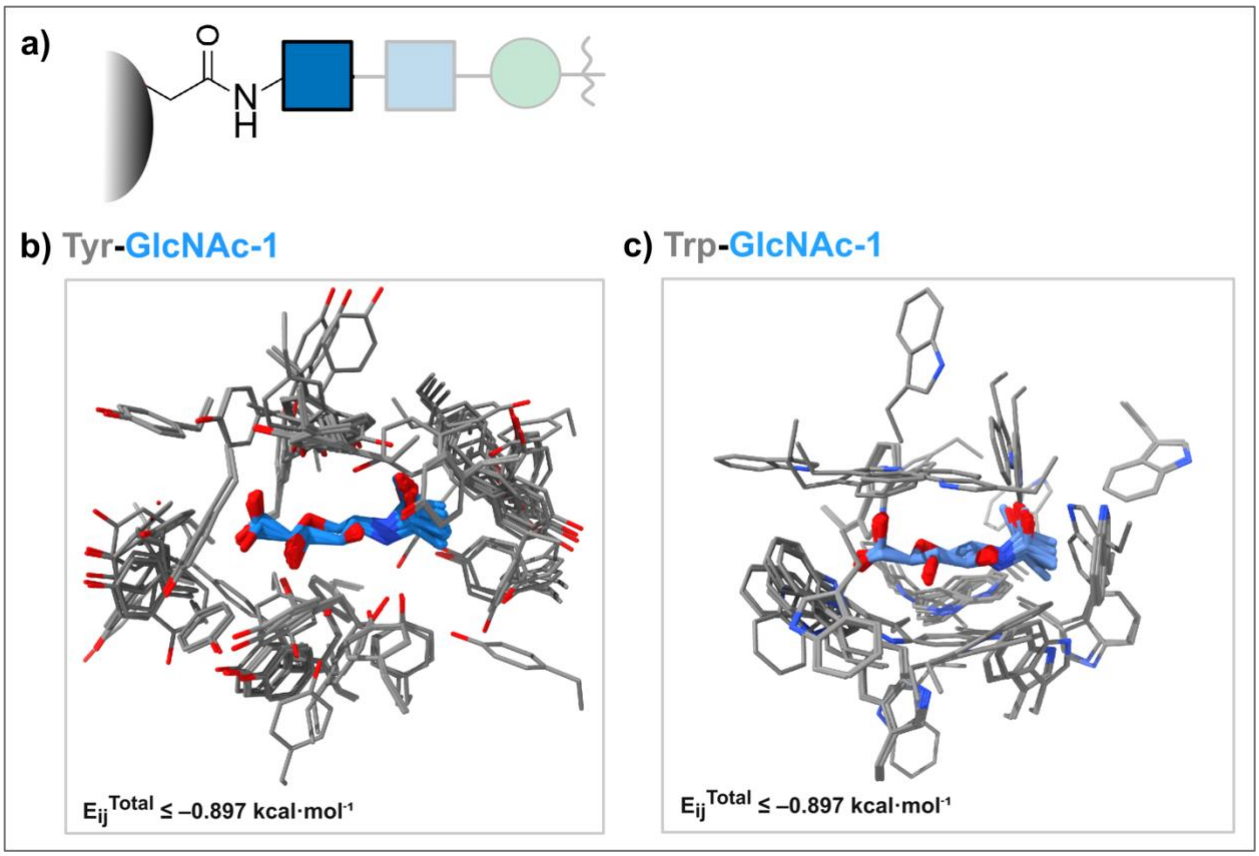


Figure 6. **Visualization of the aromatic protein residues that formed favorable interactions with non-core-fucosylated GlcNAc-1 residues. a)** SNFG representation structure of the N-glycan trisaccharide core, with GlcNAc-1 emphasized to indicate the glycan residue involved in favorable interactions with aromatic residues. **b)** GlcNAc-1—Tyrosine interactions with a total interaction energy (E_{ij}^{total}) ≤ –0.897 kcal·mol^{-1} are shown, with GlcNAc-1 residues aligned to show the diversity of the positions position of the aromatic residues in relation to the glycan. **c)** GlcNAc-1—Tyrosine interactions with a total interaction energy (E_{ij}^{total}) ≤ –0.897 kcal·mol^{-1} are shown, with GlcNAc-1 residues aligned.

Comparison of the amino acid enrichment in protein residues that formed favorable interactions with the GlcNAc-1 and fucose of core-fucosylated glycans revealed that the glycan

residues engaged in distinct interactions with the protein surface. Arginine was significantly enriched in interactions with both GlcNAc-1 and core fucose, whereas glycine, serine, and valine were enriched in GlcNAc-1 interactions (Figure 6a). Notably, aromatic residues phenylalanine and tryptophan were enriched in interactions with both glycan residues, but only significantly in interactions with core fucose (Figure 6b).

Decomposition of the van der Waals and electrostatic components of MM/GBSA revealed that the van der Waals contribution dominated the total interaction energies calculated between the protein surface and both non-core-fucosylated and core-fucosylated glycans, despite highly favorable electrostatic interactions calculated in the gas phase (Figure 4-5). These interactions were offset by the energetic cost of desolvating these residues, which was quantified by the Generalized Born implicit solvent model (Still et al., 1990). It is important to note that the gas-phase electrostatic interactions may be exaggerated due choice of internal dielectric constant. This parameter represents the polarizability of the solute and influences the strength of the electrostatic interactions calculated between two residues (Hou et al., 2011; Li et al., 2018). In this study, an internal dielectric constant of 1 was applied to all glycoprotein systems to compare the interaction energies across systems. This value assumes that the glycoprotein surface is minimally polarizable, which may overemphasize the electrostatic component of the total interaction energy between protein and glycan residues. Therefore, any polarizability at the glycoprotein surface would likely weaken the gas-phase electrostatic energies, which further suggests that favorable protein—glycan interactions are driven by van der Waals interactions and hydrophobic contacts.

CHAPTER 5

CONCLUSIONS

Despite the widespread use of computational modeling to characterize the 3D structure and dynamics of glycoproteins, it is unclear whether the distribution of glycan orientations in the conformational ensemble generated by MD or other methods reflects those preferred in solution. Therefore, it is important to identify features of the protein environment that stabilize the glycan in orientations that are closely associated with the protein surface. In this study, we built upon previous surveys of the protein environment surrounding N-glycosylation sites of experimental glycoprotein structures by employing MM/GBSA, an *in silico* energy calculation method, to quantify interaction energies between N-linked glycans and neighboring protein residues. We found that the core residue of N-linked glycans, including GlcNAc-1 and core fucose, generally contributed the most to protein—glycan interactions compared to distal glycan residues. Additionally, enrichment analyses revealed amino acids that occurred more frequently in favorable contacts with the glycan core than was expected from their background distribution on the protein surface. Comparison of the amino acids enriched among residues involved in favorable interactions with GlcNAc-1 verses core fucose showed that these residues formed distinct interactions with the protein surface. While further studies are needed to determine whether the favorable protein—glycan interactions observed in static crystal structures contribute to stabilizing glycans on the protein surface, these findings provide insight into the interaction patterns that are observed across glycoprotein systems.

CHAPTER 6

REFERENCES

Adams, T. M., Zhao, P., Chapla, D., Moremen, K. W., & Wells, L. (2022). Sequential in vitro enzymatic N-glycoprotein modification reveals site-specific rates of glycoenzyme processing. *J Biol Chem*, *298*(10), 102474. https://doi.org/10.1016/j.jbc.2022.102474

Adelin, A. (2017). Biostatistics: Facing the Interpretation of 2 × 2 Tables. In *Journal of the Belgian Society of Radiology* (Vol. 101): Ubiquity Press.

Aebi, M. (2013). N-linked protein glycosylation in the ER. *Biochim Biophys Acta*, *1833*(11), 2430-2437. https://doi.org/10.1016/j.bbamcr.2013.04.001

An, H. J., Froehlich, J. W., & Lebrilla, C. B. (2009). Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol*, *13*(4), 421-426. https://doi.org/10.1016/j.cbpa.2009.07.022

Bai, L., Wang, T., Zhao, G., Kovach, A., & Li, H. (2018). The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature*, *555*(7696), 328-333. https://doi.org/10.1038/nature25755

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*, *28*(1), 235-242. https://doi.org/10.1093/nar/28.1.235

Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J., & Neumann, S. (1986). X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J*, *5*(10), 2453-2458. https://doi.org/10.1002/j.1460-2075.1986.tb04521.x

Boruah, B. M., Kadirvelraj, R., Liu, L., Ramiah, A., Li, C., Zong, G., Bosman, G. P., Yang, J. Y., Wang, L. X., Boons, G. J., Wood, Z. A., & Moremen, K. W. (2020). Characterizing human alpha-1,6-fucosyltransferase (FUT8) substrate specificity and structural similarities with related fucosyltransferases. *J Biol Chem*, *295*(50), 17027-17045. https://doi.org/10.1074/jbc.RA120.014625

Borza, B., Hajba, L., & Guttman, A. (2020). N-glycan Analysis in Molecular Medicine: Innovator and Biosimilar Protein Therapeutics. *Curr Mol Med*, *20*(10), 828-839. https://doi.org/10.2174/1566524020999201203212352

Case, D. A., Cerutti, D. S., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghazimirsaeed, M., Giambasu, G. M., Giese, T. J., Gotz, A. W., Harris, J. A., Kasavajhala, K., Lee, T. S., Li, Z., Lin, C., Liu, J., Miao, Y., Salomon-Ferrrer, R., Shen, J., Snyder, R.,…Merz, K. M., Jr. (2025). Recent Developments in Amber Biomolecular Simulations. *J Chem Inf Model*, *65*(15), 7835-7843. https://doi.org/10.1021/acs.jcim.5c01063

Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J., & Davis, S. J. (2007). Glycoprotein structural genomics: solving the glycosylation problem. *Structure*, *15*(3), 267-273. https://doi.org/10.1016/j.str.2007.01.011

Cheng, K., Zhou, Y., & Neelamegham, S. (2017). DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology*, *27*(3), 200-205. https://doi.org/10.1093/glycob/cww115

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422-1423. https://doi.org/10.1093/bioinformatics/btp163

Culyba, E. K., Price, J. L., Hanson, S. R., Dhar, A., Wong, C. H., Gruebele, M., Powers, E. T., & Kelly, J. W. (2011). Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science*, *331*(6017), 571-575. https://doi.org/10.1126/science.1198461

Davis, S. J., Crispin, M. (2010). Solutions to the Glycosylation Problem for Low- and High-Throughput Structural Glycoproteomics. In *Functional and Structural Proteomics of Glycoproteins* (pp. 127-158).

de Haan, N., Song, M., Grant, O. C., Ye, Z., Khoder Agha, F., Koed Moller Aasted, M., Woods, R. J., Vakhrushev, S. Y., & Wandall, H. H. (2023). Sensitive and Specific Global Cell Surface N-Glycoproteomics Shows Profound Differences Between Glycosylation Sites and Subcellular Components. *Anal Chem*, *95*(47), 17328-17336. https://doi.org/10.1021/acs.analchem.3c03626

Falconer, D. J., Subedi, G. P., Marcella, A. M., & Barb, A. W. (2018). Antibody Fucosylation Lowers the FcgammaRIIIa/CD16a Affinity by Limiting the Conformations Sampled by the

N162-Glycan. *ACS Chem Biol*, *13*(8), 2179-2189. https://doi.org/10.1021/acschembio.8b00342

Fogarty, C. A., & Fadda, E. (2021). Oligomannose N-Glycans 3D Architecture and Its Response to the FcgammaRIIIa Structural Landscape. *J Phys Chem B*, *125*(10), 2607-2616. https://doi.org/10.1021/acs.jpcb.1c00304

Fontana, C., & Widmalm, G. (2023). Primary Structure of Glycans by NMR Spectroscopy. *Chem Rev*, *123*(3), 1040-1102. https://doi.org/10.1021/acs.chemrev.2c00580

Freeze, H. H., & Kranz, C. (2010). Endoglycosidase and glycoamidase release of N-linked glycans. *Curr Protoc Protein Sci*, *Chapter 12*, Unit12 14. https://doi.org/10.1002/0471140864.ps1204s62

Gao, C., Park, M. S., & Stern, H. A. (2010). Accounting for ligand conformational restriction in calculations of protein-ligand binding affinities. *Biophys J*, *98*(5), 901-910. https://doi.org/10.1016/j.bpj.2009.11.018

Gazaway, E., Kandel, R., Grant, O. C., & Woods, R. J. (2025). Are N-linked glycans intrinsically disordered? *Current Opinion in Structural Biology*, *93*, 103118. https://doi.org/https://doi.org/10.1016/j.sbi.2025.103118

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*, *10*(5), 449-461. https://doi.org/10.1517/17460441.2015.1032936

Gohlke, H., & Case, D. A. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem*, *25*(2), 238-250. https://doi.org/10.1002/jcc.10379

Gordon, R. D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C. A., Vizitiu, D., Withers, S. G., & Rini, J. M. (2006). X-ray crystal structures of rabbit N-acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *J Mol Biol*, *360*(1), 67-79. https://doi.org/10.1016/j.jmb.2006.04.058

Gorelik, A., Randriamihaja, A., Illes, K., & Nagar, B. (2017). A key tyrosine substitution restricts nucleotide hydrolysis by the ectoenzyme NPP5. *FEBS J*, *284*(21), 3718-3726. https://doi.org/10.1111/febs.14266

Gorny, M. K., Sampson, J., Li, H., Jiang, X., Totrov, M., Wang, X. H., Williams, C., O'Neal, T., Volsky, B., Li, L., Cardozo, T., Nyambi, P., Zolla-Pazner, S., & Kong, X. P. (2011). Human anti-V3 HIV-1 monoclonal antibodies encoded by the VH5-51/VL lambda genes define a conserved antigenic structure. *PLoS One*, *6*(12), e27780. https://doi.org/10.1371/journal.pone.0027780

Grant, O. C., Montgomery, D., Ito, K., & Woods, R. J. (2020). Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci Rep*, *10*(1), 14991. https://doi.org/10.1038/s41598-020-71748-7

Gronenborn, A. M., & Clore, G. M. (1985). Investigation of the solution structures of short nucleic acid fragments by means of nuclear overhauser enhancement measurements. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *17*, 1-32. https://doi.org/https://doi.org/10.1016/0079-6565(85)80004-2

Hadden, J. A., Tessier, M. B., Fadda, E., & Woods, R. J. (2015). Calculating Binding Free Energies for Protein-Carbohydrate Complexes. *1273*, 431-466. https://research.ebsco.com/linkprocessor/plink?id=7b4a5b6b-18d0-3300-974a-ea4280d963f5

Hammond, C., Braakman, I., & Helenius, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc Natl Acad Sci U S A*, *91*(3), 913-917. https://doi.org/10.1073/pnas.91.3.913

Hang, I., Lin, C. W., Grant, O. C., Fleurkens, S., Villiger, T. K., Soos, M., Morbidelli, M., Woods, R. J., Gauss, R., & Aebi, M. (2015). Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi. *Glycobiology*, *25*(12), 1335-1349. https://doi.org/10.1093/glycob/cwv058

Hawkins, G. D., Cramer, C. J., & Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry*, *100*(51), 19824-19839. https://doi.org/10.1021/jp961710n

Hirschberg, K., Miller, C. M., Ellenberg, J., Presley, J. F., Siggia, E. D., Phair, R. D., & Lippincott-Schwartz, J. (1998). Kinetic analysis of secretory protein traffic and characterization of golgi to plasma membrane transport intermediates in living cells. *J Cell Biol*, *143*(6), 1485-1503. https://doi.org/10.1083/jcb.143.6.1485

Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on

molecular dynamics simulations. *J Chem Inf Model*, *51*(1), 69-82. https://doi.org/10.1021/ci100275a

Ives, C. M., Singh, O., D'Andrea, S., Fogarty, C. A., Harbison, A. M., Satheesan, A., Tropea, B., & Fadda, E. (2024). Restoring protein glycosylation with GlycoShape. *Nat Methods*, *21*(11), 2117-2127. https://doi.org/10.1038/s41592-024-02464-7

Izadi, S., Anandakrishnan, R., & Onufriev, A. V. (2014). Building Water Models: A Different Approach. *J Phys Chem Lett*, *5*(21), 3863-3871. https://doi.org/10.1021/jz501780a

Jakob, C. A., Burda, P., Roth, J., & Aebi, M. (1998). Degradation of misfolded endoplasmic reticulum glycoproteins in Saccharomyces cerevisiae is determined by a specific oligosaccharide structure. *J Cell Biol*, *142*(5), 1223-1233. https://doi.org/10.1083/jcb.142.5.1223

Jo, S., Lee, H. S., Skolnick, J., & Im, W. (2013). Restricted N-glycan conformational space in the PDB and its implication in glycan structure modeling. *PLoS Comput Biol*, *9*(3), e1002946. https://doi.org/10.1371/journal.pcbi.1002946

Jo, S., Qi, Y., & Im, W. (2016). Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins. *Glycobiology*, *26*(1), 19-29. https://doi.org/10.1093/glycob/cwv083

Kadirvelraj, R., Yang, J. Y., Sanders, J. H., Liu, L., Ramiah, A., Prabhakar, P. K., Boons, G. J., Wood, Z. A., & Moremen, K. W. (2018). Human N-acetylglucosaminyltransferase II substrate recognition uses a modular architecture that includes a convergent exosite. *Proc Natl Acad Sci U S A*, *115*(18), 4637-4642. https://doi.org/10.1073/pnas.1716988115

Kirschner, K. N., & Woods, R. J. (2001). Solvent interactions determine carbohydrate conformation. *Proc Natl Acad Sci U S A*, *98*(19), 10541-10545. https://doi.org/10.1073/pnas.191362798

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem*, *29*(4), 622-655. https://doi.org/10.1002/jcc.20820

Kizuka, Y. (2024). Regulation of intracellular activity of N-glycan branching enzymes in mammals. *J Biol Chem*, *300*(7), 107471. https://doi.org/10.1016/j.jbc.2024.107471

Kornfeld, R., & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem*, *54*, 631-664. https://doi.org/10.1146/annurev.bi.54.070185.003215

Lee, T. S., Cerutti, D. S., Mermelstein, D., Lin, C., LeGrand, S., Giese, T. J., Roitberg, A., Case, D. A., Walker, R. C., & York, D. M. (2018). GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J Chem Inf Model*, *58*(10), 2043-2050. https://doi.org/10.1021/acs.jcim.8b00462

Lees, W. D., Stejskal, L., Moss, D. S., & Shepherd, A. J. (2017). Investigating Substitutions in Antibody-Antigen Complexes Using Molecular Dynamics: A Case Study with Broad-spectrum, Influenza A Antibodies. *Front Immunol*, *8*, 143. https://doi.org/10.3389/fimmu.2017.00143

Lemieux, R. U., Koto, S., & Voisin, D. (1979). The Exo-Anomeric Effect. In *Anomeric Effect* (Vol. 87, pp. 17-29). AMERICAN CHEMICAL SOCIETY. https://doi.org/doi:10.1021/bk-1979-0087.ch002
10.1021/bk-1979-0087.ch002

Lemmin, T., & Soto, C. (2019). Glycosylator: a Python framework for the rapid modeling of glycans. *BMC Bioinformatics*, *20*(1), 513. https://doi.org/10.1186/s12859-019-3097-6

Li, Y., Cong, Y., Feng, G., Zhong, S., Zhang, J. Z. H., Sun, H., & Duan, L. (2018). The impact of interior dielectric constant and entropic change on HIV-1 complex binding free energy prediction. *Struct Dyn*, *5*(6), 064101. https://doi.org/10.1063/1.5058172

Li, Z., Song, L. F., Li, P., & Merz, K. M., Jr. (2020). Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J Chem Theory Comput*, *16*(7), 4429-4442. https://doi.org/10.1021/acs.jctc.0c00194

Liu, K., Tan, S., Jin, W., Guan, J., Wang, Q., Sun, H., Qi, J., Yan, J., Chai, Y., Wang, Z., Deng, C., & Gao, G. F. (2020). N-glycosylation of PD-1 promotes binding of camrelizumab. *EMBO Rep*, *21*(12), e51444. https://doi.org/10.15252/embr.202051444

Murray, A. N., Chen, W., Antonopoulos, A., Hanson, S. R., Wiseman, R. L., Dell, A., Haslam, S. M., Powers, D. L., Powers, E. T., & Kelly, J. W. (2015). Enhanced Aromatic Sequons Increase Oligosaccharyltransferase Glycosylation Efficiency and Glycan Homogeneity. *Chem Biol*, *22*(8), 1052-1062. https://doi.org/10.1016/j.chembiol.2015.06.017

Nagae, M., Morita-Matsumoto, K., Arai, S., Wada, I., Matsumoto, Y., Saito, K., Hashimoto, Y., & Yamaguchi, Y. (2014). Structural change of N-glycan exposes hydrophobic surface of human transferrin. *Glycobiology*, *24*(8), 693-702. https://doi.org/10.1093/glycob/cwu033

Onufriev, A. V., & Case, D. A. (2019). Generalized Born Implicit Solvent Models for Biomolecules. *Annu Rev Biophys*, *48*, 275-296. https://doi.org/10.1146/annurev-biophys-052118-115325

Park, S. J., Lee, J., Qi, Y., Kern, N. R., Lee, H. S., Jo, S., Joung, I., Joo, K., Lee, J., & Im, W. (2019). CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology*, *29*(4), 320-331. https://doi.org/10.1093/glycob/cwz003

Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A., & Wormald, M. R. (2004). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology*, *14*(2), 103-114. https://doi.org/10.1093/glycob/cwh008

Quintana, J. I., Delgado, S., Rabano, M., Azkargorta, M., Florencio-Zabaleta, M., Unione, L., Vivanco, M. D., Elortza, F., Jimenez-Barbero, J., & Arda, A. (2024). The impact of glycosylation on the structure, function, and interactions of CD14. *Glycobiology*. https://doi.org/10.1093/glycob/cwae002

Roe, D. R., & Brooks, B. R. (2020). A protocol for preparing explicitly solvated systems for stable molecular dynamics simulations. *J Chem Phys*, *153*(5), 054123. https://doi.org/10.1063/5.0013849

Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, *9*(7), 3084-3095. https://doi.org/10.1021/ct400341p

Rogals, M. J., Eletsky, A., Huang, C., Morris, L. C., Moremen, K. W., & Prestegard, J. H. (2022). Glycan onformation in the heavily glycosylated protein, CEACAM1. *ACS Chem Biol*, *17*(12), 3527-3534. https://doi.org/10.1021/acschembio.2c00714

Schneider, M., Al-Shareffi, E., & Haltiwanger, R. S. (2017). Biological functions of fucose in mammals. *Glycobiology*, *27*(7), 601-618. https://doi.org/10.1093/glycob/cwx034

Sengupta, A., Li, Z., Song, L. F., Li, P., & Merz, K. M., Jr. (2021). Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J Chem Inf Model*, *61*(2), 869-880. https://doi.org/10.1021/acs.jcim.0c01390

Shenkman, M., Ron, E., Yehuda, R., Benyair, R., Khalaila, I., & Lederkremer, G. Z. (2018). Mannosidase activity of EDEM1 and EDEM2 depends on an unfolded state of their glycoprotein substrates. *Commun Biol*, *1*, 172. https://doi.org/10.1038/s42003-018-0174-8

Shields, R. L., Lai, J., Keck, R., O'Connell, L. Y., Hong, K., Meng, Y. G., Weikert, S. H., & Presta, L. G. (2002). Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J Biol Chem*, *277*(30), 26733-26740. https://doi.org/10.1074/jbc.M202069200

Slynko, V., Schubert, M., Numao, S., Kowarik, M., Aebi, M., & Allain, F. H. (2009). NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J Am Chem Soc*, *131*(3), 1274-1281. https://doi.org/10.1021/ja808682v

Sood, A., Gerlits, O. O., Ji, Y., Bovin, N. V., Coates, L., & Woods, R. J. (2018). Defining the Specificity of Carbohydrate–Protein Interactions by Quantifying Functional Group Contributions. *Journal of Chemical Information and Modeling*, *58*(9), 1889-1901. https://doi.org/10.1021/acs.jcim.8b00120

Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T., & Dermody, T. S. (2014). The sweet spot: defining virus-sialic acid interactions. *Nat Rev Microbiol*, *12*(11), 739-749. https://doi.org/10.1038/nrmicro3346

Stewart-Jones, G. B., Soto, C., Lemmin, T., Chuang, G. Y., Druz, A., Kong, R., Thomas, P. V., Wagh, K., Zhou, T., Behrens, A. J., Bylund, T., Choi, C. W., Davison, J. R., Georgiev, I. S., Joyce, M. G., Kwon, Y. D., Pancera, M., Taft, J., Yang, Y.,…Kwong, P. D. (2016). Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. *Cell*, *165*(4), 813-826. https://doi.org/10.1016/j.cell.2016.04.010

Sticht, H., Pickford, A. R., Potts, J. R., & Campbell, I. D. (1998). Solution structure of the glycosylated second type 2 module of fibronectin. *J Mol Biol*, *276*(1), 177-187. https://doi.org/10.1006/jmbi.1997.1528

Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, *112*(16), 6127-6129. https://research.ebsco.com/linkprocessor/plink?id=5dd75559-0d9c-3aa9-acca-0bc48ca7ff5b

Suga, A., Nagae, M., & Yamaguchi, Y. (2018). Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation. *Glycobiology*, *28*(10), 774-785. https://doi.org/10.1093/glycob/cwy059

Sun, X., Li, Q., Wu, Y., Wang, M., Liu, Y., Qi, J., Vavricka, C. J., & Gao, G. F. (2014). Structure of influenza virus N7: the last piece of the neuraminidase "jigsaw" puzzle. *J Virol*, *88*(16), 9197-9207. https://doi.org/10.1128/JVI.00805-14

Syed, R. S., Reid, S. W., Li, C., Cheetham, J. C., Aoki, K. H., Liu, B., Zhan, H., Osslund, T. D., Chirino, A. J., Zhang, J., Finer-Moore, J., Elliott, S., Sitney, K., Katz, B. A., Matthews, D. J., Wendoloski, J. J., Egrie, J., & Stroud, R. M. (1998). Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature*, *395*(6701), 511-516. https://doi.org/10.1038/26773

Tanji, H., Ohto, U., Shibata, T., Taoka, M., Yamauchi, Y., Isobe, T., Miyake, K., & Shimizu, T. (2015). Toll-like receptor 8 senses degradation products of single-stranded RNA. *Nature Structural & Molecular Biology*, *22*(2), 109-115. https://doi.org/10.1038/nsmb.2943

Thaysen-Andersen, M., & Packer, N. H. (2012). Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology*, *22*(11), 1440-1452. https://doi.org/10.1093/glycob/cws110

Thieker, D. F., Hadden, J. A., Schulten, K., & Woods, R. J. (2016). 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology*, *26*(8), 786-787. https://doi.org/10.1093/glycob/cww076

Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., & Simmerling, C. (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput*, *16*(1), 528-552. https://doi.org/10.1021/acs.jctc.9b00591

Tsai, Y. X., Chang, N. E., Reuter, K., Chang, H. T., Yang, T. J., von Bulow, S., Sehrawat, V., Zerrouki, N., Tuffery, M., Gecht, M., Grothaus, I. L., Colombi Ciacchi, L., Wang, Y. S., Hsu, M. F., Khoo, K. H., Hummer, G., Hsu, S. D., Hanus, C., & Sikora, M. (2024). Rapid simulation of glycoprotein structures by grafting and steric exclusion of glycan conformer libraries. *Cell*, *187*(5), 1296-1311 e1226. https://doi.org/10.1016/j.cell.2024.01.034

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J.,

Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E.,…SciPy, C. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*, *17*(3), 261-272. https://doi.org/10.1038/s41592-019-0686-2

Wang, C., Greene, D., Xiao, L., Qi, R., & Luo, R. (2017). Recent Developments and Applications of the MMPBSA Method. *Front Mol Biosci*, *4*, 87. https://doi.org/10.3389/fmolb.2017.00087

Weis, W., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J., & Wiley, D. C. (1988). Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature*, *333*(6172), 426-431. https://doi.org/10.1038/333426a0

Wiberg, K. B., Murcko, M. A., Laidig, K. E., & MacDougall, P. J. (1990). Origin of the Gauche Effect in substituted ethanes and ethenes. *The Journal of Physical Chemistry*, *94*(18), 6956-6959. https://doi.org/10.1021/j100381a008

Wolfe, S. (1972). Gauche effect. Stereochemical consequences of adjacent electron pairs and polar bonds. *Accounts of Chemical Research*, *5*(3), 102-111. https://doi.org/10.1021/ar50051a003

Wolfe, S., Myung-Hwan, W., & Mitchell, D. J. (1979). On the magnitudes and origins of the "anomeric effects", "exo-anomeric effects", "reverse anomeric effects", and C・X and C・Y bond -lengths in XCH2YH molecules. *Carbohydrate Research*, *69*(1), 1-26. https://doi.org/https://doi.org/10.1016/S0008-6215(00)85747-X

Woods, R. J. (2018). Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem Rev*, *118*(17), 8005-8024. https://doi.org/10.1021/acs.chemrev.8b00032

Woods, R. J., Edge, C. J., & Dwek, R. A. (1994). Protein surface oligosaccharides and protein function. *Nat Struct Biol*, *1*(8), 499-501. https://doi.org/10.1038/nsb0894-499

Woods, R. J., Pathiaseril, A., Wormald, M. R., Edge, C. J., & Dwek, R. A. (1998). The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. *Eur J Biochem*, *258*(2), 372-386. https://doi.org/10.1046/j.1432-1327.1998.2580372.x

Word, J. M., Lovell, S. C., Richardson, J. S., & Richardson, D. C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol*, *285*(4), 1735-1747. https://doi.org/10.1006/jmbi.1998.2401

Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R., Smolyar, A., Reinherz, E. L., & Wagner, G. (1995). Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science*, *269*(5228), 1273-1278. https://doi.org/10.1126/science.7544493

Xiang, Y., Karaveg, K., & Moremen, K. W. (2016). Substrate recognition and catalysis by GH47 alpha-mannosidases involved in Asn-linked glycan maturation in the mammalian secretory pathway. *Proc Natl Acad Sci U S A*, *113*(49), E7890-E7899. https://doi.org/10.1073/pnas.1611213113

Yamaguchi, Y., & Barb, A. W. (2020). A synopsis of recent developments defining how N-glycosylation impacts immunoglobulin G structure and function. *Glycobiology*, *30*(4), 214-225. https://doi.org/10.1093/glycob/cwz068

Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., Aoki, K., Kellman, B. P., Bridger, R., Barouch, D. H., Brindley, M. A., Lewis, N. E., Tiemeyer, M., Chen, B., Woods, R. J., & Wells, L. (2020). Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microbe*, *28*(4), 586-601 e586. https://doi.org/10.1016/j.chom.2020.08.004

Zhou, T., Frabutt, D. A., Moremen, K. W., & Zheng, Y. H. (2015). ERManI (Endoplasmic Reticulum Class I alpha-Mannosidase) Is Required for HIV-1 Envelope Glycoprotein Degradation via Endoplasmic Reticulum-associated Protein Degradation Pathway. *J Biol Chem*, *290*(36), 22184-22192. https://doi.org/10.1074/jbc.M115.675207

Zhu, W., Zhou, Y., Guo, L., & Feng, S. (2024). Biological function of sialic acid and sialylation in human health and disease. *Cell Death Discov*, *10*(1), 415. https://doi.org/10.1038/s41420-024-02180-3

CHAPTER 7

APPENDIX

Supplementary Information

Supplementary Table 1

PDB	n. Glycans analyzed in this study.	PDB	n. Glycans analyzed in this study.	PDB	n. Glycans analyzed in this study.
1AGM	2	3SJF	1	5I6S	1
1APZ	1	3T6Q	2	5IFP	3
1BJI	1	3TI4	1	5MEJ	2
1GYA	1	3TIA	4	5NBS	11
1INX	1	3U7B	5	5VEN	2
1IQQ	1	3UUE	1	5VEO	1
1JND	1	3V9E	2	6B9O	2
1LR5	4	3W3G	5	6C02	2
1NXC	1	3W52	1	6ELC	1
1TG7	4	4DKC	2	6KRL	2
2B8H	4	4GOS	1	6TCV	1
2CIV	1	4IIB	13	6Z30	1
2CL2	1	4JHV	2	6Z7A	1
2HL6	2	4MWJ	1	6Z7B	2
2HRG	1	4N6N	1	7A0Q	1
2IH8	5	4NI3	2	7AQZ	2
2IQ7	1	4NYQ	1	7BYS	1
2Z66	4	4QN4	4	7DE5	1
3CY4	2	4R08	8	7ESK	1
3EQN	2	4R0A	2	7U4F	1
3HN3	4	4ZZT	1	7WH6	2
3LPP	3	5A7M	5	8BHH	4
3OG2	2	5ANN	4	8D0P	1
3PFX	1	5E64	3	8IDP	8
3PXL	2	5FJI	14	8OK4	1
3QVP	1	5GSQ	4	8PCH	1

Table S1: List of PDBs containing non-core-fucosylated glycans that contain at least the βMan1-4βGlcNAc1-4βGlcNAc core. This list was generated by the GlyFinder web-tool available on

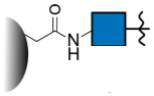
GLYCAM-Web, and the results were filtered to only include PDB structures with at least a 2.5 Å resolution and an average oligosaccharide B-factor of less than or equal to 30.

Supplementary Table 2

PDB	n. Glycans analyzed in this study.	PDB	n. Glycans analyzed in this study.
1AH1	2	5U81	1
1CDR	1	5VAA	2
1CXP	2	5VEO	1
1LK2	1	7JZU	1
1PPF	2	7K43	3
1PPG	2	7KI4	3
1UVQ	1	7LXX	1
2XXL	3	7LXZ	6
2Z7F	2	7LY2	3
3Q76	2	7N6U	2
3UJI	1	7N8H	3
3ZS1	1	7SOC	1
4DL1	8	7TLY	1
4HQ1	1	7UHB	1
4HZW	2	7XCK	1
4X0L	1		

Table S2: List of PDBs containing non-core-fucosylated glycans that contain at least the Fucα1-6-βGlcNAc core. This list was generated by the GlyFinder web-tool available on GLYCAM-Web, and the results were filtered to only include PDB structures with at least a 2.5 Å resolution and an average oligosaccharide B-factor of less than or equal to 30.

Supplementary Table 3

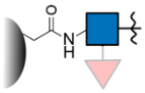


Amino acid	Interaction Count	Total interaction count	Background count	Total background count	χ^2	p-value
a) E_{ij}^{total}						
ALA	36	914	3403	47548	13.6	$2x10^{-4}$
ARG	40		1980		0.05	0.815
ASN	39		3029		6.34	0.012
ASP	77		3015		6.17	0.013
CYS	3		76		0.70	0.403
CYX	5		793		6.28	0.012
GLN	48		2174		0.80	0.372
GLU	41		2635		1.72	0.190
GLY	44		3683		10.4	0.001
HIE	17		1088		0.56	0.455
HID	0		53		0.25	0.614
ILE	55		2116		4.79	0.029
LEU	48		3838		9.29	0.002
LYS	22		1883		5.32	0.021
MET	10		554		0	0.966
PHE	44		1834		1.95	0.162
PRO	38		2597		2.72	0.099
SER	83		3838		1.10	0.295
THR	85		3561		3.97	0.046
TRP	43		933		32.8	$1x10^{-8}$
TYR	76		1805		47.9	$4x10^{-12}$
VAL	60		2660		1.42	0.234
b) $E_{ij}^{electrostatic}$						
ALA	3	336	3436	48126	18.8	$1x10^{-5}$
ARG	32		1988		23.0	$2x10^{-6}$
ASN	7		3061		9.58	0.002
ASP	107		2985		363	0
CYS	0		79		0.00	0.948
CYX	1		797		3.01	0.083
GLN	21		2201		1.78	0.182
GLU	41		2635		27.7	$1x10^{-7}$
GLY	12		3715		7.51	0.006
HIE	7		1098		0	0.953
HID	0		53		0	1
ILE	12		2159		0.46	0.499
LEU	0		3886		28.4	$1x10^{-7}$
LYS	28		1877		16.2	$6x10^{-5}$
MET	0		564		3.03	0.082
PHE	6		1872		3.42	0.064
PRO	0		2635		18.4	$2x10^{-5}$
SER	29		3892		0.07	0.792
THR	6		3640		15.2	$1x10^{-4}$
TRP	4		972		0.78	0.377
TYR	9		1872		1.01	0.315
VAL	11		2709		3.06	0.080

c) $E_{ij}^{van\ der\ Waals}$					
ALA	44		3395	13.5	$2x10^{-4}$
ARG	45		1975	0.01	0.930
ASN	49		3019	4.85	0.028
ASP	70		3022	0.08	0.774
CYS	3		76	0.36	0.546
CYX	6		792	7.06	0.008
GLN	55		2167	0.84	0.358
GLU	37		2639	7.97	0.005
GLY	50		3677	12.8	$4\ x10^{-4}$
HIE	18		1087	1.33	0.248
HID	1	1054	52	0	1
ILE	64		2107	6.01	0.014
LEU	69		3817	2.97	0.085
LYS	22		1883	9.20	0.002
MET	16		548	0.88	0.348
PHE	52		1826	2.96	0.086
PRO	49		2586	1.15	0.284
SER	90		3831	0.23	0.630
THR	102		3544	6.87	0.009
TRP	47		929	31.4	$2\ x10^{-8}$
TYR	92		1789	66.5	0
VAL	73		2647	3.26	0.071

Table S3: a) Appearance of each amino acid type among all surface protein residues that formed favorable interactions with **non-core-fucosylated GlcNAc-1** residues, compared with their appearance in the background population of surface protein residues. Favorable interactions were defined as a total interaction energy (E_{ij}^{total}) with GlcNAc-1 ≤ -0.897 kcal·mol^{-1}. **b-c)** Appearance of each amino acid among surface protein residues that formed favorable gas-phase electrostatic (b) and van der Waals interactions (c) with non-core-fucosylated GlcNAc-1 residues and in the background surface protein residues. Favorable electrostatic interactions were defined as $E_{ij}^{electrostatic}$ ≤ -1.47 kcal·mol^{-1}, and favorable van der Waals interactions were defined as $E_{ij}^{van\ der\ Waals}$ ≤ -0.326 kcal·mol^{-1}.

Supplementary Table 4



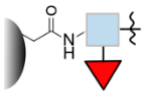
Amino acid	Interaction Count	Total interaction count	Background count	Total background count	χ^2	p-value
a) E_{ij}^{total}						
ALA	5	214	1914	31148	4.72	0.030
ARG	24		1643		13.7	$2x10^{-4}$
ASN	7		1656		1.39	0.239
ASP	5		1514		2.42	0.120
CYS	0		20		0	1
CYX	2		785		1.58	0.208
GLN	3		1670		5.84	0.016
GLU	15		1415		2.43	0.119
GLY	27		2139		10.1	0.002
HIE	1		411		0.62	0.430
HID	0		11		0	1
ILE	3		1605		5.40	0.020
LEU	10		2593		3.26	0.071
LYS	1		1212		5.81	0.016
MET	0		529		2.74	0.098
PHE	16		1442		3.27	0.071
PRO	1		1862		10.6	0.001
SER	43		2528		38.9	0
THR	7		2285		4.60	0.032
TRP	5		472		0.49	0.485
TYR	15		1326		3.29	0.070
VAL	24		2116		5.86	0.016
b) $E_{ij}^{electrostatic}$						
ALA	1	233	1918	31129	12.2	$5x10^{-4}$
ARG	139		1528		1366	0
ASN	3		1660		6.75	0.009
ASP	2		1517		7.24	0.007
CYS	0		20		0	1
CYX	0		787		5.05	0.025
GLN	7		1666		2.08	0.149
GLU	4		1426		3.73	0.054
GLY	0		2166		16.3	$5x10^{-5}$
HIE	0		412		2.19	0.139
HID	0		11		0	1
ILE	8		1600		1.06	0.304
LEU	0		2603		20.2	$7x10^{-6}$
LYS	58		1155		273	0
MET	0		529		3.07	0.080
PHE	0		1458		10.4	0.001
PRO	0		1863		13.8	$2x10^{-4}$
SER	7		2564		7.73	0.005
THR	2		2290		13.5	$2x10^{-4}$
TRP	0		477		2.67	0.102
TYR	1		1340		7.57	0.006
VAL	1		2139		14.1	$2x10^{-4}$

c) $E_{ij}^{van\ der\ Waals}$					
ALA	5		1914	8.55	0.003
ARG	34		1633	24.6	$7x10^{-7}$
ASN	10		1653	1.38	0.239
ASP	6		1513	3.94	0.047
CYS	0		20	0	1
CYX	3		784	1.85	0.174
GLN	8		1665	2.99	0.084
GLU	18		1412	1.81	0.178
GLY	36		2130	14.5	$1x10^{-4}$
HIE	1		411	1.33	0.249
HID	0	281	11	0	1
ILE	4		1604	7.25	0.007
LEU	13		2590	4.55	0.033
LYS	2		1211	6.76	0.009
MET	0		529	3.89	0.049
PHE	19		1439	2.39	0.122
PRO	3		1860	11.2	0.001
SER	54		2517	44.3	0
THR	8		2284	7.68	0.006
TRP	8		469	2.50	0.114
TYR	21		1320	6.32	0.012
VAL	28		2112	3.92	0.048

31081

Table S4: a) Appearance of each amino acid type among all surface protein residues that formed favorable interactions with **core-fucosylated GlcNAc-1** residues, compared with their appearance in the background population of surface protein residues. Favorable interactions were defined as a total interaction energy (E_{ij}^{total}) with GlcNAc-1 ≤ -0.817 kcal·mol^{-1}. **b-c)** Appearance of each amino acid among surface protein residues that formed favorable gas-phase electrostatic (b) and van der Waals interactions (c) with non-core-fucosylated GlcNAc-1 residues and in the background surface protein residues. Favorable electrostatic interactions were defined as $E_{ij}^{electrostatic}$ ≤ -1.67 kcal·mol^{-1}, and favorable van der Waals interactions were defined as $E_{ij}^{van\ der\ Waals}$ ≤ -0.206 kcal·mol^{-1}.

Supplementary Table 5



Amino acid	Interaction Count	Total interaction count	Background count	Total background count	χ^2	p-value
a) E_{ij}^{total}						
ALA	6	197	1913	31165	2.74	0.098
ARG	22		1645		12.3	$4x10^{-4}$
ASN	10		1653		0	1
ASP	14		1505		1.74	0.188
CYS	0		20		0	1
CYX	9		778		2.64	0.104
GLN	7		1666		0.92	0.339
GLU	10		1420		0.03	0.859
GLY	21		2145		3.78	0.052
HIE	0		412		1.72	0.190
HID	0		11		0	1
ILE	3		1605		4.58	0.032
LEU	10		2593		2.30	0.130
LYS	2		1211		3.60	0.058
MET	0		529		2.45	0.117
PHE	16		1442		4.63	0.031
PRO	0		1863		11.5	0.001
SER	16		2555		0	1
THR	7		2285		3.59	0.058
TRP	13		464		30.8	$3x10^{-8}$
TYR	12		1329		1.18	0.277
VAL	19		2121		2.06	0.152
b) $E_{ij}^{electrostatic}$						
ALA	0	139	1919	31223	8.06	0.005
ARG	0		1667		6.81	0.009
ASN	6		1657		0.11	0.741
ASP	68		1451		579	0
CYS	0		20		0	1
CYX	1		786		1.17	0.280
GLN	5		1668		0.52	0.469
GLU	48		1382		281	0
GLY	0		2166		9.31	0.002
HIE	0		412		0.98	0.322
HID	0		11		0.00	1
ILE	0		1608		6.52	0.011
LEU	0		2603		11.6	0.001
LYS	0		1213		4.62	0.032
MET	0		529		1.48	0.223
PHE	10		1448		1.50	0.220
PRO	0		1863		7.78	0.005
SER	1		2570		9.40	0.002
THR	0		2292		9.95	0.002
TRP	0		477		1.26	0.262
TYR	0		1341		5.23	0.022
VAL	0		2140		9.18	0.002

c) $E_{ij}^{van\ der\ Waals}$					
ALA	8	222	1911	31140	2.04	0.153
ARG	22		1645		8.48	0.004
ASN	13		1650		0.05	0.827
ASP	11		1508		0	1
CYS	0		20		0	1
CYX	9		778		1.59	0.207
GLN	3		1670		6.25	0.012
GLU	11		1419		0.01	0.903
GLY	20		2146		1.23	0.268
HIE	0		412		2.04	0.153
HID	0		11		0.00	1
ILE	5		1603		3.23	0.072
LEU	14		2589		0.92	0.338
LYS	3		1210		3.16	0.076
MET	0		529		2.88	0.090
PHE	23		1435		15.2	$1x10^{-4}$
PRO	0		1863		13.1	$3x10^{-4}$
SER	21		2550		0.32	0.572
THR	11		2281		1.49	0.221
TRP	14		463		31.0	$3x10^{-8}$
TYR	14		1327		1.78	0.182
VAL	20		2120		1.35	0.245

Table S5: a) Appearance of each amino acid type among all surface protein residues that formed favorable interactions with **core-fucose** residues, compared with their appearance in the background population of surface protein residues. Favorable interactions were defined as a total interaction energy (E_{ij}^{total}) with GlcNAc-1 ≤ -0.817 kcal·mol^{-1}. **b-c)** Appearance of each amino acid among surface protein residues that formed favorable gas-phase electrostatic (b) and van der Waals interactions (c) with non-core-fucosylated GlcNAc-1 residues and in the background surface protein residues. Favorable electrostatic interactions were defined as $E_{ij}^{electrostatic}$ ≤ -1.67 kcal·mol^{-1}, and favorable van der Waals interactions were defined as $E_{ij}^{van\ der\ Waals}$ ≤ -0.206 kcal·mol^{-1}.