

# Journal of Visualized Experiments

## Bioinformatics resources for the study of glycan-mediated protein interactions

--Manuscript Draft--

<b>Article Type:</b>	Invited Results Article - JoVE Produced Video
<b>Manuscript Number:</b>	JoVE63356R2
<b>Full Title:</b>	Bioinformatics resources for the study of glycan-mediated protein interactions
<b>Corresponding Author:</b>	Frederique Lisacek Swiss Institute of Bioinformatics Geneva, Geneva SWITZERLAND
<b>Corresponding Author's Institution:</b>	Swiss Institute of Bioinformatics
<b>Corresponding Author E-Mail:</b>	frederique.lisacek@sib.swiss
<b>Order of Authors:</b>	Frederique Lisacek
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Please specify the section of the submitted manuscript.	Biology
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (\$1400)
Please indicate the <b>city, state/province, and country</b> where this article will be <b>filmed</b> . Please do not use abbreviations.	Geneva, Switzerland
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the <a href="#">Author License Agreement</a>
Please confirm that you have read and agree to the terms and conditions of the video release that applies below:	I agree to the <a href="#">Video Release</a>
Please provide any comments to the journal here.	

**TITLE:**

Bioinformatics Resources for the Study of Glycan-Mediated Protein Interactions

**AUTHORS AND AFFILIATIONS:**

Frédérique Lisacek<sup>1,2,3</sup>

<sup>1</sup>Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland

<sup>2</sup>Computer Science Department, University of Geneva, CH-1227 Geneva, Switzerland

<sup>3</sup>Section of Biology, University of Geneva, CH-1205 Geneva, Switzerland

**Corresponding Author:**

Frédérique Lisacek (frederique.lisacek@sib.swiss)

**SUMMARY:**

This protocol illustrates how to explore, compare, and interpret human protein glycomes with online resources.

**ABSTRACT:**

The Glyco@Expasy initiative was launched as a collection of interdependent databases and tools spanning several aspects of knowledge in glycobiology. In particular, it aims at highlighting interactions between glycoproteins (such as cell surface receptors) and carbohydrate-binding proteins mediated by glycans. Here, major resources of the collection are introduced through two illustrative examples centered on the N-glycome of the human Prostate Specific Antigen (PSA) and the O-glycome of human serum proteins. Through different database queries and with the help of visualization tools, this article shows how to explore and compare content in a continuum to gather and correlate otherwise scattered pieces of information. Collected data are destined to feed more elaborate scenarios of glycan function. Glycoinformatics introduced here is, therefore, proposed as a means to either strengthen, shape, or refute assumptions on the specificity of a protein glycome in a given context.

**INTRODUCTION:**

Glycans, proteins to which they are attached (glycoproteins) and proteins to which they bind (lectins or carbohydrate-binding proteins) are the main molecular actors at the cell surface<sup>1</sup>. Despite this central role in cell-cell communication, large-scale studies, including glycomics, glycoproteomics, or glycan-interactomics data are still scarce compared to their counterpart in genomics and proteomics.

Until recently, methods for characterizing the branching structures of complex carbohydrates while still being conjugated to the carrier protein had not been developed. The biosynthesis of glycoproteins is a non-template-driven process in which the monosaccharide donors, the accepting glycoprotein substrates, and the glycosyltransferases and glycosidases play an interactive role. The resulting glycoproteins can bear complex structures with multiple branching points where each monosaccharide component can be one of the several types present in nature<sup>1</sup>. The non-template-driven process imposes biochemical analysis as the only option for

generating oligosaccharide structural data. The analytical process of glycan structures attached to a native protein is often challenging as it requires sensitive, quantitative, and robust technologies to determine monosaccharide composition, linkages, and branching sequences<sup>2</sup>.

In this context, mass spectrometry (MS) is the most widely used technique in glycomics and glycoproteomics experiments. As time goes, these are carried out in higher throughput settings and data is now accumulating in databases. Glycan structures in various formats<sup>3</sup>, populate GlyTouCan<sup>4</sup>, the universal glycan data repository where each structure is associated with a stable identifier irrespective of the level of precision with which the glycan is defined (e.g., possibly missing linkage type or ambiguous composition). Very similar structures are collected but their minor differences are clearly reported. Glycoproteins are described and curated in GlyConnect<sup>5</sup> and GlyGen<sup>6</sup>, two databases cross-referencing each other. MS data supporting structural pieces of evidence are increasingly stored in GlycoPOST<sup>7</sup>. For a wider coverage of online resources, chapter 52 of the reference manual, Essentials of Glycobiology, is dedicated to glycoinformatics<sup>8</sup>. Interestingly, glycopeptide identification software has proliferated in recent years<sup>9,10</sup> though not to the benefit of reproducibility. The latter concern prompted the leaders of the HUPO GlycoProteomics Initiative (HGI) to set a software challenge in 2019. The MS data obtained from processing complex mixtures of N- and O-glycosylated human serum proteins in CID, ETD, and ETHcD fragmentation modes, were made available to competitors whether software users or developers. The full report on the results of this challenge<sup>11</sup> is only outlined here. To begin with, a spread of identifications was observed. It was mainly interpreted as caused by the diversity of methods implemented in search engines, of their settings, and how outputs were filtered, and peptide “counted”. The experimental design may also have put some software and approaches at a (dis)advantage. Importantly, participants using the same software reported inconsistent results, thereby highlighting serious reproducibility issues. It was concluded by comparing different submissions that some software solutions perform better than others and some search strategies yield better results. This feedback is likely to guide the improvement of automated glycopeptide data analysis methods and will in turn, impact database content.

The expansion of glycoinformatics led to creating web portals that provide information and access to multiple similar or complementing resources. The most recent and up-to-date are described in a chapter of the Comprehensive Glycoscience book series<sup>12</sup> and through cooperation, a solution to data sharing and information exchange is offered in an open access mode. One such portal was developed which was originally called Glycomics@ExpASY<sup>13</sup> and renamed Glyco@Expasy, following the major overhaul of the Expasy platform<sup>14</sup> that has hosted a large collection of tools and databases used across several -omics for decades, the most popular item being UniProt<sup>15</sup>—the universal protein knowledgebase. Glyco@Expasy offers a didactic discovery of the purpose and usage of databases and tools, based on a visual categorization and a display of their interdependencies. The following protocol illustrates procedures to explore glycomics and glycoproteomics data with a selection of resources from this portal that makes the connection between glycoproteomics and glycan-interactomics explicit *via* glycomics. As it is, glycomics experiments produce structures where monosaccharides are fully defined and linkages partially or fully determined, but their protein site attachment is poorly, if at all, characterized. In contrast, glycoproteomics experiments generate precise site attachment information but with

a poor resolution of glycan structures, often limited to monosaccharide compositions. This information is pieced together in the GlyConnect database. Furthermore, search tools in GlyConnect can be used to detect potential glycan ligands which are described along with the proteins recognizing them in UniLectin<sup>16</sup>, linked to GlyConnect *via* glycans. The protocol presented here is divided into two sections to cover questions specific to N-linked and O-linked glycans and glycoproteins.

## PROTOCOL:

NOTE: A device with an Internet connection (larger screen preferred) and an up-to-date Web browser such as Chrome or Firefox is required. Using Safari or Edge may not be as reliable.

### 1. From a protein N-glycome in GlyConnect to a lectin of UniLectin

#### 1.1. Accessing resources from Glyco@Expasy

NOTE: The procedure described here is to access GlyConnect but can be applied to accessing any resource recorded in the platform.

1.1.1. Go to <https://glycoproteome.expasy.org/glycomics-expasy> and consider the bubble chart on the right showing different categories such as **Glycoconjugates** or **Glycan Binding**. In the leftmost menu that reflects the categories in the bubbles, check the **Glycoproteins** box so that the bubble chart on the right immediately zooms in the bubble matching that category.

NOTE: Green bubbles are tools and yellow bubbles are databases. Clicking on either one zooms in again to provide details on the resource. Before doing so, the user may want to understand the dependencies of that resource to others.

1.1.2. To get the information on dependencies, move from the **Resource Thematic Classification** tab to the **Resource Dependency Wheel** tab. Place the mouse on **GlyConnect** in the wheel to check its level of integration with other sources (**Figure 1**).

1.1.3. Go back to the **Resource Thematic Classification** tab to reach the **GlyConnect** bubble as in step 1.1.1 and click on it (**Supplementary Figure 1**) to display the GlyConnect homepage in a new tab that shows the statistics of the content in the latest release of the database.

NOTE: A color scheme detailed in **Table 1** matches the different types of information stored in the database. This color code is valid in all entity pages in GlyConnect and is consistent throughout. The homepage also displays four sections dedicated to focused datasets such as those describing the glycosylation of the Sars-Cov-2 spike protein (COVID-19) or extensively detailing human milk oligosaccharides (HMO). These will not be explored in this protocol.

#### 1.2. Exploring the contextual information of a protein N-glycome

NOTE: All glycan structures in GlyConnect are displayed in three alternative and commonly used formats: (1) Symbol Nomenclature For Glycans (SNFG)<sup>17</sup> (2) IUPAC condensed<sup>18</sup>, and (3) Oxford<sup>19</sup>. In contrast, there is no standard notation to express glycan composition. In GlyConnect, the following code is used: Hex for hexose, HexNAc for N-Acetylhexosamine, dHex for fucose, and NeuAc for sialic acids. For the sake of simplicity, visualizing tools rely on a condensed notation: H for hexose, N for N-Acetylhexosamine, F for fucose, and S for sialic acids. Additionally, small letters designate modifications such as “a” for acetylation, “p” for phosphorylation, and “s” for sulfation, for the most frequent of these so-called substituents.

1.2.1. To view and explore the N-glycome of human Prostate Specific Antigen (PSA), from the GlyConnect homepage, proceed as follows.

NOTE: The glycosylation of human PSA has been studied over the years, especially in the context of prostate cancer. The GlyConnect database stores three references<sup>20–22</sup>, which combine glycomics and glycoproteomics data. Note that the results provided here were obtained with the September 2021 release of GlyConnect. Ulterior use of the database may yield slightly different statistics due to frequent data updates.

1.2.2. Select the **PROTEIN** button to open the protein view of the database. In the protein view page, type **prostate** in the search window. Look for the two entries listed in the output distinguishing two isoforms of PSA with distinct pI values. Click on **790** (Id column) corresponding to the common isoform of PSA.

NOTE: Look for the top multi-colored bar that shows summary information extracted from the published work in the scheme detailed above. Several options for navigation are possible as described below.

1.2.3. On the top multi-colored bar, click on the **SOURCE** button in green to display the sample types from which the published data were processed: **Urine** and **Seminal Fluid**. To browse this information further, click on either of these sample types. The same applies to any item that appears when clicking on a colored button.

1.2.4. To check the health-related content of the database, click on the **DISEASE** button, which contains two items, one of which is **Prostate Cancer** that links to the corresponding dedicated disease page in GlyConnect. The summary for that page shows that three large-scale studies have reported 319 compositions on 1,087 sites found in 308 human proteins.

1.2.5. Click on the **STRUCTURE** button to view the full list of 135 structures associated with PSA from glycomics data. Click on the **COMPOSITION** button for the associated 78 compositions determined by glycoproteomics experiments. Click on any structure or composition to obtain further details.

NOTE: Details such as the list of alternative proteins carrying the particular structure or the list of structures matching the composition can be obtained. PSA is known to have only one N-glycosylation site at Asn-69 (only one item counted for the brown **SITE** button).

1.2.6. To reduce the ambiguity of compositions, click on **SUGGESTED STRUCT** below a selected composition (for example, Hex:6 HexNAc:3 NeuAc:1). A suggestion is made each time the monosaccharide count coincides with that of a structure listed above (**Figure 2**).

NOTE: The Hex:6 HexNAc:3 NeuAc:1 composition generated by a glycoproteomics experiment is matched to four higher resolution structures from the glycomics data. In the case of PSA, there is no site ambiguity to resolve since only Asn-69 is glycosylated.

1.2.7. To fully explore the protein page, view further details on the right side of the page (**Figure 3**).

1.2.7.1. View the default **3QUM** PDB (Protein Data Bank<sup>23</sup>) entry for PSA that is shown with two complex glycans attached to each monomer (**Figure 3**) or the alternative **2ZCK** entry, which is also available because of an attached carbohydrate. The second entry shows a single chain.

NOTE: Both entries are visualized with the 3D LiteMol plugin<sup>24</sup> that displays glycans in SNFG-3D notation adopted in the PDB-RCSB.

1.2.7.2. Click on the corresponding links of other cross-references to explore relevant functional information from major proteomics databases, such as UniProt (**Figure 3**).

### 1.3. Visualizing and correlating the contextual information of a protein N-glycome

NOTE: As seen in the previous section, long lists of structures or compositions can be hard to apprehend as a whole and GlyConnect relies on two different tools to visualize key information, namely, GlyConnect Octopus and GlyConnect Compozitor (the first one expands the summary information captured in colored buttons and the second brings out structural dependencies in terms of a structure/composition being contained in another). As illustrated below, GlyConnect Octopus explores associations between the various entities stored in the database through highlighting multiple or single connections as a reflection of the database content.

1.3.1. Perform a GlyConnect Octopus search to confirm the presence of common structural traits such as hybrid core structures and highly frequent sialic acid-containing structures in the diversity of glycans attached to PSA, as described below.

1.3.2. Go to the Octopus homepage <https://glyconnect.expasy.org/octopus/>. Keep the **N-linked** tab selected by default. Move to the **Cores** subtab and click on the **Hybrid** icon. Move to the **Properties** subtab and click on the **Sialylated** icon. Click on the green **Search** button below.

NOTE: The search results are graphically displayed as relationships between three categories of items. By default, the center list matches the query for compositions, the left collection spans related proteins, and the right one spans related glycans.

1.3.3. In the displayed graph of relationships, hover over **H6N4F1S1** to highlight links to six proteins and three structures. Contrast this by hovering over **H6N4F2S1** that singles out the two isoforms of PSA (both referred to as UniProt ID: KLK3\_HUMAN) and one structure (ID: 10996). Hover over the structure ID to show its SNFG representation and click on it to open the corresponding page (**Supplementary Figure 2**).

1.3.4. Change the nodes of the Octopus to any other topic describing the context of glycosylation. The color code remains the same as the one described earlier (see **Table 1**).

1.3.4.1. Change the **Center Nodes to Tissues** to display 15 options in the middle of the graph, many of which are body fluids. Look for all the associations between proteins and glycans matching the query depending on tissue information. Place the cursor on **Urine** or **Seminal Fluid** in the middle of the graph to view different associations (**Figure 4A,B**).

1.3.4.2. Change the **Center Nodes to Disease** to display 13 options, one of which is **Prostate Cancer**. The only protein associated is PSA (KLK3\_HUMAN) (**Supplementary Figure 3**).

NOTE: A closer look at the PSA N-glycome shown in the protein page singles out the very high frequency of a terminal NeuAc(a?-?)Gal(b?-?)GlcNAc substructure in many cases on structures with two or three antennae. Another Octopus can be generated on that basis as described below.

1.3.5. Click on the **Clear** button to refresh the search. Move to the **Properties** subtab and click on the **Bi-antennary** icon. Move to the **Determinants** subtab and click on the **3-Sialyl-LN (type 2)** icon. Click on the green **Search** button below.

1.3.6. Check the Octopus-retrieved associations with bi-antennary glycans containing a terminal 3-Sialyl-LN (type 2) motif, i.e., NeuAc(a1-3)Gal(b1-4)GlcNAc. Change the **Center Nodes to Tissues** for easier reading and hover over **KLK3\_HUMAN** to directly connect **Seminal Fluid** with PSA common isoform and seven structures (**Supplementary Figure 4**).

NOTE: The second visualization tool, GlyConnect Compozitor, performs the scan of potential relationships between each and every composition in a list thereof (see below). A relationship is defined as differing from only one monosaccharide between two compositions. These identified relationships plotted in a graph expose the (dis)continuity of a glycome.

1.3.7. Use GlyConnect Compozitor to perform the scan of potential relationships between each and every composition in a list thereof, as illustrated below for the case of PSA.

NOTE: GlyConnect Compozitor processes compositions in association with a context. It offers distinct tabs for querying GlyConnect, e.g., **Proteins**, **Sources**, **Cell Lines**, **Diseases** that are self-

explanatory to qualify a context. This is illustrated here with PSA as follows.

1.3.8. Go back to the protein page of PSA: <https://glyconnect.expasy.org/browser/proteins/790>. On the right side of the PSA entry page, click on the Compozitor link. Ensure that the Compozitor search fields are pre-filled with the details of the Id 790 entry in the **Protein** tab (**Protein: Prostate-specific Antigen, Species: Homo sapiens, and Glycan Type: N-linked**).

1.3.9. Click on the **Add to Selection** button to retrieve data from the database and display the graph of connected compositions. Deselect the **Include Virtual Nodes** option. Click on the **Compute Graph** button to display a graph showing a well-connected set of 78 compositions representing the PSA N-glycome, and a bar plot showing the main characteristics of the glycans.

1.3.10. Hover over the purple bar in the bar plot, which locates all sialylated structures in the graph to reveal an observable bias toward sialylated structures.

1.3.11. Remain in the main **Protein** tab and select **Prostate-specific antigen - high Pi isoform (psah)** in the **Protein (name)** field.

NOTE: The **Glycan Type** and **Glycan Site** fields are automatically filled.

1.3.12. Click on the **Add to Selection** button to retrieve data from the database that amounts to 57 compositions. Click on the **Compute Graph** button to generate the superimposed graphs of both isoforms and assess the differences in glycomes of the two PSA isoforms. Hover over node labels to prompt the display of the number of structures corresponding to the compositions/labels (**Figure 5**).

#### 1.4. Glycan-binding information in UniLectin

NOTE: Recall the determinant tested in the Octopus, described as NeuAc(a2-3)Gal(b1-4). By definition, it is an established binding part of a glycan structure and, as such, can be searched in the UniLectin3D database<sup>25</sup>.

1.4.1. Go to <https://www.unilectin.eu/> and click on the **UniLectin3D** button. Alternatively, directly go to the page: <https://www.unilectin.eu/unilectin3D/>. Click on the **Glycan Search** button to open this page: [https://www.unilectin.eu/unilectin3D/glycan\\_search](https://www.unilectin.eu/unilectin3D/glycan_search) (**Supplementary Figure 6**).

1.4.2. Click on the purple diamond representing a sialic acid, which prompts the display of all glycan-binding motifs ending with a sialic acid stored in the database. The top part of that collection of motifs contains the NeuAc(a2-3)Gal(b1-4)GlcNAc motif investigated earlier (**Supplementary Figure 7**).

1.4.3. Click on the **NeuAc(a2-3)Gal(b1-4)GlcNAc** motif to prompt the display of all lectins for which a 3D structure confirming the interaction with NeuAc(a2-3)Gal(b1-4)GlcNAc is known. The



result by default shows lectins in all species. Use the **Search by Field** option to limit the view to human-centric information.

1.4.4. Click on the **Search by Field** option. In the **species** field, type **Homo sapiens**. Click on the **Explore X-ray Structures** button to filter out the original list. Only one entry remains, i.e., the human galectin-8. Click on the **View the 3D Structure and Information** button on the upper-right corner of the listed item to display detailed information of human galectin-8 interacting with NeuAc(a2-3)Gal(b1-4)GlcNAc.

1.4.5. Access the structural information on human galectin-8 displayed on the page with two different viewers.

1.4.5.1. Hold the mouse to turn the molecule around and bring the ligand to the fore with the Litemol software<sup>26</sup> integrated to show the lectin 3D structure. Mouse over one of the listed interactions on the left to update the view on the right and locate where that particular interaction acts in the structure with the PLIP software<sup>27</sup> integrated to detail atomic interactions between the lectin and the ligand (**Figure 6**).

1.4.6. Click on any green button that links to the corresponding entries in UniProt, PDB (European or American sites), and GlyConnect to explore these cross-references.

## 2. Exploring and comparing O-glycomes in GlyConnect

### 2.1. Browsing the HGI challenge high confidence dataset

NOTE: The HGI dataset mentioned in the introduction is stored in the GlyConnect database. It contains 163 N- and 23 O-glycopeptides found in 37 glycoproteins considered as a high confidence list. GlyConnect Compozitor<sup>28</sup> is key to assessing glycome data consistency. Importantly, Compozitor allows for virtual nodes (shown in gray) when only one intermediary step is needed to connect the isolated nodes. In that way, virtual nodes tighten the graph and can be interpreted as structures potentially missed in the experimental results.

2.1.1. Browse the HGI dataset from the GlyConnect homepage by going directly to the reference page of the article: <https://glyconnect.expasy.org/browser/references/2943>.

NOTE: The summary in the colored buttons partially reflects the figures provided in the article. Yet, if only 69 unique peptides are listed, this reflects multiple associations between peptides and sites or structures. In the article, a glycopeptide is defined as a unique combination of a peptide and a composition. In GlyConnect, glycosites are first considered, and they are described as a combination of a peptide with structures. This explains the discrepancy in figures between GlyConnect and the above citation.

2.1.2. Check the high frequency of occurrence of N-linked compositions, such as Hex:5 HexNAc:4 NeuAc:2, identified on 42 sites in 43 peptides as opposed to the frequent uniqueness of most O-linked compositions identified on 1 site in 1 peptide.

2.1.3. Click on the Compozitor link on the right side of the reference entry page to assess the consistency of the dataset. Ensure that the Compozitor tool directly processes the DOI of the reference and fills the search field with **reference=10.1101/2021.03.14.435332** in the **Advanced** tab of the tool. Type **&glycan\_type=O-linked** after the DOI number to narrow down the search to O-linked glycans, so that the query becomes: **reference=10.1101/2021.03.14.435332&glycan\_type=O-linked**

2.1.4. Click on the **Add to Selection** button to retrieve data from the database (there are 20 O-linked compositions). Keep the **Include Virtual Nodes** option selected. Click on the **Compute Graph** button to display the graph of connected compositions. This result highlights several gaps in the expected continuity of glycan biosynthesis with nine virtual nodes required to complete the graph (**Figure 7**).

## 2.2. Comparing with the O-glycome of a selected serum protein in GlyConnect

NOTE: To assess whether the gaps can be filled by data stored in GlyConnect, one O-glycosylated protein out of the 37 listed with the reference was selected. In the dataset, Inter-alpha-trypsin inhibitor heavy chain H4 (Q14624) is reported to be an O-glycosylated on Thr-725.

2.2.1. Go to the **Protein** tab of GlyConnect Compozitor (see step 2.1.3). From the **Protein** list, select **Inter-alpha-trypsin inhibitor heavy chain H4**. Ensure that the **Species** selection is **Homo sapiens** by default. Deselect **N-linked** in the **Glycan Type**. Select only **Thr-725** in the **Site** list by first clicking on the minus sign to the left of **Site** to deselect all the sites, and then selecting only **Thr-725** from the list.

2.2.2. Click on the **Add to Selection** button (note that six compositions are associated with Thr-725). Click on the **Compute Graph** button to display the graph of connected compositions (**Supplementary Figure 8**).

2.2.3. Observe the displayed graph, which shows the 17 unique compositions out of the 20 O-linked compositions of the article dataset in blue and the three unique ones out of six in the database in red. In other words, the overlap between the two sources is present in three compositions that are represented in magenta. Note that a 45° rotation of the graph is generated automatically.

NOTE: The number of virtual nodes is reduced by one. As it turns out, H2N2S1 missing in the 20 O-linked compositions of the article dataset and represented as a virtual node is now filled with an additional composition associated with Thr-725 of Inter-alpha-trypsin inhibitor heavy chain H4 in the database. This simplifies the topology of the graph because two other virtual nodes are rendered useless since they were alternative options for filling the gap between H1N2S1 and

H2N2S2. Yet, a second composition imported from the database would be isolated if not for the creation of two new alternative virtual nodes H2N2F1S1 and H1N2F2S1.

2.2.4. To make sense of the virtual nodes, check whether the corresponding compositions are present in GlyConnect. To do this, click on the **Export** button below the graph. Select **Virtual** only by deselecting all other options. Click on the clipboard icon to copy the selection of 8 compositions.

2.2.5. Paste the selection in the query window of Compozitor's **Custom** tab. Select **O-linked** in the **Glycan Type** field. Set the **Selection Label** in the **Compositions** field to, for example, VN to name the list of 8 compositions. Click on the **Add to Selection** button, and then on the **Compute Graph** button. All virtual nodes are now displayed as green nodes (**Figure 8**).

#### REPRESENTATIVE RESULTS:

The first part of the protocol (section 1) showed how to investigate the specificity or the commonality of N-glycans attached on Asn-69 of the human Prostate Specific Antigen (PSA) using the GlyConnect platform. Tissue-dependent (urine and seminal fluid), as well as isoform-dependent (normal and high pI) variations in glycan expression, were emphasized using two visualizing tools (**Figure 4 and Figure 5**).

First, GlyConnect Octopus, which displays associations between entities stored in the database, provided the opportunity to explore contextual information *via* (1) selecting different entities to be shown in the Octopus and (2) clicking on links to examine related entries. The outcome was distinctive associations depending on the tissue.

Second, GlyConnect Compozitor, originally designed to define/refine a composition file for glycopeptide identification, was used to assess glycan expression in two known PSA isoforms (normal and high pI). The comparison of each isoform glycomes produced a well-connected graph singling out four nodes (compositions), two of which are characteristic of the high pI isoform. Even though the glycome overlap is significant, the glycan property bar chart showed a drop of sialylation from the common to the high pI isoform (**Supplementary Figure 5**).

Furthermore, the exploration of UniLectin3D singles out galectin-8 as a possible reader of the PSA glycome since the latter contains many structures with a NeuAc(a2-3)Gal(b1-4)GlcNAc terminal epitope. This provides a lead to look into and cannot be considered as final evidence. Nonetheless, PSA and galectins are known to play an essential role in prostate cancer<sup>29</sup> and the specific role of Galectin-8 was recently highlighted<sup>30</sup>. The first part of the protocol correlates structural (glycoproteomics) and functional (binding) data to establish a likely scenario for protein-protein interactions mediated by glycans.

In the second part of the protocol (section 2), a high-quality set of O-glycan compositions associated with a particular tissue (human serum) was examined and compared to the GlyConnect database content, thereby offering the option of customizing a glycan composition file for the refined identification of glycopeptides (**Figure 7 and Figure 8**). It could rely on the

minimal set of 20 compositions available from one dataset (HGI challenge results) or be enhanced with 23 to 26 items rationally collected in GlyConnect to strengthen the consistency of the set.

#### FIGURE AND TABLE LEGENDS:

**Table 1: Color scheme associated with each entity of the GlyConnect database and valid throughout.**

**Figure 1: Dependency wheel of Glyco@Expasy instantiated for GlyConnect.**

**Figure 2: Suggested glycan structure for a selected glycan composition.** Suggested glycan structure from a glycomics experiment for a glycan composition of a glycoproteomic experiment targeting the same glycoprotein, here human Prostate Specific Antigen (PSA), as proposed in the GlyConnect page for PSA (ID: 790).

**Figure 3: Side right menu of the GlyConnect page for PSA.** Clickable cross-references to other major databases and display with LiteMol glycan plugin of existing 3D structure in the PDB.

**Figure 4: The output of GlyConnect Octopus showing tissue-dependent associations between proteins and glycans.** The query **Hybrid AND Sialylated** has returned all compositions matching these criteria and each composition links together the associated information about proteins and glycans as recorded in the database. Note that by default **Species** is set to *Homo sapiens* but this option is modifiable. Here, GlyConnect Octopus displays all human proteins (left nodes) carrying hybrid and sialylated glycan structures (right nodes) with the tissues in which they are expressed (center nodes). **(A)** The associations with urine are highlighted showing two proteins: choriogonadotropin (GLHA\_HUMAN) and PSA common isoform (KLK3\_HUMAN) connected to scattered (heterogeneous) glycan structures. **(B)** The associations with seminal fluid are highlighted showing two protein isoforms of PSA (KLK3\_HUMAN) connected to grouped (similar) glycan structures.

**Figure 5: The output of GlyConnect Compozitor showing the superimposed N-glycomes of the two isoforms of PSA.** Compositions in condensed notation label each node. The glycans associated with the common isoform are represented as blue nodes and those of the high pI isoform as red nodes. The overlap between glycomes is shown as magenta nodes. Numbers inside the nodes represent the number of glycan structures matching the labeled composition according to the content of the GlyConnect database regarding PSA. The Compozitor graph shown has been slightly modified from the raw output to disentangle the network which is generated by the D3.js library. This is easy to do as any node can be dragged in the browser window space wherever a user wishes, and the paths can thus be shortened or stretched. User can type a specific composition in the **Zoom On** field in the top-right corner to zoom in and center the graph on the corresponding node.

**Figure 6: Summary entry of the human galectin-8 with NeuAc(a2-3)Gal(b1-4)GlcNAc binding details.** Clicking on the green **View the 3D Structure and Information** button (indicated with a

red ellipse) opens a new page in which a close-up on residue interactions is displayed with the PLIP application (indicated by a red arrow).

**Figure 7: The output of GlyConnect Compozitor showing the O-glycome of the human serum high confidence dataset of the HGI challenge.** Without virtual nodes (see text), the connectivity of that graph is low.

**Figure 8: The output of GlyConnect Compozitor showing the possibility of completion of the O-glycome of the human serum high confidence dataset of the HGI challenge, using the GlyConnect database content.** Accessing the content of the entire GlyConnect database using the Compozitor's **Custom** tab reveals that compositions corresponding to the virtual nodes are mapped with existing defined structures as highlighted in the node labels. The node size represents the number of references stored in the database and reporting the corresponding composition. The numeric label of nodes denotes the number of corresponding structures stored in GlyConnect. Selected compositions appear to have zero to eighteen possible matches in the database. In fact, these nodes are only virtual as a reflection of the content of experimental datasets. It is recommended to refine the information in the graph to test the realism of these additional nodes.

**Supplementary Figure 1: Bubble chart of the Glyco@Expasy homepage.** Zooming in the bubble chart of the Glyco@Expasy homepage to focus on the **glycoprotein** category. Software shown in green bubbles and databases in yellow bubbles. Clicking on any bubble summarizes the purpose of the resource.

**Supplementary Figure 2: Octopus-retrieved associations matching the query depending on composition.** Default GlyConnect Octopus display of human proteins (left nodes) carrying hybrid and sialylated glycan structures (right nodes) with matching compositions (center nodes). Composition H6N4F12S1 appears unique to both PSA isoforms (KLK3\_HUMAN). Clicking on the unique structure ID (10996) opens the corresponding page with details showing that the two isoforms are indeed the only proteins carrying this particular glycan.

**Supplementary Figure 3: Octopus-retrieved associations matching the query depending on the disease.** GlyConnect Octopus display of all human proteins (left nodes) carrying hybrid and sialylated glycan structures (right nodes) with the diseases in which they are expressed (center nodes). The associations with prostate cancer are highlighted showing the common isoform of PSA (KLK3\_HUMAN).

**Supplementary Figure 4: Octopus-retrieved associations matching the query depending on tissue information.** GlyConnect Octopus display of all human proteins (left nodes) carrying bi-antennary glycan structures, including the NeuAc(a1-3)Gal(b1-4)GlcNAc motif (right nodes) with the tissues in which they are expressed (center nodes). The associations with seminal fluid are highlighted showing only the common isoform of PSA (KLK3\_HUMAN) and seven structures.

**Supplementary Figure 5: The output of GlyConnect Compozitor showing the superimposed N-**

**glycomes of the two isoforms of PSA.** Compositions in condensed notation are labeling each node. The glycans associated with the common isoform are represented as blue nodes and those of the high pI isoform as red nodes. The overlap between glycomes is shown as magenta nodes. Numbers inside the nodes represent the number of glycan structures matching the labeled composition according to the content of the GlyConnect database regarding PSA. Mousing over the bar chart of glycan properties shows the correspondence between the frequency and the nodes as orange bubbles. Nearly all the PSA common isoform nodes are covered. This frequency drops in the high pI isoform.

**Supplementary Figure 6: Glycan search interface in UniLectin3D.** Clicking on the sialic acid SNFG symbol (circled in red) launches the search for all ligands that contain NeuAc, stored in UniLectin3D.

**Supplementary Figure 7: Excerpt of the output of the search for all ligands that contain NeuAc.** The NeuAc(a2-3)Gal(b1-4)GlcNAc motif of interest is circled in red.

**Supplementary Figure 8: The output of GlyConnect Compozitor showing the O-glycome of the HGI dataset superimposed with the one in GlyConnect.** The output of GlyConnect Compozitor showing the O-glycome of the human serum high confidence dataset of the HGI challenge in blue superimposed with the O-glycome of one O-glycosylated protein out of the 37 listed with the reference, i.e., inter-alpha-trypsin inhibitor heavy chain H4 with additional information contained in GlyConnect. This enhances the connectivity of the graph.

## **DISCUSSION:**

### **GlyConnect Octopus as a tool for revealing unexpected correlations**

GlyConnect Octopus was originally designed to query the database with a loose definition of glycans. Indeed, the literature often reports the main characteristics of glycans in a glycome such as being fucosylated or sialylated, being made of two or more antennae, etc. Furthermore, glycans whether N- or O-linked are classified in cores, as detailed in the reference manual Essentials of Glycobiology<sup>1</sup>, that are also often cited in published articles. Finally, glycan epitopes such as blood group antigens are yet another property sought in structures and potentially singled out for typing a glycan. In the end, it may be relevant to the search for common or distinct characteristics of a glycome expressed in a specific tissue or selected species. In that sense, the collected information should be used as a source of new assumptions as opposed to unique facts.

### **GlyConnect Compozitor as a tool for shaping a glycan composition set**

Browsing structural information as described in a protein page has limitations because lists tend to obscure the relationships between itemized structures as well as those between compositions. GlyConnect Octopus attends to the former and GlyConnect Compozitor to the latter. A careful look at structures listed in most GlyConnect entries reveals the existence of common substructures. Yet this information is not easy to grasp visually without the help of a dedicated viewer.

The content of the glycan composition file supporting the identification of the glycan moiety as

a key parameter of glycopeptide identification software was established by analyzing the results of the HGI challenge. Most classical proteomics search engines accommodate the selection of glyco-based modifications from a collection that derives from data collected in databases/repositories or the literature. Other glycoproteomics dedicated tools use the knowledge of glycan biosynthesis. In this way, the composition file is theoretically defined as the result of expected enzymatic activity. In the end, there are as many composition files as there are search engines and the overlap between them is highly variable. Nonetheless, learning from past experience in proteomics, especially when posttranslational modifications are accounted for, reveals that the performance of search engines is correlated with limiting the search space<sup>31</sup>. Similar observations are made in glycoproteomics and GlyConnect Compozitor was designed to support educated composition data selection, the importance of which was previously discussed<sup>32</sup>.

The usage of this tool was incompletely illustrated in the protocol especially regarding the **Advanced** tab in which queries that directly launch programmatic access to GlyConnect *via* its API (Application Programming Interface) can be expressed. For example, typing **taxonomy=homo sapiens&glycanType=N-linked&tissue=urine&disease=prostate cancer** in the query window of the **Advanced** tab is equivalent to filling the corresponding fields in the **Source** tab (selecting **Homo sapiens** in **Species**, **Urine** in **Tissue**, and **N-linked** in **Glycan Type**) and the **Disease** tab (selecting **Homo sapiens** in **Species**, **Prostate cancer** in **Disease** and **N-linked** in **Glycan Type**). In other words, it provides in one step a result that would require several selections.

Finally, while the creation of virtual nodes is explained in the protocol, their potential redundancy needs an additional comment. Two concurrent options may be indistinguishable because the simulated action of enzymes in the graph does not account for the chronology of enzyme activities. That is why Compozitor suggests two paths through two virtual nodes to bridge two unconnected nodes corresponding to monosaccharide counts with up to two differences. The inclusion of new data often provides missing links. The user is always free to consider or dismiss virtual nodes, by (un)tickling the **Include Virtual Nodes** box.

#### **Known databases and software limitations**

Overall, as with any navigation on the Web, the protocols described above occasionally lead to a non-existent page, often due to an update of a site or a conflict of updates between two sites. In this case and, in fact, all cases where the navigation is not flowing, the easiest is to send a note to the Expaty helpdesk whose efficiency has significantly contributed to the portal's success in the past 28 years.

The content of GlyConnect is biased as a reflection of the current unbalances in the literature. The majority of publications report N-glycosylation in mammals and the database is richer in human N-glycoproteins. Nonetheless, we have been asked in the past to include lesser common datasets and to remain completely open to receiving advice and suggestions.

Besides, Compozitor is currently limited to the comparison of three composition datasets. A major revision of the **Determinant** subtab in the Octopus is planned. Resources of Glyco@Expaty

need regular updates and some may not be carried out in due course; nonetheless, warnings and/or announcements are published when it so happens.

Partner portals, known as GlyGen (<https://www.glygen.org>) and GlyCosmos (<https://www.glycosmos.org>), provide different options and tools. Ultimately, browsing and searching information on either of the options entails a high level of subjectivity and largely depends on users' habits and concerns. We can only hope that our solution suits a part of the community.

The input of glycoscience is growing in life science projects and studies establishing the role of glycans in health issues are continuously produced. The recent focus on Sars-Cov-2 revealed yet again the importance of glycosylated proteins, especially in structural approaches<sup>33</sup>. Glycoinformatics supports glycoscientists in daily tasks of data analysis and interpretation.

#### ACKNOWLEDGMENTS:

The author warmly acknowledges past and present members of the Proteome Informatics Group involved in developing the resources used in this tutorial, specifically, Julien Mariethoz and Catherine Hayes for GlyConnect, François Bonnardel for UniLectin, Davide Alocci, and Frederic Nikitin for the Octopus, and Thibault Robin for Compozitor and final touch on Octopus.

The development of the glyco@Expasy project is supported by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation (SERI) and is currently complemented by the Swiss National Science Foundation (SNSF: 31003A\_179249). ExPASy is maintained by the Swiss Institute of Bioinformatics and hosted at the Vital-IT Competency Center. The author also acknowledges Anne Imberty for outstanding cooperation on the UniLectin platform jointly supported by ANR PIA Glyco@Alps (ANR-15-IDEX-02), Alliance Campus Rhodanien Co-funds (<http://campusrhodanien.unige-cofunds.ch>) Labex Arcane/CBH-EUR-GS (ANR-17-EURE-0003).

#### DISCLOSURES:

The authors declare no conflicts of interest.

#### REFERENCES:

1. Essentials of Glycobiology. at <<http://www.ncbi.nlm.nih.gov/books/NBK310274/>>. Cold Spring Harbor Laboratory Press. Cold Spring Harbor (NY) (2015).
2. Gray, C. J. et al. Advancing solutions to the carbohydrate sequencing challenge. *Journal of the American Chemical Society*. **141** (37), 14463–14479 (2019).
3. Tsuchiya, S., Yamada, I., Aoki-Kinoshita, K. F. GlycanFormatConverter: a conversion tool for translating the complexities of glycans. *Bioinformatics*. **35** (14), 2434–2440 (2018).
4. Fujita, A. et al. The international glycan repository GlyTouCan version 3.0. *Nucleic Acids Research*. **49** (D1), D1529–D1533 (2021).
5. Alocci, D. et al. GlyConnect: glycoproteomics goes visual, interactive, and analytical. *Journal of Proteome Research*. **18** (2), 664–677 (2019).
6. York, W. S. et al. GlyGen: computational and informatics resources for glycoscience.



655 *Glycobiology*. **30** (2), 72–73 (2020).

656 7. Watanabe, Y., Aoki-Kinoshita, K. F., Ishihama, Y., Okuda, S. GlycoPOST realizes FAIR principles  
657 for glycomics mass spectrometry data. *Nucleic Acids Research*. **49** (D1), D1523–D1528 (2020).

658 8. Campbell, M. P., Aoki-Kinoshita, K. F., Lisacek, F., York, W. S., Packer, N. H. Glycoinformatics.  
659 *Essentials of Glycobiology* at <<http://www.ncbi.nlm.nih.gov/books/NBK453097/>> (2015).

660 9. Cao, W. et al. Recent advances in software tools for more generic and precise intact  
661 glycopeptide analysis. *Molecular & Cellular Proteomics*. **20**, 100060 (2021).

662 10. Mariethoz, J., Hayes, C., Lisacek, F. Glycan compositions with Compozitor to enhance  
663 glycopeptide identification. *Proteomics Data Analysis*. **2361**, 109–127 (2021).

664 11. Kawahara, R. et al. Community evaluation of glycoproteomics informatics solutions reveals  
665 high-performance search strategies of serum glycopeptide analysis. *Nature Methods*. **18**, 1304–  
666 1316 (2021).

667 12. Lisacek, F., Aoki-Kinoshita, K. F., Vora, J. K., Mazumder, R., Tiemeyer, M. Glycoinformatics  
668 resources integrated through the GlySpace Alliance. *Comprehensive Glycoscience*. **1**, 507–521  
669 (2021).

670 13. Mariethoz, J. et al. Glycomics@ExPASy: bridging the gap. *Molecular & Cellular Proteomics*. **17**  
671 (11), 2164–2176 (2018).

672 14. Duvaud, S. et al. Expasy, the swiss bioinformatics resource portal, as designed by its users.  
673 *Nucleic Acids Research*. **49**(W1), W216–W227 (2021)

674 15. The UniProt Consortium et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic*  
675 *Acids Research*. **49** (D1), D480–D489 (2021).

676 16. Bonnardel, F., Perez, S., Lisacek, F., Imberty, A. Structural database for lectins and the  
677 UniLectin web platform. *Lectin Purification and Analysis*. **2132**, 1–14 (2020).

678 17. Neelamegham, S. et al. Updates to the symbol nomenclature for glycans guidelines.  
679 *Glycobiology*. **29** (9), 620–624 (2019).

680 18. Sharon, N. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature  
681 of glycoproteins, glycopeptides and peptidoglycans: JCBN recommendations 1985.  
682 *Glycoconjugate Journal*. **3** (2), 123–133 (1986).

683 19. Harvey, D. J. et al. Proposal for a standard system for drawing structural diagrams of N- and  
684 O-linked carbohydrates and related compounds. *Proteomics*. **9** (15), 3796–3801 (2009).

685 20. Song, E., Mayampurath, A., Yu, C. -Y., Tang, H., Mechref, Y. Glycoproteomics: identifying the  
686 glycosylation of prostate specific antigen at normal and high isoelectric points by LC–MS/MS.  
687 *Journal of Proteome Research*. **13** (12), 5570–5580 (2014).

688 21. Moran, A. B. et al. Profiling the proteoforms of urinary prostate-specific antigen by capillary  
689 electrophoresis – mass spectrometry. *Journal of Proteomics*. **238**, 104148 (2021).

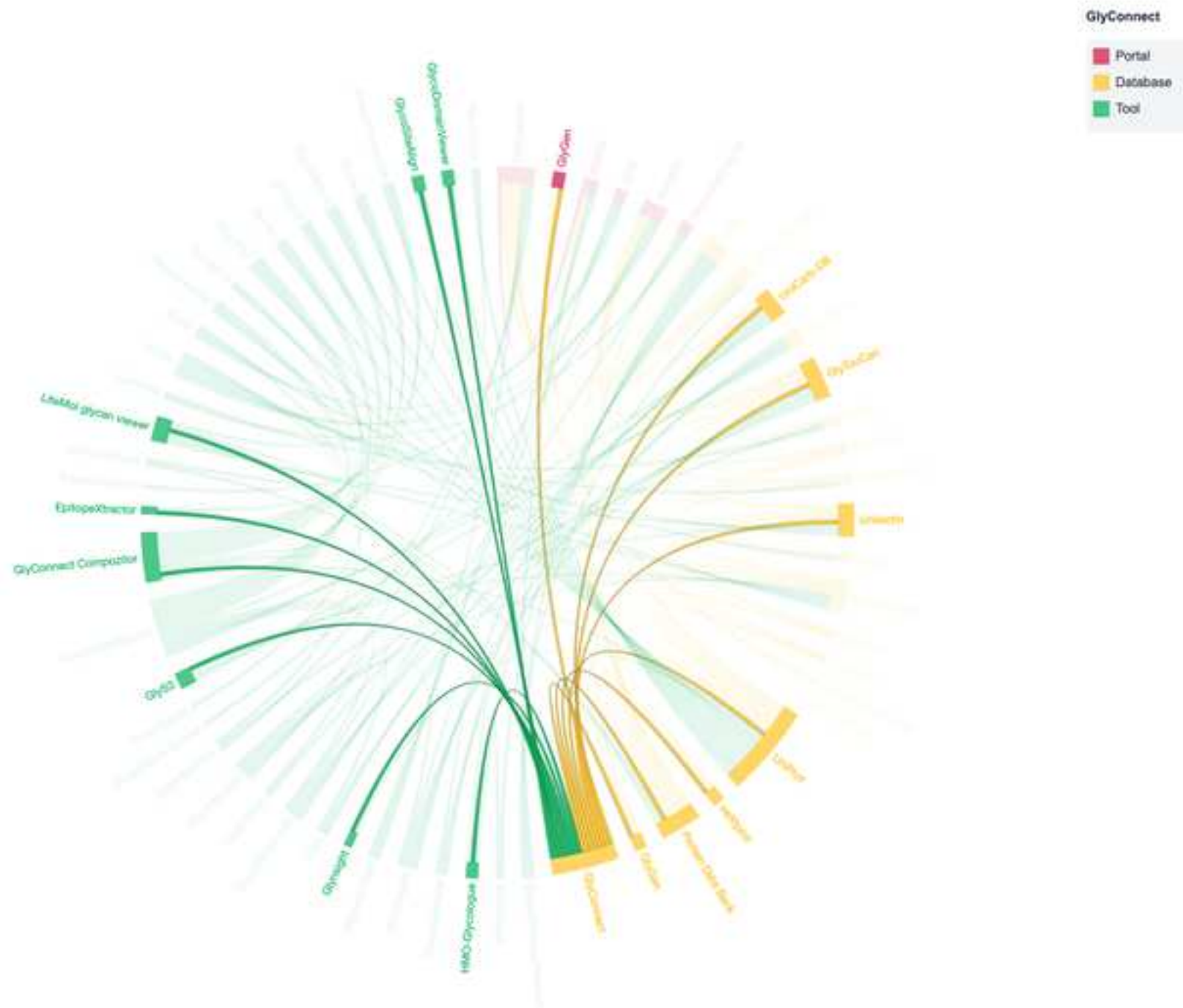
690 22. Wang, W. et al. High-throughput glycopeptide profiling of prostate-specific antigen from  
691 seminal plasma by MALDI-MS. *Talanta*. **222**, 121495 (2021).

692 23. wwPDB consortium et al. Protein Data Bank: the single global archive for 3D macromolecular  
693 structure data. *Nucleic Acids Research*. **47** (D1), D520–D528 (2019).

694 24. Sehnal, D., Grant, O. C. Rapidly display glycan symbols in 3D structures: 3D-SNFG in LiteMol.  
695 *Journal of Proteome Research*. **18** (2), 770–774 (2019).

696 25. Bonnardel, F. et al. UniLectin3D, a database of carbohydrate binding proteins with curated  
697 information on 3D structures and interacting ligands. *Nucleic Acids Research*. **47** (D1), D1236–  
698 D1244 (2019).

26. Sehna, D. et al. LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nature Methods*. **14** (12), 1121–1122 (2017).
27. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Research*. **43** (W1), W443–W447 (2015).
28. Robin, T., Mariethoz, J., Lisacek, F. Examining and fine-tuning the selection of glycan compositions with GlyConnect Compozitor. *Molecular & Cellular Proteomics*. **19** (10), 1602–1618 (2020).
29. Compagno, D. et al. Glycans and galectins in prostate cancer biology, angiogenesis and metastasis. *Glycobiology*. **24** (10), 899–906 (2014).
30. Gentilini, L. D. et al. Stable and high expression of Galectin-8 tightly controls metastatic progression of prostate cancer. *Oncotarget*. **8** (27), 44654–44668 (2017).
31. Schwämmle, V., Verano-Braga, T., Roepstorff, P. Computational and statistical methods for high-throughput analysis of post-translational modifications of proteins. *Journal of Proteomics*. **129**, 3–15 (2015).
32. Khatri, K., Klein, J. A., Zaia, J. Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Analytical and Bioanalytical Chemistry*. **409** (2), 607–618 (2017).
33. Sztain, T. et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chemistry*. **13**, 963–968 (2021)





Hex:6 HexNAc:3 NeuAc:1 / N-Linked (avg mass : 1891.7126)



SRC (1) ▾



REF (1) ▾



SITE (1) ▾



PEPT (1) ▾

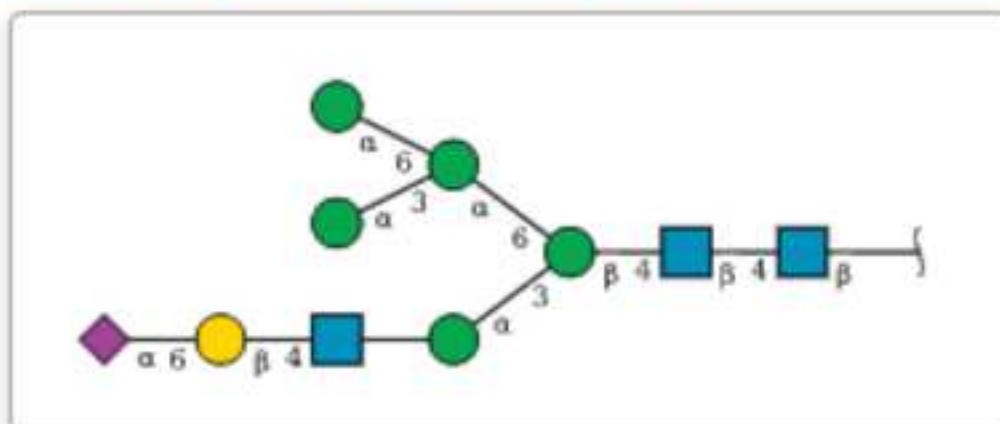


SUGGESTED STRUCT (4) ▾

## ■ Suggested structure

■ N-Linked / Hybrid / Structure 9644

■ N-Linked / Hybrid / Structure 10739



N-Linked / Hybrid

## Cross-references

UniProtKB : [P07288](#)

neXtProt : [NX\\_P07288](#)

GeneCards : [KLK3](#)

GlyGen : [P07288](#)

Compozitor : [Prostate-specific antigen \(psa-1\)](#)

## 3D structure from PDB

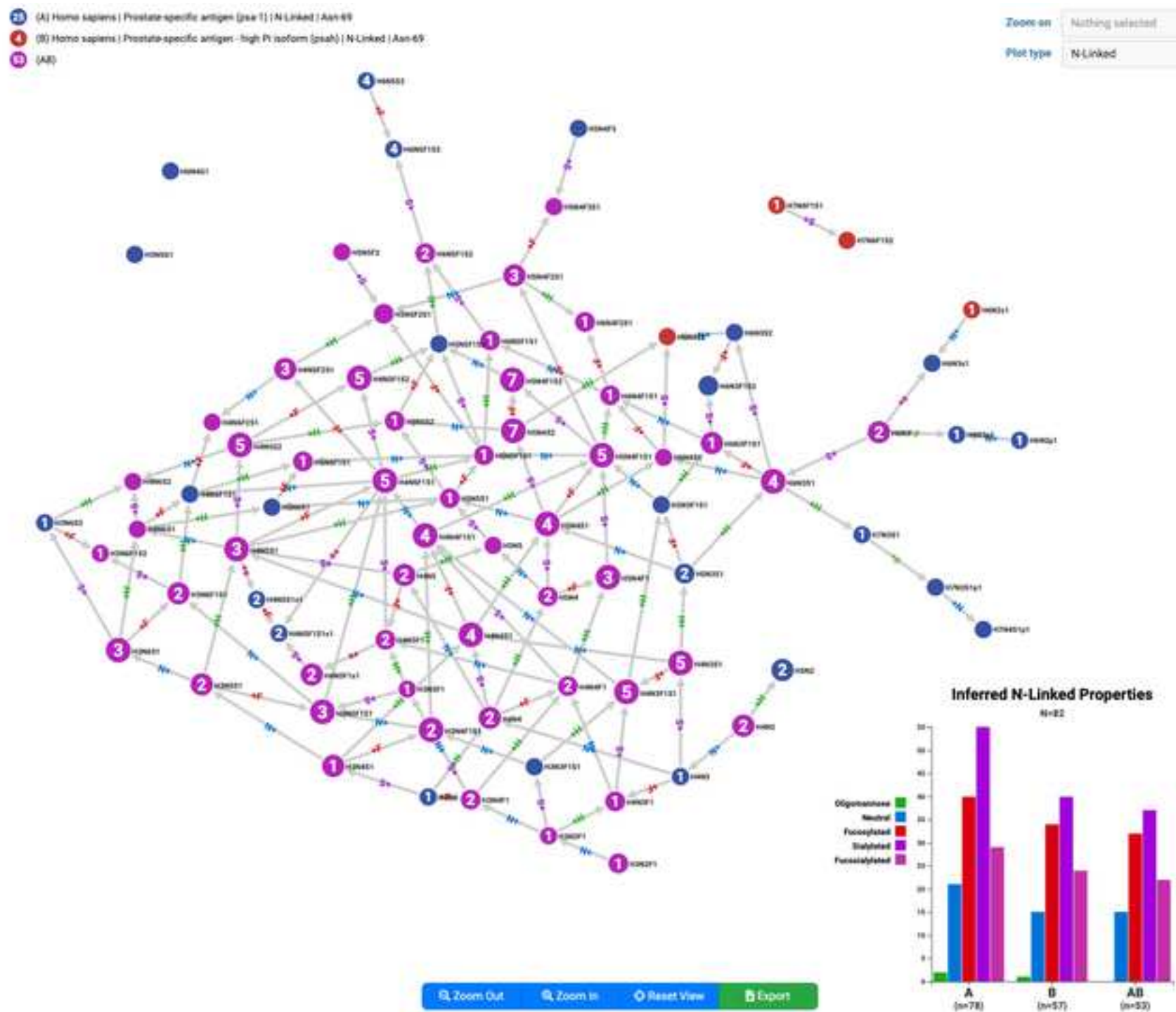
Select a PDB entry

3QUM









SWISS MODEL PLIP INTERACTIONS VIEWER

NAG- 9 residues within 4Å:  
GAL-  
SIA.1: 4 PLIP interactions:  
4 interactions with chain A

- ☒ Hydrogen bonds: A:R.45, A:R.45, A:R.69, A:N.79, A:E.89, A:R.69, A:R.69, A:Q.47, A:W.86
- ☒ Water bridges: A:E.89, A:E.89
- ☒ Salt bridges:

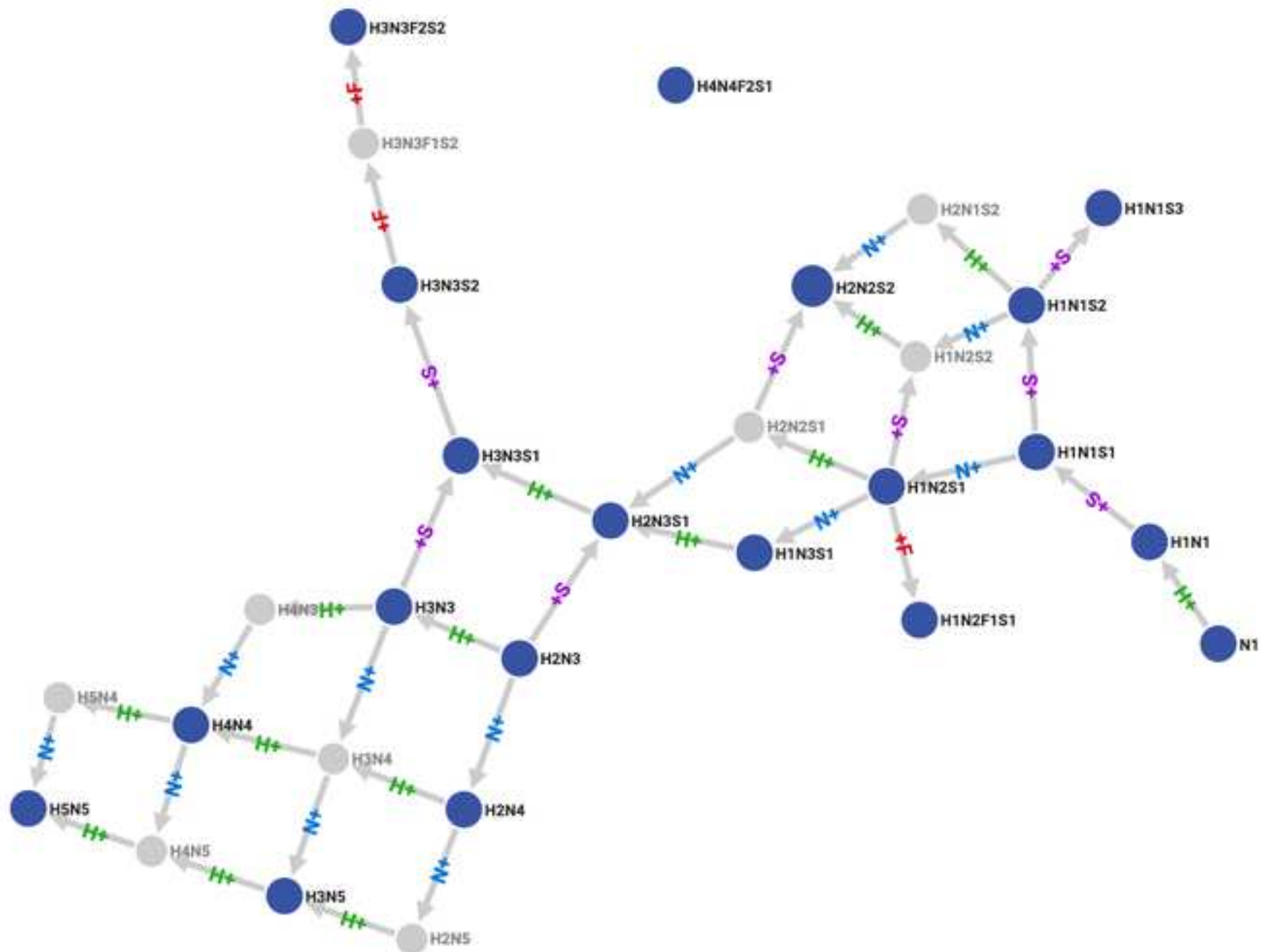
3VKO galectin-8 Homo sapiens

View the 3D structure and information

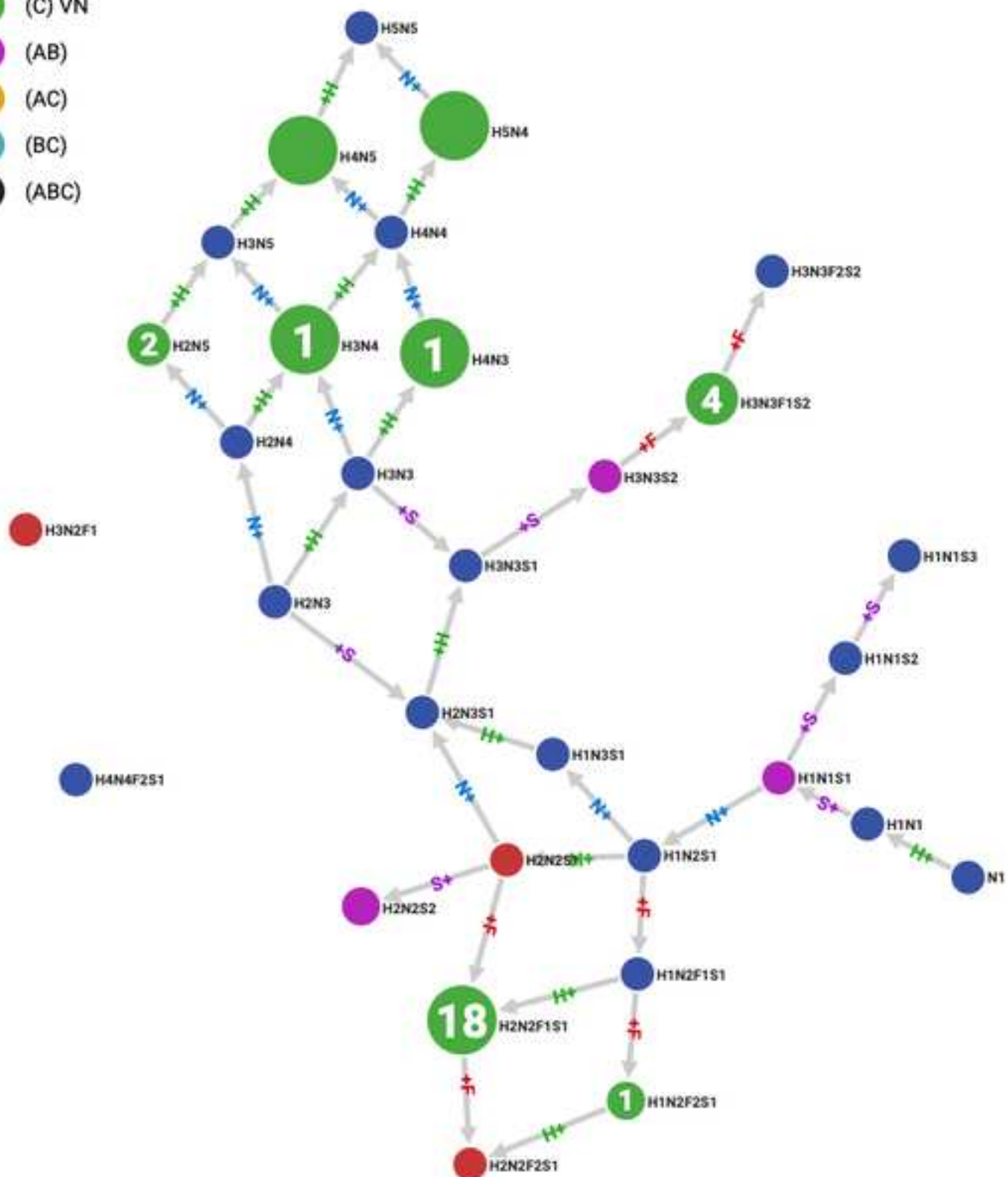
fold	b-sandwich / ConA-like	origin	Animal lectins
class	galectin-like	species	<i>Homo sapiens</i>
family	galectin-8	comments	N-term domain complex with sialyllactosamine
resolution (Å)	2.08	IUPAC condensed	NeuAc(a2-3)Gal(b1-4)GlcNAc



20 (A) reference=10.1038/s41592-021-01309-x&glycan\_type=O-linked  
Virtual



- 17 (A) reference=10.1038/s41592-021-01309-x&glycan\_type=O-linked
- 3 (B) Homo sapiens | Inter-alpha-trypsin inhibitor heavy chain h4 | O-Linked | Thr-725
- 8 (C) VN
- 3 (AB)
- 0 (AC)
- 0 (BC)
- 0 (ABC)



red	light orange	green	light blue	purple	pink	dark blue
species	protein	tissue source	structure	composition	disease	reference

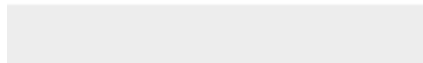
brown	dark orange
glycosite	peptide



[Click here to access/download](#)

**Table of Materials**

[JoVE\\_Materials\\_glycoinformatics.xls](#)



Dear Editor,

I was very pleased to get positive feedback from both reviewers who had no concern about my manuscript but have almost lost patience with the editorial comments that basically required more than a full day of work ending up in a completely reshuffled manuscript that I cannot read anymore due to the density of track changes. I did all I could but cannot devote any more time to cosmetic changes that suit your journal more than my idea of pedagogy. I do appreciate that you are expert in making videos but reediting everything when two reviewers are satisfied with what I produced is somehow frustrating. I have never received 20 points of editorial comments before somehow because the work is usually more shared. It is rather defeating. My time is also precious and, in the end, I will pay the bill for work that I have done.

In a nutshell, a bit of your cooperation would be appreciated so that I am alleviated.

Regards

f.Lisacek

Please find the detailed answers in the text:

**Editorial comments:**

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

[Done but there was none in the first place](#)

2. Please rephrase the Summary to clearly describe the protocol and its applications in complete sentences between 10-50 words.

[Done](#)

3. Please rephrase the Abstract to more clearly state the goal of the protocol in 150-300 words.

[The abstract was doubled in size](#)

4. Please revise the following lines to avoid previously published work: 37-40,52-56,63-66,251,259-260,270-274, 281-288,329-332,334-335,340-341,365-368,394-409. Please refer to the iThenticate report attached.

[Most of this was done, except for explanations on notations that cannot be modified \(e.g., lines 110-115\) and clear instructions \(e.g, lines 254-259\). It is actually difficult at times to write completely different sentences when the basic instructions to perform an operation are the same. There aren't that many ways to instruct a user to click on a button and even when the example is completely different, the button remains the same. In any event, conflict is minimal as cited sources were written by me.](#)

5. Please revise the text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.) especially in the Protocol section.

[Done](#)

6. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials.

Cannot understand what you are referring to. I only cited web browser names not mentioning the companies that develop them. As pointed by reviewer #1 the information in line 532 is essential...

7. Please adjust the numbering of the Protocol to follow the JoVE Instructions for Authors. For example, 1 should be followed by 1.1 and then 1.1.1 and 1.1.2 if necessary. Please refrain from using bullets or dashes.

Every step has been renumbered. I did not use any bullet or dash in the first place.

8. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “Note.” However, notes should be concise and used sparingly. Please include all safety procedures and use of hoods, etc.

All instructions are now in the imperative tense

9. Line 166-177/225-235/270-288/325-341: Please use brief notes or consider moving the lines to the Discussion section. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.

The text was substantially reshuffled to comply with this recommendation. However, there were places where additional explanations were essential (e.g., lines 208-210) and text could not be removed.

10. The Protocol should be made up almost entirely of discrete steps without large paragraphs of text between sections. Please simplify the Protocol so that individual steps contain only 2-3 actions per step and a maximum of 4 sentences per step.

Done this to the best of my ability not to threaten understanding

11. Line 134-136: Please ensure that the Protocol section consists of numbered steps. We cannot have non-numbered paragraphs/steps/headings/subheadings.

It was done in that case but a few other cases could not be solved as explained in point 9

12. Please note that your protocol will be used to generate the script for the video and must contain everything that you would like shown in the video. Please add more details to your protocol steps. Please ensure you answer the “how” question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action. Please add more specific details (e.g., button clicks for software actions, numerical values for settings, etc) to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol.

Cannot do much more than what I did in the new version

13. Please include a one-line space between each protocol step and then highlight up to 3 pages of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader.

This task is currently unmanageable with all the track changes. In any event, I am not sure I can define steps that would be more essential than others.

14. Please remove the embedded figure(s) from the manuscript. All figures should be uploaded separately to your Editorial Manager account. Each figure must be accompanied by a title and a description after the Representative Results of the manuscript text.

I do not understand this instruction that is contradictory in itself and with the template that you provide

15. Please remove the embedded Table from the manuscript. All tables should be uploaded separately to your Editorial Manager account in the form of an .xls or .xlsx file. Each table must be accompanied by a title and a description after the Representative Results of the manuscript text.

16. Please submit each figure individually as a vector image file to ensure high resolution throughout production: (.psd, ai, .eps.).

I do not have the means to do this. All my figures are screenshots so the resolution is as good as what my computer does when it captures the screen

17. Please remove the titles and Figure Legends from the uploaded figures. The information provided in the Figure Legends after the Representative Results is sufficient.

Only the figure number occurs, no title no legend. Did not know what to correct.

18. Please include a title and brief description for the supplementary figures/coding files in the Figure and Table Legends section.

I thought that was done. I do not know what else to add

19. As we are a methods journal, please revise the Discussion to explicitly cover the following in detail in 3-6 paragraphs with citations:

- a) Critical steps within the protocol
- b) Any modifications and troubleshooting of the technique
- c) Any limitations of the technique
- d) The significance with respect to existing methods
- e) Any future applications of the technique

This was already the case in the first version and was completed in the corrected version.

20. Please revise the table of the essential supplies, reagents, and equipment. The table should include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file. Please sort the Materials Table alphabetically by the name of the material.

How is this relevant to this manuscript?

---

### **Reviewers' comments:**

#### **Reviewer #1:**

Manuscript Summary:

As forefront glycoinformatics leader, Frederique Lisacek has provided the comprehensive



overview on the tools available at Glyco@Expasy. The described steps are clear, concise and easy to follow for the researchers entering this field.  
I strongly recommend this manuscript for publication.

Major Concerns:  
None

Minor Concerns:

- Possibly few references can be added for the Introduction, particularly for the lines 25 to 40  
[Done](#)

- Line 56: Please correct glycans to glycopeptides.  
[This line was erased to lighten the text. This level of detail was not useful in the introduction.](#)

- Line 111 and 113 - hexosamine should be addressed as N-acetylhexosamine to avoid ambiguity  
[Done](#)

- Glyconnect Octopus' - Step 5: Glycan composition needs to be corrected from H6N4F1S2 to H6N4F2S1  
[Done](#)

- Author has provided information about the compatibility of the web resource in Line 532. But from a user perspective it would be great to have this at the start of the protocol.  
[This is not my choice but guided by the template. I am happy to move this information upstream as suggested by the reviewer.](#)

## **Reviewer #2:**

Manuscript Summary:

The manuscript talks about the glycomics and/or glyco-intrectomics resources present and there usage through examples. The author has explained Glyco@Expasy resource in the manuscript. The manuscript is well written with proper usage of examples and step wise instructions to use the resource. The sections are well supported by results and figures. The author has also talked about the limitations of the study which is helpful for the users while using or during interpretation of the results.

Major Concerns:  
None

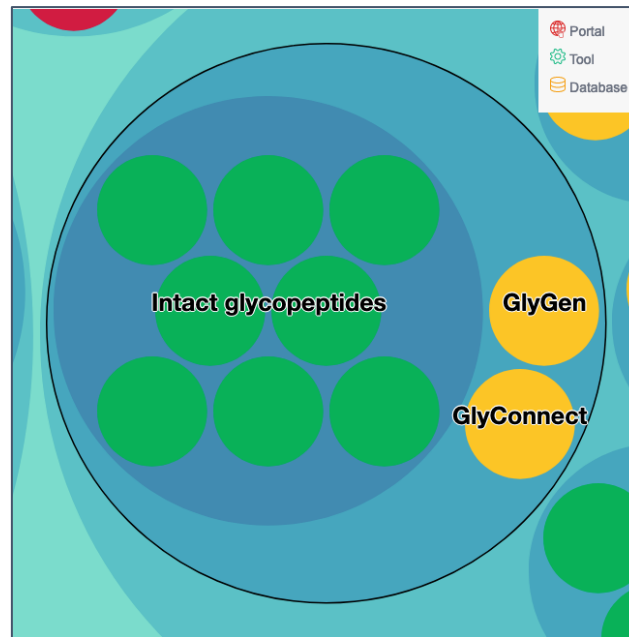
Minor Concerns:

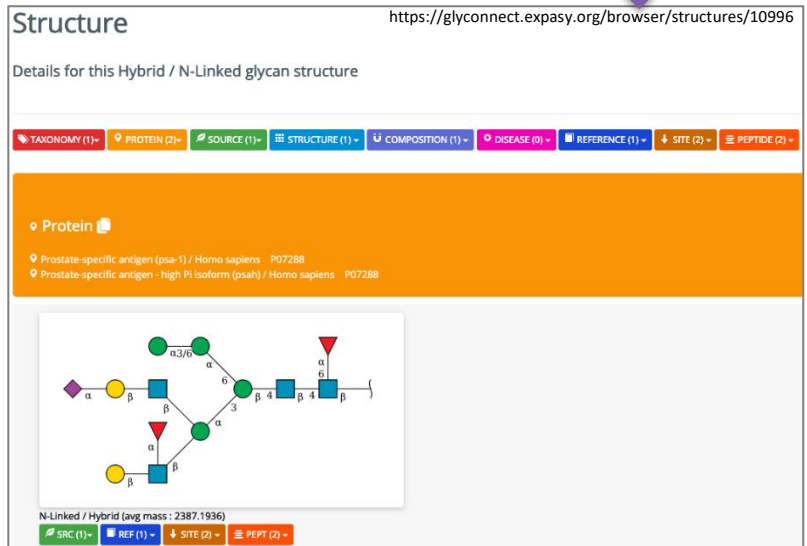
It would be helpful if the interpretation of the results can be explained in detail so that it can help the beginners in the field.

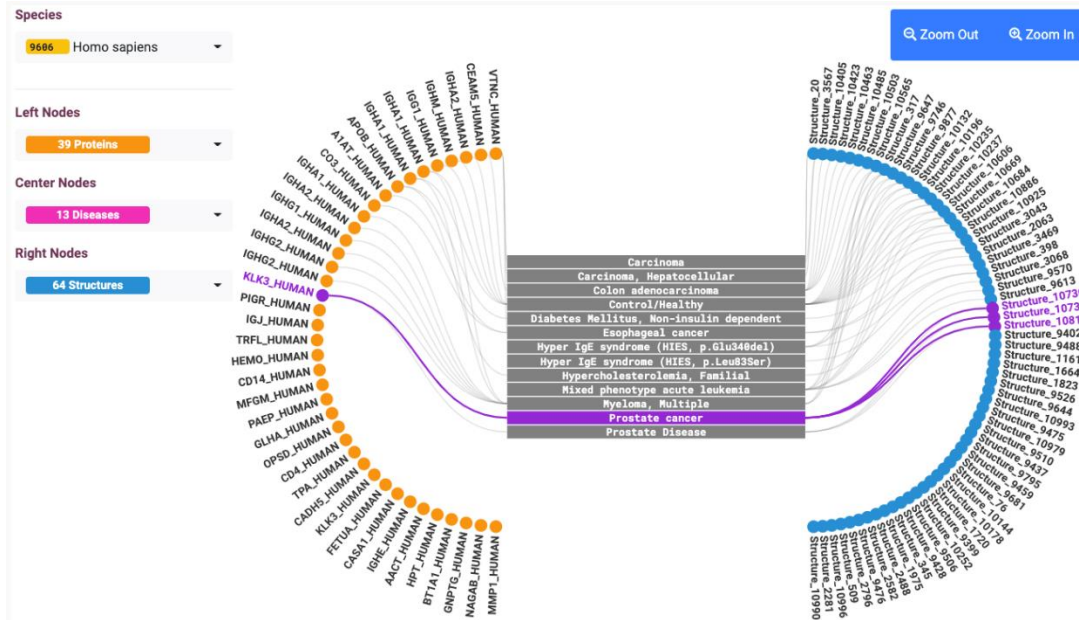
[Additional text was input to that end](#)

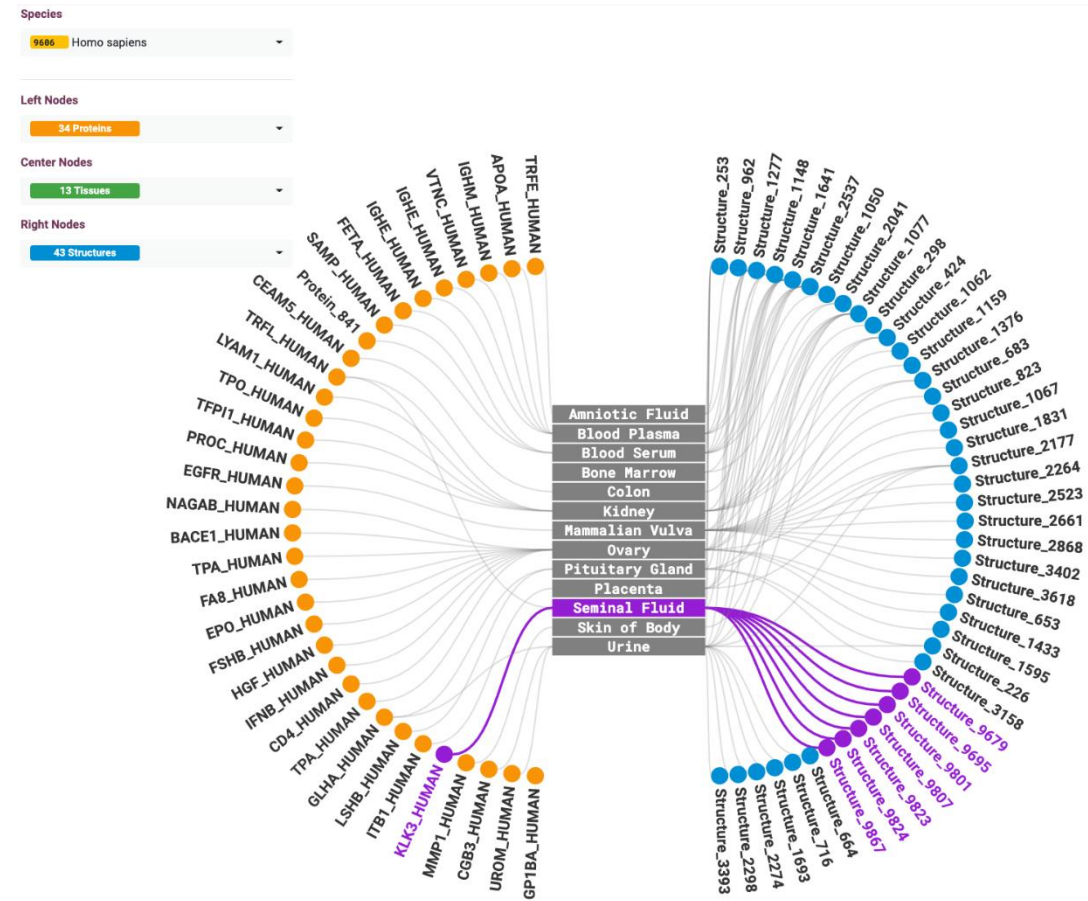
Video Produced by Author: Less than 50 MB. If your video is greater than 50 MB, click "offline" as the delivery method and our

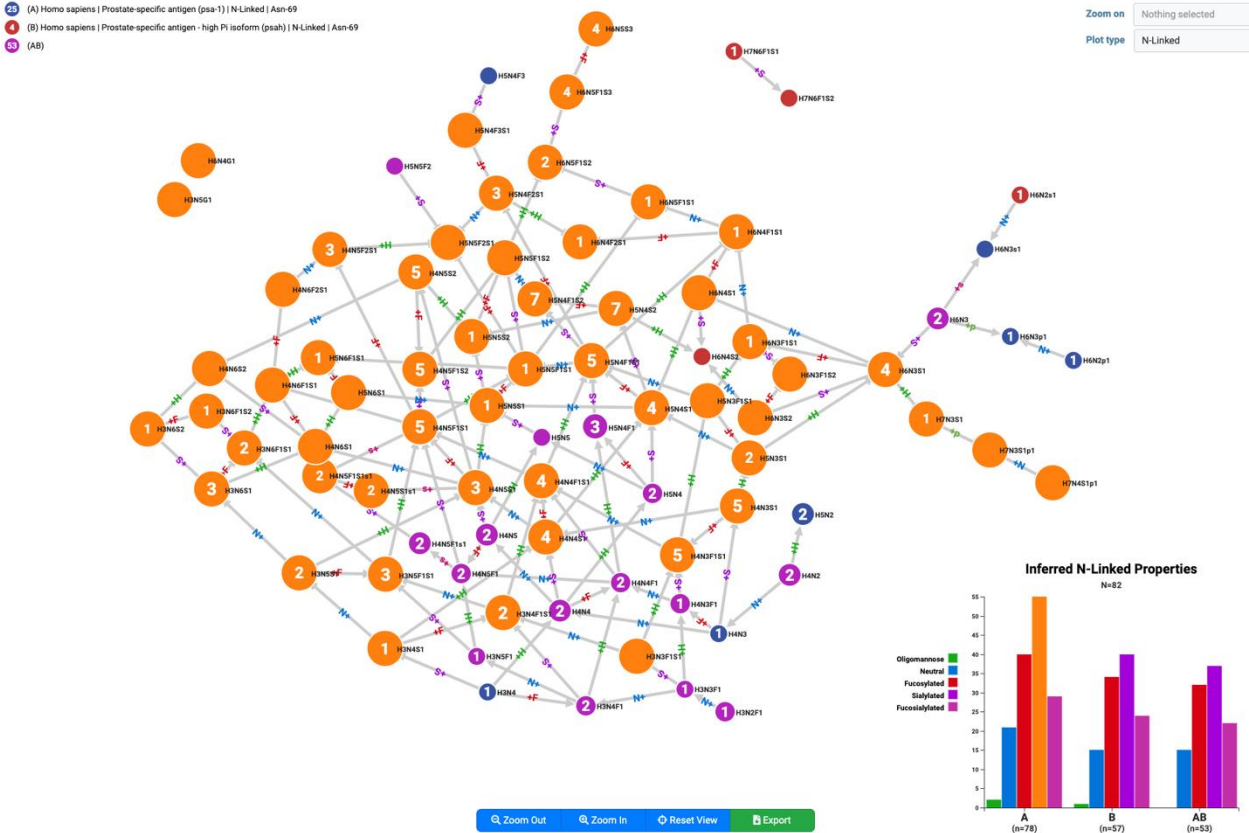
This piece of the submission is being sent via mail.




































UniLectin3D glycan search

D-Abep 	D-Arap 	D-Fruf 	D-Frup 	D-Fucp 	D-Galnq 
D-Galf 	D-Galp 	D-GalpNAc 	D-Glcp 	D-GlcpN 	D-GlcpNAc 
D-Manp 	D-ManpN 	D-ManpNAc 	D-Murp 	D-NeupNAc 	D-Neup5Gc 
D-Talp 	D-Tyvp 	D-Xylp 	DD-manHepp 	L-Fucp 	L-Galp 
L-Rhap 	LD-manHepp 