



Comparative ligand structural analytics illustrated on variably glycosylated MUC1 antigen–antibody binding

Christopher B. Barnett^{*1}, Tharindu Senapathi¹ and Kevin J. Naidoo^{*1,2}

Full Research Paper

Open Access

Address:

¹Scientific Computing Research Unit and Department of Chemistry, University of Cape Town, Rondebosch, 7701, South Africa and

²Infectious Disease and Molecular Medicine, Faculty of Health Science, University of Cape Town, Rondebosch, 7701, South Africa

Email:

Christopher B. Barnett* - Chris.Barnett@uct.ac.za; Kevin J. Naidoo* - Kevin.Naidoo@uct.ac.za

* Corresponding author

Keywords:

binding; conformation; Galaxy; glycoprotein; in silico

Beilstein J. Org. Chem. **2020**, *16*, 2540–2550.

<https://doi.org/10.3762/bjoc.16.206>

Received: 09 June 2020

Accepted: 30 September 2020

Published: 13 October 2020

This article is part of the thematic issue "GlycoBioinformatics".

Guest Editors: N. H. Packer, F. Lisacek and N. Karlsson

© 2020 Barnett et al.; licensee Beilstein-Institut.

License and terms: see end of document.

Abstract

When faced with the investigation of the preferential binding of a series of ligands against a known target, the solution is not always evident from single structure analysis. An ensemble of structures generated from computer simulations is valuable; however, visual analysis of the extensive structural data can be overwhelming. Rapid analysis of trajectory data, with tools available in the Galaxy platform, can be used to understand key features and compare differences that inform the preferential ligand structure that favors binding. We illustrate this informatics approach by investigating the in-silico binding of a peptide and glycopeptide epitope of the glycoprotein Mucin 1 (MUC1) binding with the antibody AR20.5. To study the binding, we performed molecular dynamics simulations using OpenMM and then used the Galaxy platform for data analysis. The same analysis tools are applied to each of the simulation trajectories and this process was streamlined by using Galaxy workflows. The conformations of the antigens were analyzed using root-mean-square deviation, end-to-end distance, Ramachandran plots, and hydrogen bonding analysis. Additionally, RMSF and clustering analysis were carried out. These analyses were used to rapidly assess key features of the system, interrogate the dynamic structure of the ligand, and determine the role of glycosylation on the conformational equilibrium. The glycopeptide conformations in solution change relative to the peptide; thus a partially pre-structuring is seen prior to binding. Although the bound conformation of peptide and glycopeptide is similar, the glycopeptide fluctuates less and resides in specific conformers for more extended periods. This structural analysis which gives a high-level view of the features in the system under observation, could be readily applied to other binding problems as part of a general strategy in drug design or mechanistic analysis.

Introduction

A typical sequence of events in research and discovery is noticing a critical biological interaction, searching for structural data, and then searching for the molecular rationale. This is the

connection between biology, chemical biology, and chemistry. The Galaxy project is a popular open web-based platform for accessible, reproducible, and transparent computational

research [1]. Originally built to support bioinformatics, Galaxy now supports a much more expansive community including proteomics [2], metabolomics [3], cheminformatics [4], glycoinformatics [5], and chemistry [6]. Of value to these communities are the broad range of tools and ways to connect tools (workflows) in Galaxy that enable diverse, multidisciplinary research. In this paper, we show how an informatics approach provides a high-level overview, thus enabling rapid observations of changes in molecular details pertinent to the system under investigation. We apply this approach to the binding of glycosylated molecules for the well-known system of mucin binding to the AR20.5 murine antibody.

The binding of glycosylated biomolecules is of increasing interest as glycans are found to be involved in cellular functioning and messaging. The mucins, which are cell surface-associated glycoproteins, are found in mucous secretions and are heavily O-glycosylated [7]. Mucins serve several functions: including protecting the body from pathogens by forming chemical barriers and cellular signaling. Mucin 1 (MUC1) is tethered to the cellular membrane and is found to be aberrantly glycosylated and overexpressed in several epithelial cancers [8]. Further, it is thought to participate in the hyperactivation of selected intracellular signal transduction pathways that promote tumorigenicity [9]. MUC1 is a cancer biomarker that can be detected by serum biomarker assays (such as the CA15-3 test [10,11]). The mode of binding between MUC1 and antibodies has received much attention, and the specificity of this interaction is of interest in improving the performance of these biomarker assays [12,13].

The extracellular domain of MUC1 contains a variable number of tandem repeats (VNTR). The VNTR region is comprised of a repeating sequence of 20 amino acids (–His-Gly-Val-**Thr-Ser**-Ala-Pro-Asp-**Thr**-Arg-Pro-Ala-Pro-Gly-**Ser-Thr**-Ala-Pro-Pro-Ala–)_n, and there are five sites where O-glycosylation may occur (indicated in bold). In cancerous cells, the glycans tend to be truncated or have additional sialylation [14]. For example, in mammary epithelial cells, the mixture of O-glycans that glycosylate mucins are *extended core 2* structures, while in breast cancer cells, O-glycan mass decreases (hypoglycosylation), and there is an increase in abundance of *sialylated core 1* [15]. The upregulation of Tn (α GalNAc) and STn (α NeuAc-2,6- α GalNAc) antigens are commonly associated with cancerous cells [14].

Movahedin et al. confirmed that the glycosylation of MUC1 influences its binding to the AR20.5 murine antibody [16], specifically the Tn-antigen binds more strongly than the nonglycosylated antigen. AR20.5 is known to bind a specific epitope within the MUC1 VNTR domain. Thus, a synthetic 8-amino

acid peptide (APDTRPAP) and the corresponding Tn glycopeptide were synthesized. It was found from the co-crystallization of the AR20.5 antigen-binding fragment (Fab) with the MUC1 peptide and glycopeptide that the glycan moiety of the glycopeptide did not bind to the antibody (Figure 1 and PDB ID:5T6P, 5T78). This is unusual considering that in previous experiments of murine antibody SM3 that Brooks [17] found the glycan forms part of the epitope and binds directly to the antibody. Movahedin et al. hypothesized that the glycan modulates the conformation of the peptide portion of the antigen and does not bind directly.

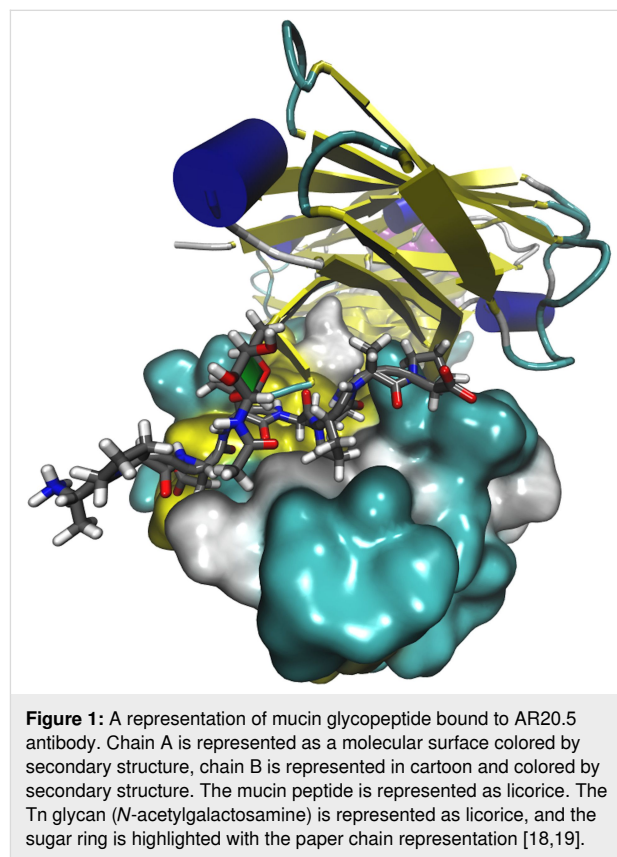


Figure 1: A representation of mucin glycopeptide bound to AR20.5 antibody. Chain A is represented as a molecular surface colored by secondary structure, chain B is represented in cartoon and colored by secondary structure. The mucin peptide is represented as licorice. The Tn glycan (*N*-acetylgalactosamine) is represented as licorice, and the sugar ring is highlighted with the paper chain representation [18,19].

Previous studies have shown that O-glycosylation may provide increased physical stability [20], rigid conformations for protein stability [21], induce the formation of stiff and extended peptide conformations [22], and may affect peptide conformations near the glycosylation site and at distant sites [23]. In glycopeptide enkephalin analogs, the only observed conformational effects due to O-glycosylation were on the residue of attachment and its neighboring residue [24]. While for prion peptides, the O-glycosylation (α -GalNAc) is able to affect the structural transition and suppresses the formation of amyloid fibril formation [25]. The solution structure of O-glycosylated prion peptide was not shifted significantly, with only minor shifts seen in the vicinity of the glycosylation site. Yet there is a

stabilization of the β -structure relative to the random coil and the effects of the glycosylation were hypothesized to relate to the conformational properties of the peptides in solution (as opposed to their equilibrium structures in solution) [25].

A comprehensive structural study of the O-glycosylation-induced changes in a mucin octapeptide showed that the peptide conformation depended on the extent of glycosylation. Glycosylation induces small changes in protein structure and shifts it from a random to a more turn-like structure [26]. Kirnasky et al. noted that O-glycosylation slightly affected the conformational equilibrium of the peptide backbone near the glycosylated residue for a 15-residue mucin peptide. The APDTRP fragment resembled an S-shaped bend and a clustering of low-energy conformations revealed structural similarities between glycosylated and nonglycosylated peptides [23].

The work by Movahedin et al. and others [14,16] provides a foundation for further investigation into the binding of glycopeptide antigens to antibodies using computational modeling. Molecular dynamics (MD) simulations and analysis thereof are a well-known ingredient of the in-silico process for mechanistic screening of glycopeptide fragment binding to antibodies. In this work, the peptide only antigen (Ala-Pro-Asp-Thr-Arg-Pro-Ala-Pro, APDTRPAP) and the Tn glycosylated antigen (APDT(Tn)RPAP) are considered in solution and complex with the AR20.5 antibody. The Tn-antigen is of interest as it is often found upregulated in breast cancer [11,13]. We use MD simulations to investigate the conformational behavior of (glyco)peptide antigens bound to the AR20.5 antibody and to investigate the hypothesis that the glycan modulates the conformation of the peptide portion of the antigen. Primarily showcasing a structural analytics approach, we aim to use the tools and workflows available as part of the Galaxy project to analyze MD simulations to find out if the sugar moiety of the Tn-antigen binds directly to the antibody. Further, if the sugar does not bind directly (as found previously), then we will use these analyses to observe how the sugar modulates binding.

Methods

The inputs, simulation scripts, Galaxy workflows (a series of tools and dataset actions that run in sequence), and data for these simulations are available at <https://github.com/chrisbarnettster/bjoc-paper-2020-sm>.

Simulation

There is an increasing number of software available to assist with the building up of glycosylated biomolecular systems. As opposed to manual preparation, there are glycan-specific tools and toolkits such as doGlycans [27], Glycosylator [28], and online platforms such as GLYCAM-WEB [29] and CHARMM-

GUI [30]. In this work, the CHARMM-GUI server [30] which includes several helper tools (PDB Manipulator [31] and Glycan Reader [32,33]), was used to build these systems and generate input files [34] for use with OpenMM.

Five systems were built in CHARMM-GUI based on initial structures from the Protein Data bank (PDB ID:5T6P, 5T78). The assumption was made that the Tn-antigen binds as per the PDB structure, and other modes of binding are not possible. The solvated receptor, solvated antigens (both the nonglycosylated and Tn-antigen), and a solvated complex (with both antigens) were built in 0.15 M KCl aqueous solution at 310.15 K (physiological temperature). Missing amino acid residues were added. Energy minimization and MD (equilibration and production) simulations were performed using OpenMM [35] and the CHARMM36 force field [36] using the OpenCL platform with mixed precision. Equilibration and production dynamics were carried out as per the scripts provided with CHARMM-GUI, except for adjustments to the time step and number of iterations. The calculations were carried out using Nvidia V100 GPUs.

The equilibration step included 5000 steps of minimization followed by 25000 steps of NVT dynamics (constant volume and temperature) with a time step of 0.001 ps. The particle mesh Ewald (PME) method was used. Nonbonded interactions were cut-off using the force-switching method from 10 Å to 12 Å, and hydrogen bonding constraints applied. During equilibration, the protein backbone and side chains were restrained (force constants of 400.0 and 40.0 kJ mol⁻¹ nm⁻² were used, respectively). The production dynamics were simulated using an *NpT* ensemble and using a time step of 0.002 ps. The antigen-antibody complex in solution was run for 210 ns, while the antigen was run for 500 ns. The antibody was run for 100 ns.

Analysis

The majority of the analyses was carried out using Galaxy, the popular open web-based platform for bioinformatics and computational data analysis, which enables the creation of repeatable analysis pipelines (workflows). There are several well-known molecular dynamics analysis packages (MDAnalysis [37], Bio3D [38] and MDTraj [39]) which are available as computational chemistry analysis tools in Galaxy [6], and these were used to analyze the molecular dynamics trajectories.

The root-mean-square deviation (RMSD) is calculated to measure the stability and conformation of a set of selected atoms. The RMSD is a standard measure of the structural distance between coordinate sets that measures the average distance between a group of atoms [40]. The peptide portion of the antigens was selected for analysis. The root-mean-square-fluc-

tuation (RMSF) represents the deviation at a reference position over time and was calculated in order to measure the variability of the carbon backbone (C- α atoms were selected) of the peptide portion of the antigen (Figure 2).

The end-to-end distance (displacement length) was used as a metric to understand the mobility and conformation of the peptide portion of the antigen throughout the simulation. This is defined as the carbon–nitrogen distance between the first and last amino acid residues of the antigen. A time-series analysis provides some insight, while a histogram provides a clearer understanding of the most populated conformations (Figure 3).

A Ramachandran plot [41] is a well-known method for investigating the ϕ – ψ (phi–psi dihedral angle) preferences around protein backbones (Figure 4). All ϕ – ψ angles for the peptide portion of the antigens were measured for each frame of the simulation and aggregated per residue. The glycosidic-linkage dihedral angles of the Tn-antigen (in solution and bound to antibody) were also measured. A standard hydrogen-bonding analysis using MDAnalysis and VMD was carried out with the default angle cut-off and distance cut-off.

A cluster analysis of the peptide portion of the antigen was carried out (Figure 5) using TTClust [42]. The clusters were chosen automatically based on the carbon backbone of the peptide portion of the antigen and clustered using the Ward algorithm.

Results

The antigens were simulated in solution to understand the innate flexibility prior to binding to the antibody, and then also simulated in the complex with AR20.5 MUC1 antibody to understand the effect of glycosylation on antigen conformation during binding. With the rationale that a high-level overview can be used to understand the molecular changes, various analyses were considered: root-mean-square, end-to-end distance, clustering, ϕ – ψ backbone dihedral angles, and hydrogen-bonding interactions. These analyses focused primarily on the antigen as the antibody conformation does not change significantly in the time frame of the simulation. The peptide-only antigen will be referred to as the ‘antigen’ while the Tn-glycosylated antigen will be referred to as the ‘Tn-antigen’.

Root-mean-square-analysis

In solution (unbound), the RMSD (Figure 2) has a broad spread and a similar center for both the antigen and Tn-antigen. It is readily apparent that the glycosylated antigen has a bimodal distribution (secondary peak at 5.7 Å), indicating at least one other interesting conformation. On consideration of the RMSD for the bound antigens, a narrowing in the distributions is noted. Bound

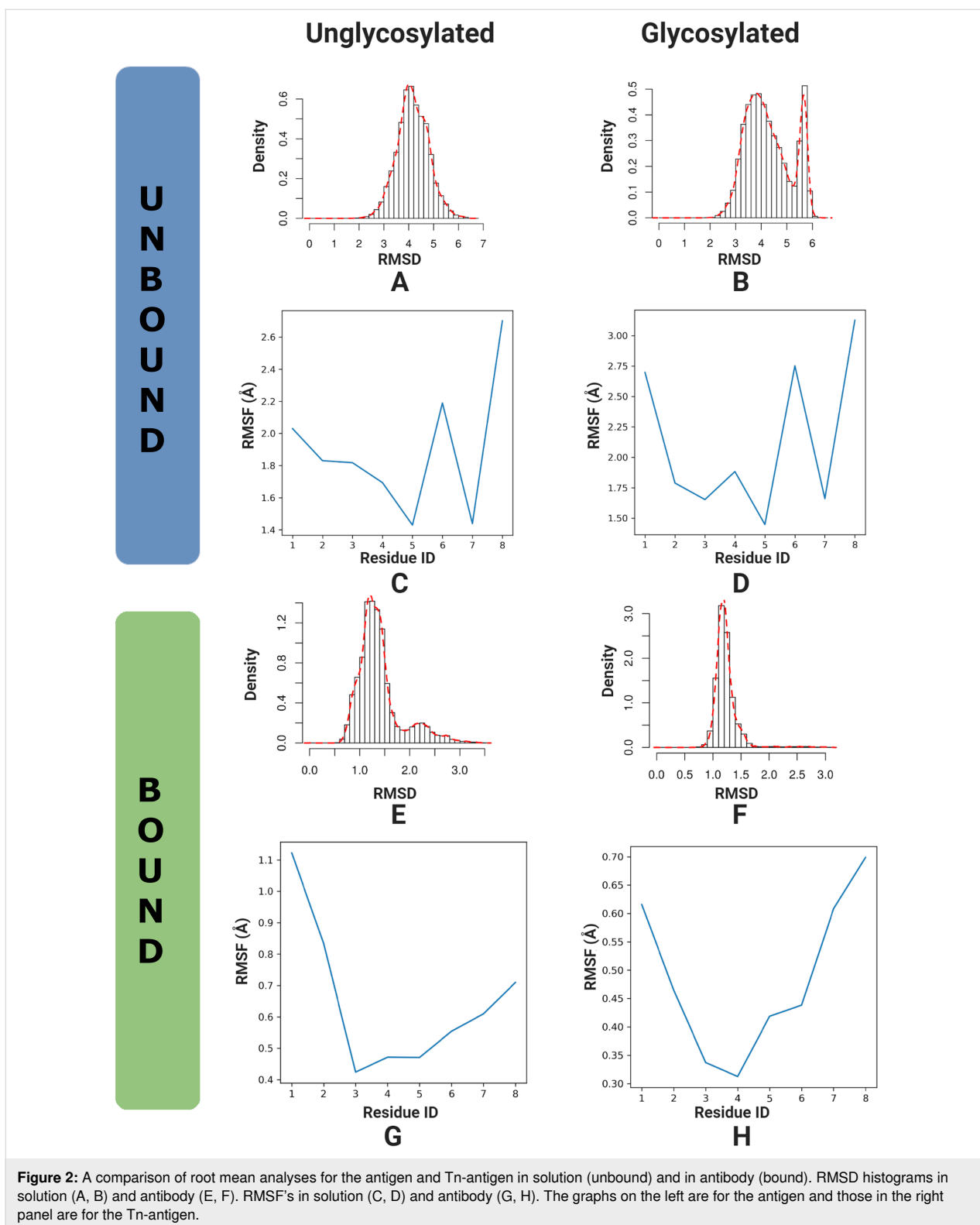
Tn-antigen (Figure 2F) has the narrowest distribution, with a spread from 0.8 Å to 1.6 Å; this unimodal distribution is centered at 1.25 Å. There is no longer a secondary peak, indicating that there is restricted movement on binding. The bound antigen (Figure 2E), instead has a bimodal distribution with a significant population centered at 1.25 Å, a minor population centered at 2.25 Å, and a broad tail that extends to 3.5 Å. While there is restricted movement on binding, the antigen shows unexpected flexibility and a secondary peak at 2.25 Å. From the RMSD, we can infer there is a much tighter range of structures for both antigens when bound than in solution (this should be apparent as there is restricted motion due to the binding of the antigen to the antibody) and the bound Tn-antigen has a more defined and stable conformation.

The RMSFs of the two antigens in solution (Figure 2C and D) have a similar trend with fluctuations ranging between 1.4 Å and 3 Å. Both have large fluctuations, especially for Ala1, Pro6, and Pro8. The Tn-antigen RMSF fluctuates more than the antigen especially for Ala1, Thr4, and Pro8, respectively. When bound, both antigens show restricted fluctuations (Figure 2G and H), with the Tn-antigen showing less fluctuation about each residue. The first and last residues still fluctuate but all RMSF values are less than 1.1 Å indicating relatively minor fluctuations occur for the C- α carbons of the peptide backbone. Another noticeable change is the shift in Pro6, which fluctuated significantly in solution, and now does not. The antigen fluctuates most at the first residue, Ala1, and least at Asp3 and Thr4, while the Tn-antigen fluctuates most for the first and last residues, Ala1 and Pro8, and least for Asp3 and Thr4.

End-to-end analysis

In solution (Figure 3A and B), the displacement lengths of the antigens have a similar range (3.0 Å to 25.0 Å vs. 6.5 Å to 25.0 Å), and both antigens adopt a wide range of conformations with a preference for extended structures. There is a tendency for the Tn-antigen to also prefer a compact conformation, as per the sampling seen at 9.5 Å in the histogram (Figure 3B). The antigen has a left-skewed distribution centered at 19.5 Å, while the Tn-antigen could be bimodal (see the sampling at 9.5 Å and 19.5 Å) or a left-skewed unimodal distribution centered at 19.5 Å.

In contrast, the bound antigens have a much narrower spread (Figure 3C and D). The end-to-end distance for the antigen ranges from 12.5 Å to 22.5 Å, with a distribution centered at 18.9 Å; while the Tn-antigen end-to-end distance ranges from 16.0 Å to 22.0 Å and is centered at 19.5 Å. This is a short peptide so the head and tail regions do fluctuate which could explain the significant spread in the end-to-end distance even



though the antigen is bound to the antibody. Nevertheless, the Tn-antigen shows a slightly narrower spread and a more compact ensemble of structures, but otherwise, the end-to-end distance is very similar for both bound antigens.

Ramachandran analysis

The ϕ - ψ angles of the antigens are considered using a Ramachandran plot. Figure 4 shows a Ramachandran plot for two key amino acids, the glycosylated threonine (Thr4) and

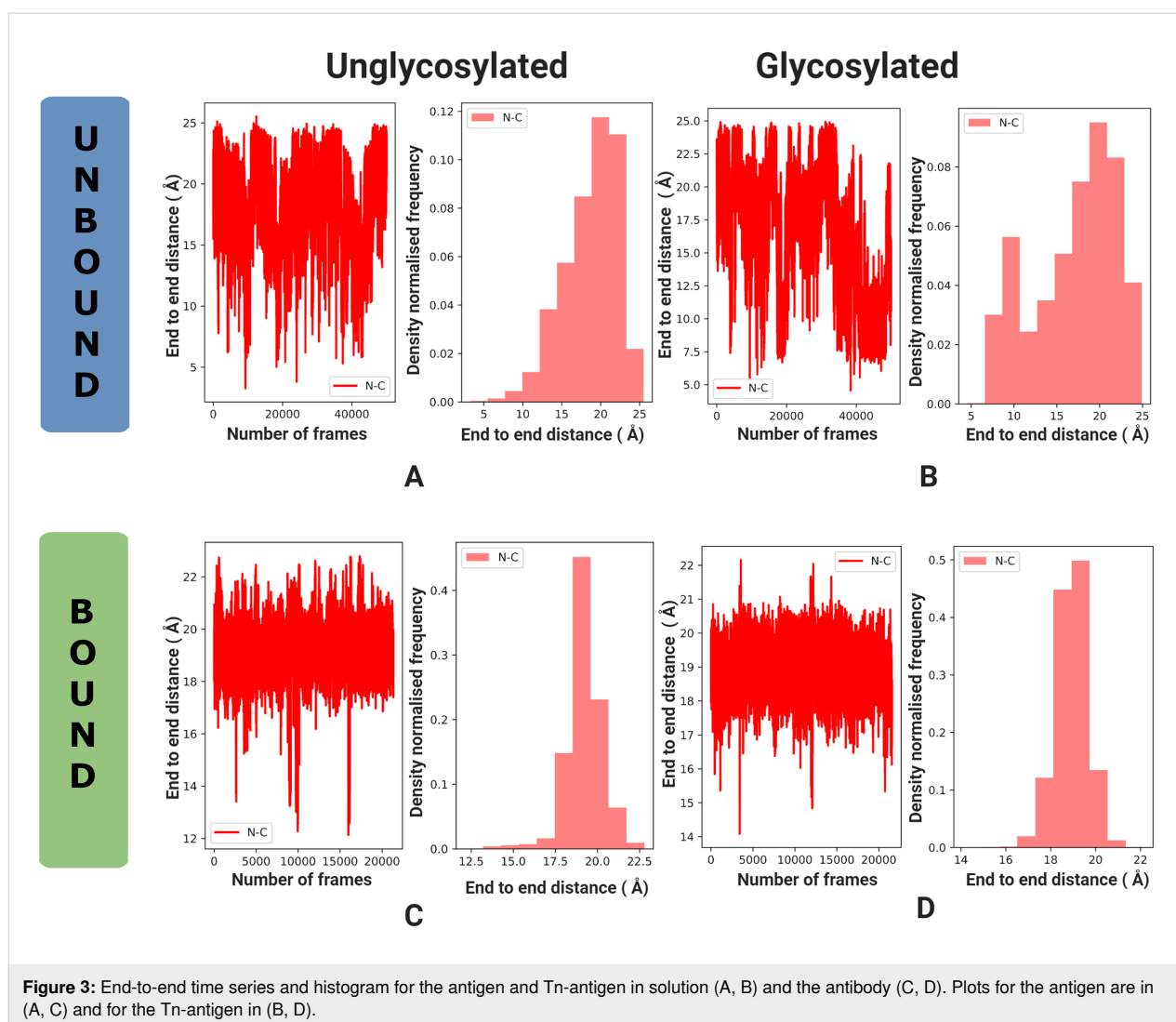


Figure 3: End-to-end time series and histogram for the antigen and Tn-antigen in solution (A, B) and the antibody (C, D). Plots for the antigen are in (A, C) and for the Tn-antigen in (B, D).

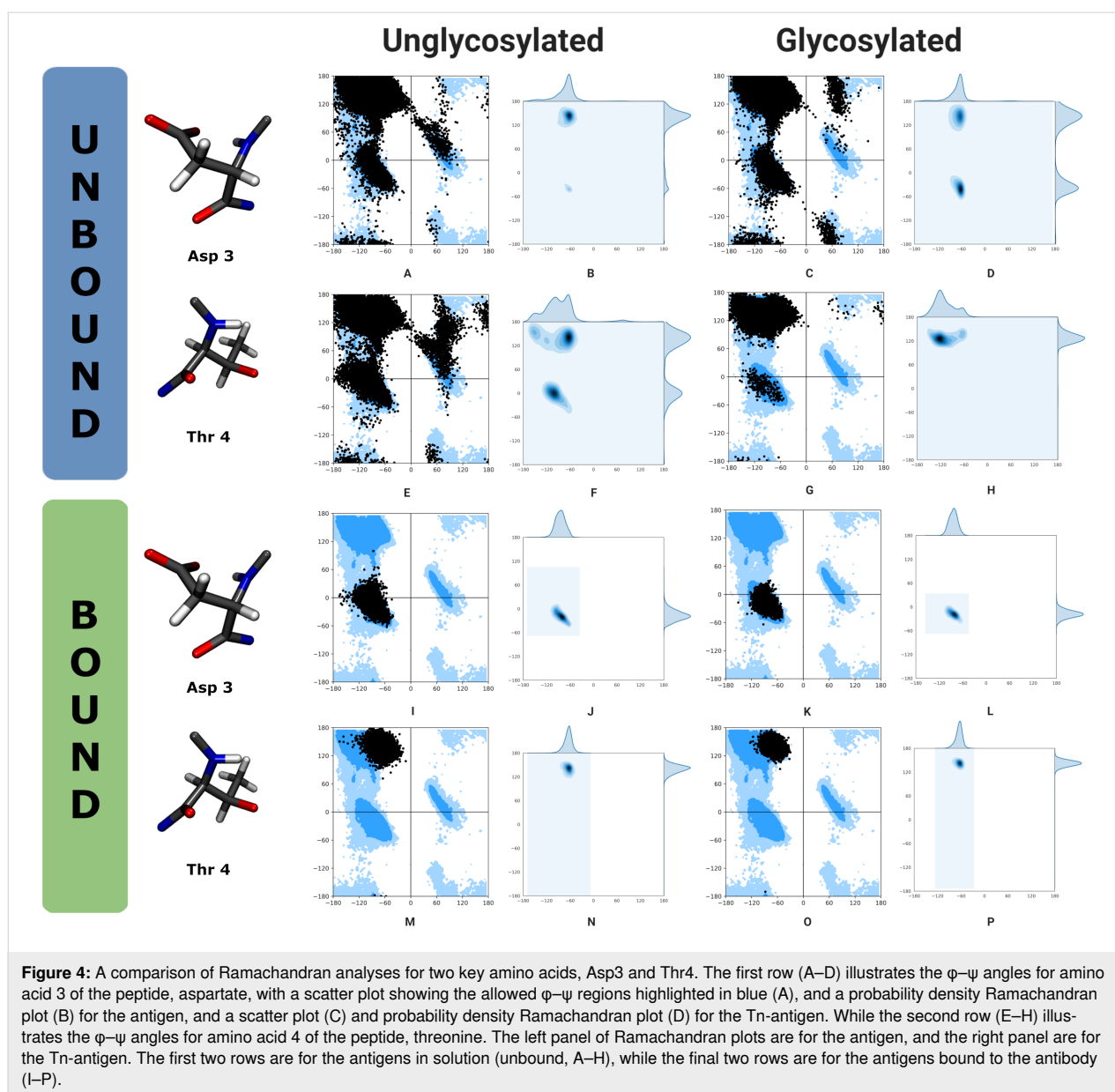
neighboring aspartate (Asp3), and considers the ϕ - ψ angles over all frames of the simulation grouped for these residues. Detailed Ramachandran plots are available (Figures S1 and S2 in Supporting Information File 1) for all residues that can be measured (residues 2–8).

Ramachandran plots show that the ϕ - ψ distribution for the antigens differs in solution but is the same when bound to the antibody. This is a prestructuring effect and is likely an important contributor to the improved binding affinities seen for the Tn-antigen.

In solution, the third residue (Asp3) prefers $(-60^\circ, 135^\circ)$ for the antigen (Figure 4A and B) with some sampling at $(-60^\circ, -40^\circ)$ and minimal sampling at $(60^\circ, 60^\circ)$. When glycosylated, the ψ sampling shifts to become a balanced bimodal distribution (Figure 4C and D) with similar sampling at $(-60^\circ, 135^\circ)$ and $(-60^\circ, -40^\circ)$, and minimal sampling seen at $(60^\circ, 160^\circ)$ and

$(60^\circ, -170^\circ)$. Note that the probability distribution gives the best indication of relevant regions. The fourth residue (Thr4) shows multimodal sampling in ϕ and a bimodal distribution in ψ , with conformers at $(-100^\circ, 0^\circ)$ and $(-60^\circ, 130^\circ)$ being preferred for the antigen (Figure 4E and F). However, when glycosylated the sampling of Thr is restricted (Figure 4G and H), with a strong preference for $(-120^\circ, 120^\circ)$ and the ψ distribution is effectively unimodal.

The antibody prefers that both antigens adopt a particular shape to fit, and this is seen in the ϕ - ψ distributions, which shift for Asp3 and Thr4. When bound, both antigens have an almost identical ϕ - ψ distribution except that the peaks are slightly narrower for the Tn-antigen. In some cases, the preference stays the same and reduced flexibility is observed, for example, Pro2 (Figure S1 and S2 in Supporting Information File 1). In other cases, the conformational preferences shift on binding but this shows no correlation to the effect of glycosylation, for example,



Pro6, Ala7 (Figure S1 and S2 in Supporting Information File 1), and finally, the conformational preference seen for glycosylation in solution aligns with the preference seen for both bound antigens, for example, Asp3 and Thr4 (Figure 4I–P).

For Asp3, the ϕ – ψ preference for both bound antigens is (-60° , -40°), which correlates with the shift seen on glycosylation in solution where the ϕ – ψ preference moved from (-60° , 135°) to sample an additional region of phase space and a combination of conformations at (-60° , -40°) and (-60° , 135°). For Thr4, the ϕ – ψ preference for both bound antigens is (-65° , 140°) which correlates with the shift seen on glycosylation in solution where the ϕ – ψ preference moved from (-100° , 0°) and (-60° , 130°) to (-120° , 120°). The antibody binds both glycosylated

and unglycosylated antigen with the same conformational preference at Asp3 and Thr4 which correlates with the preferred states seen for the glycosylated antigen in solution. There is some evidence of a pre-structuring or pre-organization effect, where O-glycosylation shifts the conformational equilibrium of the peptide towards conformations that are pre-organized for antibody binding.

A Ramachandran plot can be used to understand the role of the sugar moiety, by comparison of the dihedral angle distribution of the glycosidic linkage between the glycan and peptide portion of the Tn-antigen (Figure S3 in Supporting Information File 1). In solution, there is a preference for (70° , 100°) with limited sampling observed in the negative regions of the ψ dis-

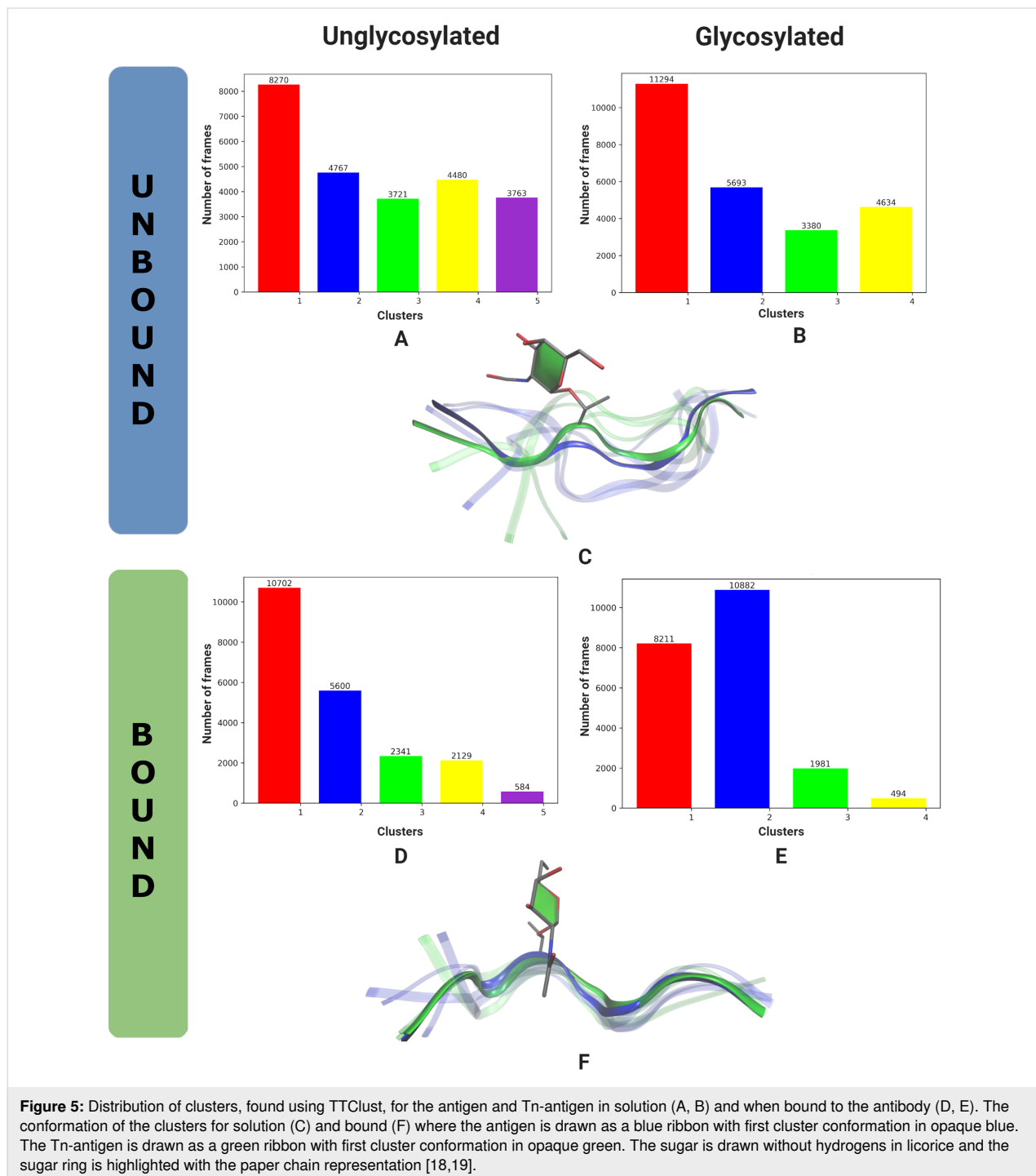
tribution. On binding, this preference is limited and changes slightly to (70°, 120°) with no sampling observed in the negative regions of the ϕ distribution.

Cluster analysis

A cluster analysis of the solution structures yields 5 clusters for the antigen and 4 clusters for the Tn-antigen (Figure 5A and B). The predominant conformer in both antigens is the extended

form (Figure 5C), while for the Tn-antigen, the fourth cluster exhibits a more compact conformation (a transparent green conformer in Figure 5C) as noted in previous analysis.

A cluster analysis of the bound antigens yields 5 clusters for the antigen and 4 clusters for the Tn-antigen (Figure 5D and E). The predominant conformer in both antigens is similar (Figure 5F), as noted in previous analysis. For the antigen, the



first cluster dominates (43%) with the second cluster about half as many members (22%), and the third cluster accounting for 9% of all conformations analyzed. For the Tn-antigen, the first and second clusters dominate accounting for respectively 33% and 44% of all conformations analyzed.

The cluster analysis indicates key conformations of the antigens seen in solution and when bound. In solution, the Tn-antigen can adopt a compact conformation while both antigens adopt extended structures when bound to the antibody. When considering the population count (Figure S5D and E) and residence time of the clusters (Figures S4 and S5 in Supporting Information File 1), the bound Tn-antigen is able to stay resident in the dominant conformation without regularly flipping to other conformations.

Hydrogen bonding

The specifics of intermolecular interactions can also be considered, and here we utilized a hydrogen-bonding analysis to consider how the sugar moiety could interact with the antibody (Tables S1–S7 in Supporting Information File 1).

In solution, hydrogen bonds occur within the antigen between Arg5–Asp3 and Arg5–Pro8 (in order donor–acceptor) with occupancies of 31.83% and 14.32% (and 13.67%). For the Tn-antigen, the peptide portion has hydrogen bonds between Arg5–Pro8 (26.69% and 26.58%), Arg5–Asp3 (12.45%), an Arg5–Pro2 interaction is observed with an occupancy of 7.13%, and an intramolecular hydrogen bond between the C3 alcohol and the carbonyl of the *N*-acetyl moiety of the GalNAc has an occupancy of 6.92%. A shift in hydrogen-bonding populations on glycosylation and the appearance of the Arg5–Pro2 (7.13%) interaction aligns with the compact structure noted previously for the Tn-antigen.

When bound, additional intramolecular hydrogen bonds are observed for the Tn-antigen with interactions between the GalNAc–Thr4 (NH of the acetyl group to carbonyl group) and GalNAc–GalNAc (NH of the acetyl group and the C3 alcohol with the carbonyl of the *N*-acetyl moiety), which occur with occupancies of 23.04% and 29.08%, respectively. These two hydrogen bonds may play a crucial role in maintaining the conformation of the Tn-antigen. There are no intramolecular hydrogen bonds between the peptide moiety of the antigens; these are replaced by hydrogen-bonding between the antigen and chain A of the antibody. The following hydrogen bonds occur between the antigen and antibody: Arg5–Glu39 (141.21%, above 100% as counting both acceptor sites on Arg), Lys58–Asp3 (44.44%), Tyr37–Pro2 (42.55%), Arg55–Asp3 (38.11%), and Tyr54–Asp3 (28.51%). The following hydrogen bonds occur between the Tn-antigen and chain A of the antibody: Arg5–Glu39 (137.49%,

above 100% as counting both acceptor sites on Arg), Lys58–Asp3 (42.80%), Tyr37–Pro2 (46.73%), Arg55–Asp3 (37.77%), and Tyr54–Asp3 (31.44%). A hydrogen bond (0.15%) was observed between the hydroxy group of Tyr100 of chain B of the antibody and the 6-hydroxy group of the GalNAc. While seemingly short-lived, it occurs with some frequency throughout the simulation (see Figure S6 in Supporting Information File 1). Movahedin et al. hypothesized that the glycan modulates the conformation of the peptide portion of the Tn-antigen and does not bind directly, noting that in the crystal structure GalNAc is positioned 4 Å away from the side chain of Tyr100, and indicating that any dispersion interactions would be insufficient to explain a 20-fold increase in affinity. It is unlikely that this hydrogen bond explains a 20-fold increase in affinity yet note that the mobility of the glycan moiety allows the hydrogen-bond interaction to occur. The hydrogen-bonding preferences and occupancies between the antigens and the antibody are very similar.

Discussion

RMSD, RMSF, end-to-end distance, and Ramachandran analyses support that the Tn-antigen has slightly less conformational play than the nonglycosylated antigen when bound to the antibody. The analysis of the ϕ – ψ preference showed that the antibody binds both glycosylated and unglycosylated antigen with the same conformational preference (at Asp3 and Thr4) as that of the glycosylated antigen in solution. There is some evidence of a prestructuring or preorganization effect, where O-glycosylation shifts the conformational equilibrium of the peptide towards conformations that are preorganized for antibody binding. This should decrease the overall entropic penalty upon binding, and therefore would explain an increased binding affinity for the glycosylated antigen.

A cluster analysis showed that the dominant conformation for the bound antigens are similar. Intramolecular hydrogen-bonding interactions within GalNAc were more dominant in the antibody (have a higher occupation) than in solution. An intramolecular hydrogen bond within the Tn-antigen between the GalNAc–Thr4 (NH of the acetyl group to carbonyl group) may be responsible for maintaining the conformation of the Tn-antigen. The role of the sugar in excluding water was not investigated. A short-lived intermolecular hydrogen bond (0.15%) was observed between Tyr100 and GalNAc, and this is unlikely to be significant. These results correlated with the hypothesis put forward previously that glycosylation alters the conformational equilibrium of the antigen.

Conclusion

We have shown how an informatics approach can be used to rapidly obtain key indicators of structural features for under-

standing the molecular level behavior of a system. We illustrated this informatics approach for the binding of glycosylated molecules, in particular for variably glycosylated mucin in solution and when bound to an antibody. RMSD, end-to-end distance, Ramachandran analysis, and hydrogen-bonding analyses were carried out using the Galaxy platform. Additionally, RMSF and cluster analysis were carried out. These analyses were used to gain rapid insight into the behavior of the system. The solution conformations of the Tn-antigen and the antigen were generally extended, yet the Tn-antigen was found to sample a more compact conformation. When bound to the antibody, both antigens had considerably less freedom than when in solution, as expected, and the Tn-antigen had less conformational play. However, this was not the result of hydrogen-bonding interactions between the glycan and the antibody or significantly different interactions between the peptide portion of the Tn-antigen and the antibody. Instead, contributing factors included an intramolecular hydrogen-bonding interaction between GalNAc and Thr4, and a preorganization effect (seen from Ramachandran analysis), where O-glycosylation shifted the conformational equilibrium of the peptide towards conformations that are preorganized for antibody binding. The results agreed with previous findings that glycosylation may affect peptide conformations near the glycosylation site and correlated with the hypothesis that glycosylation alters the conformational equilibrium of the antigen. This structural analysis which gives a high-level view of the features in the system under observation, could be readily applied to other binding problems as part of a general strategy in drug design or mechanistic analysis.

Supporting Information

Supporting Information File 1

Additional molecular dynamics analyses.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-16-206-S1.pdf>]

Acknowledgements

We would like to acknowledge the Galaxy community, the Galaxy Europe team, and the Galaxy computational chemistry team on GitHub. We thank the University of Cape Town eResearch (for support and use of the ilifu data centre) and the Centre for High Performance Computing (for the use of their GPU cluster).

Funding

We thank the University of Cape Town Research Committee and the National Research Foundation of South Africa (Grant Numbers 115215 and 116362) for funding.

ORCID® iDs

Christopher B. Barnett - <https://orcid.org/0000-0002-1467-5741>

Tharindu Senapathi - <https://orcid.org/0000-0002-3277-4022>

Kevin J. Naidoo - <https://orcid.org/0000-0002-9898-3708>

References

- Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Čech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B. A.; Guerler, A.; Hillman-Jackson, J.; Hiltemann, S.; Jalili, V.; Rasche, H.; Soranzo, N.; Goecks, J.; Taylor, J.; Nekrutenko, A.; Blankenberg, D. *Nucleic Acids Res.* **2018**, *46*, W537–W544. doi:10.1093/nar/gky379
- Stewart, P. A.; Kuenzi, B. M.; Mehta, S.; Kumar, P.; Johnson, J. E.; Jagtap, P.; Griffin, T. J.; Haura, E. B. The Galaxy Platform for Reproducible Affinity Proteomic Mass Spectrometry Data Analysis. In *Mass Spectrometry of Proteins: Methods and Protocols*; Evans, C. A.; Wright, P. C.; Noirel, J., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2019; pp 249–261. doi:10.1007/978-1-4939-9232-4_16
- Davidson, R. L.; Weber, R. J. M.; Liu, H.; Sharma-Oates, A.; Viant, M. R. *GigaScience* **2016**, *5*, s13742-016-0115-8. doi:10.1186/s13742-016-0115-8
- Bray, S. A.; Lucas, X.; Kumar, A.; Grüning, B. A. *J. Cheminf.* **2020**, *12*, 40. doi:10.1186/s13321-020-00442-7
- Barnett, C. B.; Aoki-Kinoshita, K. F.; Naidoo, K. J. *Bioinformatics* **2016**, *32*, 3005–3011. doi:10.1093/bioinformatics/btw341
- Senapathi, T.; Bray, S.; Barnett, C. B.; Grüning, B.; Naidoo, K. J. *Bioinformatics* **2019**, *35*, 3508–3509. doi:10.1093/bioinformatics/btz107
- Brockhausen, I.; Schachter, H.; Stanley, P. O-GalNAc Glycans. In *Essentials of Glycobiology*; Varki, A.; Cummings, R. D.; Esko, J. D., Eds.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2009.
- Kufe, D. W. *Oncogene* **2013**, *32*, 1073–1081. doi:10.1038/onc.2012.158
- Nath, S.; Mukherjee, P. *Trends Mol. Med.* **2014**, *20*, 332–342. doi:10.1016/j.molmed.2014.02.007
- Al-azawi, D.; Kelly, G.; Myers, E.; McDermott, E. W.; Hill, A. D. K.; Duffy, M. J.; Higgins, N. O. *BMC Cancer* **2006**, *6*, 220. doi:10.1186/1471-2407-6-220
- Williams, K. A.; Terry, K. L.; Tworoger, S. S.; Vitonis, A. F.; Titus, L. J.; Cramer, D. W. *PLoS One* **2014**, *9*, e88334. doi:10.1371/journal.pone.0088334
- Ricardo, S.; Marcos-Silva, L.; Pereira, D.; Pinto, R.; Almeida, R.; Söderberg, O.; Mandel, U.; Clausen, H.; Felix, A.; Lunet, N.; David, L. *Mol. Oncol.* **2015**, *9*, 503–512. doi:10.1016/j.molonc.2014.10.005
- Teramoto, K.; Ozaki, Y.; Hanaoka, J.; Sawai, S.; Tezuka, N.; Fujino, S.; Daigo, Y.; Kontani, K. *Ther. Adv. Med. Oncol.* **2017**, *9*, 147–157. doi:10.1177/1758834016678375
- Song, W.; Delyria, E. S.; Chen, J.; Huang, W.; Lee, J. S.; Mittendorf, E. A.; Ibrahim, N.; Radvanyi, L. G.; Li, Y.; Lu, H.; Xu, H.; Shi, Y.; Wang, L.-X.; Ross, J. A.; Rodrigues, S. P.; Almeida, I. C.; Yang, X.; Qu, J.; Schocker, N. S.; Michael, K.; Zhou, D. *Int. J. Oncol.* **2012**, *41*, 1977–1984. doi:10.3892/ijco.2012.1645
- Brockhausen, I. *EMBO Rep.* **2006**, *7*, 599–604. doi:10.1038/sj.embor.7400705
- Movahedin, M.; Brooks, T. M.; Supekar, N. T.; Gokanapudi, N.; Boons, G.-J.; Brooks, C. L. *Glycobiology* **2017**, *27*, 677–687. doi:10.1093/glycob/cww131

17. Brooks, C. L.; Schietinger, A.; Borisova, S. N.; Kufer, P.; Okon, M.; Hiram, T.; MacKenzie, C. R.; Wang, L.-X.; Schreiber, H.; Evans, S. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10056–10061. doi:10.1073/pnas.0915176107
18. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38. doi:10.1016/0263-7855(96)00018-5
19. Cross, S.; Kuttel, M. M.; Stone, J. E.; Gain, J. E. *J. Mol. Graphics Modell.* **2009**, *28*, 131–139. doi:10.1016/j.jmgs.2009.04.010
20. Chaffey, P. K.; Guan, X.; Chen, C.; Ruan, Y.; Wang, X.; Tran, A. H.; Koelsch, T. N.; Cui, Q.; Feng, Y.; Tan, Z. *Biochemistry* **2017**, *56*, 2897–2906. doi:10.1021/acs.biochem.7b00195
21. Steen, P. V. d.; Rudd, P. M.; Dwek, R. A.; Opendakker, G. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 151–208. doi:10.1080/10409239891204198
22. Jentoft, N. *Trends Biochem. Sci.* **1990**, *15*, 291–294. doi:10.1016/0968-0004(90)90014-3
23. Kirnarsky, L.; Prakash, O.; Vogen, S. M.; Nomoto, M.; Hollingsworth, M. A.; Sherman, S. *Biochemistry* **2000**, *39*, 12076–12082. doi:10.1021/bi0010120
24. Kriss, C. T.; Lou, B.-S.; Szabó, L. Z.; Mitchell, S. A.; Hruby, V. J.; Polt, R. *Tetrahedron: Asymmetry* **2000**, *11*, 9–25. doi:10.1016/S0957-4166(99)00544-3
25. Chen, P.-Y.; Lin, C.-C.; Chang, Y.-T.; Lin, S.-C.; Chan, S. I. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12633–12638. doi:10.1073/pnas.192137799
26. Wu, W.-g.; Pasternack, L.; Huang, D.-H.; Koeller, K. M.; Lin, C.-C.; Seitz, O.; Wong, C.-H. *J. Am. Chem. Soc.* **1999**, *121*, 2409–2417. doi:10.1021/ja983474v
27. Danne, R.; Poojari, C.; Martinez-Seara, H.; Rissanen, S.; Lolicato, F.; Róg, T.; Vattulainen, I. *J. Chem. Inf. Model.* **2017**, *57*, 2401–2406. doi:10.1021/acs.jcim.7b00237
28. Lemmin, T.; Soto, C. *BMC Bioinf.* **2019**, *20*, 513. doi:10.1186/s12859-019-3097-6
29. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655. doi:10.1002/jcc.20820
30. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008**, *29*, 1859–1865. doi:10.1002/jcc.20945
31. Jo, S.; Cheng, X.; Islam, S. M.; Huang, L.; Rui, H.; Zhu, A.; Lee, H. S.; Qi, Y.; Han, W.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Roux, B.; Im, W. CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues. In *Advances in Protein Chemistry and Structural Biology*; Karabencheva-Christova, T., Ed.; Biomolecular Modelling and Simulations, Vol. 96; Academic Press, 2014; pp 235–265. doi:10.1016/bs.apcsb.2014.06.002
32. Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. *Glycobiology* **2019**, *29*, 320–331. doi:10.1093/glycob/cwz003
33. Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W. *Bioinformatics* **2017**, *33*, 3051–3057. doi:10.1093/bioinformatics/btx358
34. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., III; MacKerell, A. D., Jr.; Klauda, J. B.; Im, W. *J. Chem. Theory Comput.* **2016**, *12*, 405–413. doi:10.1021/acs.jctc.5b00935
35. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLoS Comput. Biol.* **2017**, *13*, e1005659. doi:10.1371/journal.pcbi.1005659
36. Huang, J.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2013**, *34*, 2135–2145. doi:10.1002/jcc.23354
37. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319–2327. doi:10.1002/jcc.21787
38. Skjærven, L.; Yao, X.-Q.; Scarabelli, G.; Grant, B. J. *BMC Bioinf.* **2014**, *15*, 399. doi:10.1186/s12859-014-0399-6
39. McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. *Biophys. J.* **2015**, *109*, 1528–1532. doi:10.1016/j.bpj.2015.08.015
40. Bray, S. A.; Senapathi, T.; Barnett, C. B.; Grüning, B. A. *bioRxiv* **2020**, 2020.05.08.084780. doi:10.1101/2020.05.08.084780
41. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95–99. doi:10.1016/s0022-2836(63)80023-6
42. Tubiana, T.; Carvillat, J.-C.; Boulard, Y.; Bressanelli, S. *J. Chem. Inf. Model.* **2018**, *58*, 2178–2182. doi:10.1021/acs.jcim.8b00512

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>). Please note that the reuse, redistribution and reproduction in particular requires that the authors and source are credited.

The license is subject to the *Beilstein Journal of Organic Chemistry* terms and conditions: (<https://www.beilstein-journals.org/bjoc>)

The definitive version of this article is the electronic one which can be found at: <https://doi.org/10.3762/bjoc.16.206>