GlycoBrowser: A Tool for Contextual Visualization of Biological Data and Pathways Using Ontologies

Matthew Eavenson[1], Maciej Janik[1], Shravya Nimmagadda[1],
John A. Miller[1], Krys J. Kochut[1], and William S. York[2]

[1]Computer Science Department
[2]Complex Carbohydrate Research Center
University of Georgia
Athens, GA 30602
{durandal, mjanik, shravyan}@uga.edu
{jam, kochut}@cs.uga.edu
{will}@ccrc.uga.edu

Abstract. This paper introduces GlycoBrowser, a dynamic biological pathway exploration and visualization tool. Its capabilities include incremental navigation of pathways, as well as overlaying relevant experimental data about glycans and reactions within a pathway. The use of ontologies not only allows dynamic construction of pathways, but also facilitates more accurate validation of the information retrieved. Pathway exploration is initiated by means of an ontologically driven glycan structure building tool, which facilitates glycan structure construction and searching with minimal user error. Because of the complex nature of glycan structures and the difficulty which is involved in interpreting the associated data, GlycoBrowser is tailored especially to suit the needs of glycobiologists.

Keywords: GlycoBrowser, pathway browsing, metabolic pathway exploration, glycomics, transcriptomics, ontology, semantic web.

1 Introduction

The amount of data and knowledge stored in Web-accessible databases and ontologies is growing explosively. Making this information widely available in an intuitive form is becoming almost as important as creating this information in the first place. Traditionally, Web 1.0 technology, relying on servlets and Java Server Pages (JSP) to interface with relational databases, has been the preferred way to develop such online resources [3]. The approach that we are taking utilizes Web 2.0 as well as Semantic Web technology to develop a more agile means for querying and browsing biological information at a higher level than most on-line resources. The key component of our approach is the pathway exploration and visualization tool, GlycoBrowser.

GlycoBrowser is multi-faceted, allowing flexible visualization of both glycomic pathways and construction of glycan molecules. It leverages the capabilities of GlycoVault [18], along with several Web 2.0 and Semantic Web technologies (including RDF [14], OWL [16], SPARQL [20], AJAX, Javascript, and

BRAHMS [11]) to represent structural information in a way that is intuitive for glycobiologists and to overlay this information with experimental data (glycomics, proteomics, and transcriptomics analyses). GlycoBrowser provides a graphical display of glycan biosynthetic pathways and associated experimental data. Glycans are rendered "on the fly" using the standard representation endorsed by the Consortium for Functional Glycomics (CFG, http://www.functionalglycomics.org), extended to include partonomy relationships. Relevant experimental data (such as the abundance of a particular glycan in a biological sample) can be shown in associated histograms.

This paper discusses the motivation, design and implementation of GlycoBrowser and is organized as follows. Section 2 considers why more agile, flexible and high level approaches that feature visualization provide a better means of accessing biological information. In section 3, we narrow our focus and review the evolution of biochemical pathway visualization tools and where GlycoBrowser fits in. An important aspect of our approach is that we use knowledge encoded in ontologies, those available on the Web as well as ones we are developing, as discussed in section 4. Section 5 presents the design and implementation of our pathway browser and its associated structure builder. Conclusions and future work are given in section 6.

2 Motivation

A major challenge of modern biological science is to analyze and interpret the huge amount of data that is routinely collected by high-throughput techniques, as in proteomics or glycomics analysis. Due to the complexity of biological systems, it is not sufficient to simply facilitate access to specific data in these large sets; it is necessary to present these data in the context of what is known about the biological system being studied. One approach to address this requirement is to present data (such as proteomics or transcriptomics data) in the context of a metabolic or signaling pathway in a process called “data overlaying” [23]. In the domain of glycobiology (the study of complex carbohydrates and their functional roles in living organisms), such a contextual illustration of data is complicated by the fact that the pathways leading to glycan biosynthesis are extremely complex, and each of the molecules in the pathway is a complex aggregation of smaller “glycosyl residues”. Therefore, graphical representation of the pathway requires a robust internal representation of the structures along with algorithms for rendering them in a format that can be intuitively interpreted “at a glance” by the scientist.

We have developed an ontology (GlycO) that embodies knowledge regarding the structures, biosynthesis, and biological functions of complex glycans [22]. GlycO thus addresses the challenge of providing a robust internal representation of glycan structures and the metabolic pathways leading to their synthesis. In addition, GlycO comprises a repository of knowledge regarding other aspects of glycobiology, including structural and functional relationships between different glycans. GlycoBrowser allows this knowledge to be accessed and used as a context for the graphical representation of experimental data. We initially focus on two types of data: (1) qRT-PCR data, revealing the abundances of mRNA transcripts for genes involved

in glycan biosynthesis and (2) glycomics data, revealing the identities and abundances of specific glycans in a biological sample. Ultimately, GlycoBrowser will not only provide a means of overlaying data on metabolic pathways, it will provide an entry point into the diverse types of knowledge embedded in GlycO and other ontologies.

The following example illustrates the need for a tool with the capabilities of GlycoBrowser. We are interested in the relationships between the differentiation of stem cells (to form more specialized cells) and the expression of specific glycans on their surfaces. Our collaborators in the Integrated Technology Resource for Biomedical Glycomics (http://glycomics.ccrc.uga.edu) have developed and implemented powerful methods for obtaining transcriptomic and glycomic data for this experimental system. The transcriptomic data reveals the expression levels of specific enzymes involved in the metabolic pathways leading to glycan biosynthesis and the glycomics data reveals the amounts of specific glycans that are generated by these complex pathways. Manual analysis of this data indicated that specific glycans (sialylated biantennary N-glycans) are more abundant on the surfaces of differentiated cells than on undifferentiated stem cells. Moreover, some of the enzymes (sialyl transferases, which catalyze the addition of a sialic acid residue to the nascent glycan) required for the biosynthesis of these glycans are also upregulated during differentiation. The question thus arises, "Are the changes in the abundances of these glycans due solely to an increase in sialyl transferase expression, or do changes in the expression of other enzymes contribute significantly to this effect?" In order to answer this question, a glycobiologist might draw all of the relevant pathways and overlay the relevant transcriptomic and glycomic data at each step of the pathway. However, rendering this pathway is not trivial, as each of the molecules along the pathway is a complex glycan, and the pathway is highly branched. It would be difficult to find an appropriate entry point into the pathway (for drawing) and it would also be difficult to select the appropriate branch(es) of the pathway that lead to the glycans of interest.

Our initial implementation of GlycoBrowser addresses these challenges in several ways. It provides a graphical interface that allows an entry point (a specific glycan) to be selected from a large collection of structures in the knowledge base. Glycans are complex, branched molecules. Therefore, it is much more difficult to select a glycan from a collection of glycans than to select a protein sequence from a collection of protein sequences. It would be highly impractical to expect scientists to memorize the accession numbers of thousands of glycan structures, and thus it is necessary to provide a graphical tool for selecting specific glycans. Thus, a tool that enables a search for specific glycans based on their structural features, but disallows structures not found in nature would be quite useful, as it would eliminate time wasted searching for physically impossible combinations of structural features.

GlycoBrowser provides such a tool, and allows the user to define an entry point into the metabolic pathway. Complex, branched metabolic pathways are rendered using an accepted graphical representation of the glycan structures. GlycoBrowser also overlays specific transcriptomic and glycomic data at each step along the pathway. Analysis of our transcriptomic and glycomic data using GlycoBrowser demonstrated that increased expression of sialyl transferases in not the only factor leading to the increased abundances of sialylated biantennary N-glycans in differentiated cells. The abundances of precursor glycans (which are substrates for the

sialyl transferases) are also elevated, indicating that other steps in the pathway leading to the sialylated biantennary N-glycans are modified in the differentiated cells.

3 Background

Dynamic molecule and pathway construction is not a new approach. Multiple research areas impose different requirements for visual representation of pathways. As a result, there exist many tools and browsers for visualizing biological structures or pathways that can be enhanced with experimental data and other information. Among the most popular are KEGG [19], WikiPathways [13], MetaCyc [12], PathwayAssist [17], GenMAPP [7], and Cytoscape [24]. Although all these tools are designed to be comprehensive, each emphasizes different aspects of data presentation.

KEGG and WikiPathways offer static pathways visualization, pathway interconnectivity and also allow focus on many levels of granularity such as meta-pathway, pathway fragment, or details of a single compound. They serve as referral sites and repositories of known pathways. Dynamic pathway browsing is offered by MetaCyc, PathwayAssist, GenMAPP, and Cytoscape. These systems allow a user to query for pathway fragments. Their search capabilities may include finding pathways between specific molecules, gene regulators or specific reactions based on given properties. They use a database to find relevant information for visualization and offer multiple layouts and presentation schemes to present focused information.

Information overlay is another important aspect of pathway visualization. Dynamic browsers support enhancing pathway elements with accompanying experimental data or related information from biological databases and special ontologies, such as the Gene Ontology (GO) [10]. Information overlay allows scientists to easily see results of their experiments in proper biological context.

GlycoBrowser is a dynamic visualization tool created specifically to handle complex glycomics data. It supports search for glycans based on their structural features, which are selected using a graphical user interface. Contrary to unrestricted graphical editing in Glycan Builder [6], our system uses an ontology to guide the building process and restricts the user to creating only structures that fit into one of the canonical trees that represent all glycan structures in the ontology [22, 25]. A glycan located using this process may be used as a starting point for further pathway exploration which allows a user to dynamically traverse interesting pathways. Displayed glycans and reactions are then overlaid with associated experimental data. This guided approach to molecule and pathway construction minimizes user error, while also allowing a biologist to explore data, rather than simply sifting through it.

4 Underlying Ontological Representations

As mentioned in the introduction, more and more ontologies are becoming available on the Web. Along with relational databases, they make up the principal means of providing structured information on the Web. Making such information available to end users in meaningful and convenient ways has lead to massive software

development over the last decade. In particular in biosciences, many well developed ontologies are now available on the Web. For example, at the Open Biomedical Ontologies (OBO) Foundry [2], there are now over sixty ontologies registered. Finally, an important focus of our glycomics project is the study of glycans and the effect that enzymes have on their biosynthesis. We have therefore built and are in the process of populating the following two ontologies: GlycO and EnzyO [22, 25]. These ontologies were developed with Protege using the Web Ontology Language (OWL) [16]. For efficiency of pathway browsing, export to Resource Description Framework (RDF) [14] is also provided.

The Glycomics Ontology (GlycO) [25] has been preloaded with the residues that make up glycans. Then to load any glycan into the knowledge-base, one simply needs to add this new instance with links to the existing residues. The canonical approach reduces redundancy and increases the reliability of the information stored, since we at least know that the preloaded residues are correct. In addition to this information and knowledge about structures, GlycO also links structures to reactions that can make them. The population of GlycO is facilitated by GlydeII [26] which is an XML based data interchange standard for glycan structures. GlydeII facilitates validation and incorporation of new glycan structure instances by comparison to knowledge stored in GlycO. The Enzyme Ontology (EnzyO) keeps track of enzymes that catalyze the reactions which produce the glycan structures. The identities and abundance levels of glycans and enzymes are the essential components in the biochemical pathways. The ontology keeps track of basic information about enzymes (e.g., their Enzyme Commission (EC) number, their protein structure) as well as associations (e.g., with the gene that codes for it and the reactions it participates in).

5 GlycoBrowser

GlycoBrowser consists of two submodules, the Canonical Structure Builder and Pathway Browser, and utilizes three related subsystems, the GlycoVault [18] repository, SPARQL Server, and Image Server. We will describe each of these in detail, beginning with the lower level components (GlycoVault, SPARQL Server, and Image Server), and concluding with the Glyco Browser functionality and implementation.

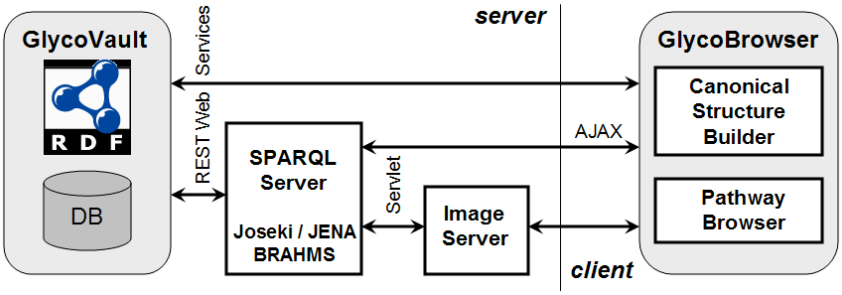

Fig. 1. System architecture

5.1 GlycoVault

GlycoVault provides a means of storing and retrieving data to support glycomics research at the Complex Carbohydrates Research Center (CCRC) at the University of Georgia. These data include quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) data as well as basic glycomics data, such as biologically relevant parameters and various types of experimental data, along with the explicit and implicit knowledge required to analyze and interpret these data. GlycoVault consists of databases, ontologies (including GlycO and EnzyO), and data files in various formats. These datasets and ontologies are accessed by the Pathway Browser via a comprehensive, yet easy to use Application Programming Interface (API). The API facilitates the development of methods for querying the knowledge and exporting the results in formats (such as XML, RDF or OWL) that can be readily incorporated by external applications. GlycoVault provides easy to use query interfaces, including support for the SQL and SPARQL query languages.

5.2 SPARQL Server

The SPARQL Server acts as an ontology storage and querying facility for the Pathway Browser. It provides access to any ontological knowledge available, whether it be the next possible branch at each step along a pathway, or any additional metadata about reactions or molecules. SPARQL was the obvious query language of choice, since we are taking an ontology-based approach and require a query language that works with RDF (or OWL). Currently, we use Joseki [1], a Jena [5] based HTTP engine that supports the SPARQL Protocol and query language. Joseki allows responses to be encoded in either XML or JavaScript Object Notation (JSON) [21]. We have chosen JSON because it meshes seamlessly with the Javascript user interface. However, as an added benefit, JSON, being a lightweight data-interchange format, is much less verbose than an XML encoding, meaning less data to transfer over the web and less bandwidth consumed. In the near future, we intend to migrate to a high-performance BRAHMS based server based on the SPARQLeR [15] extension to SPARQL. This will allow us to retrieve entire pathway fragments from the ontology, allowing fragments to be constructed in their entirety within the Pathway Browser instead of incrementally as described later.

5.3 Image Server

The Image Server dynamically constructs cartoonist [9] representations of glycan molecules, based on the underlying ontological structure. The cartoonist representation has been widely accepted and endorsed by the Consortium for Functional Glycomics. Patterns present in a cartoonist representation (as depicted in Figures 2, 3, and 4) are far more readily recognizable at-a-glance than the name of a glycan molecule, or its textual representation, making it the obvious approach for implementing a pathway browser whose purpose is to be intuitive. However, the complexity of dynamically generating cartoonist models of glycans, in tandem with

the desire for modular design, necessitated the creation of a separate, dedicated image drawing subsystem. We felt it was preferable to shift drawing to a dedicated module both for speed and for making the client more lightweight. The benefits of having a fast, dynamic, and modular image drawing subsystem connected directly to the GlycO ontology are readily apparent, as any changes to the underlying structures in the ontology will not require the generation of new static images. The resulting images are used in both the Canonical Structure Builder and Pathway Browser, and can also be utilized in any future extensions.

The image construction algorithm itself is implemented in C++, and makes use of the BRAHMS API (one of the fastest currently available RDF stores) for storage and querying of RDF data, as well as FastCGI [4], which facilitates efficient querying over the web. We utilize the GraphViz API [8] to automatically draw glycan molecules. However, we have found that carefully tweaking the GraphViz input can provide more suitable layouts for glycan molecules. All internal molecule representations are generated dynamically by recursively traversing the structure of the GlycO ontology, piecing together the component residues of a glycan molecule. The resulting graph structure is then passed to GraphViz which renders the collected nodes and edges into a cartoonist representation of a glycan. We ultimately chose GraphViz because of its easy to use API, choice of image formats (SVG, PNG, JPEG, GIF), and straightforward image customization options. Our image format of choice is PNG, owing to its non-proprietary format and improved compression characteristics over GIF.

5.4 Canonical Structure Builder

The Canonical Structure Builder serves as an entry point to pathway visualization, as well as a convenient search tool to look for glycan structures. The user may graphically construct a glycan molecule using component residues. However, in our novel approach, the construction is guided by the underlying ontological structure, thus reducing the amount of possible user error. The structure can then be matched against glycans currently represented in the ontology, either finding larger glycans which contain it, or the exact glycan itself.

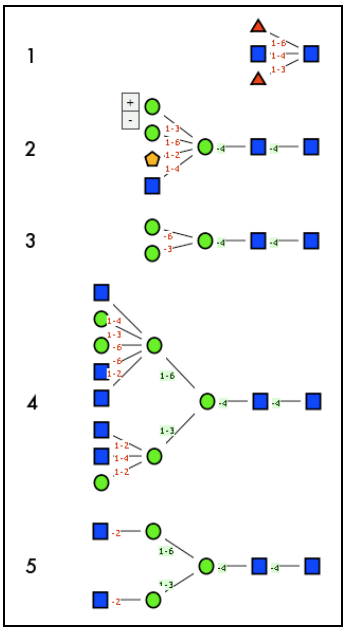

Fig. 2. Structure builder scenario

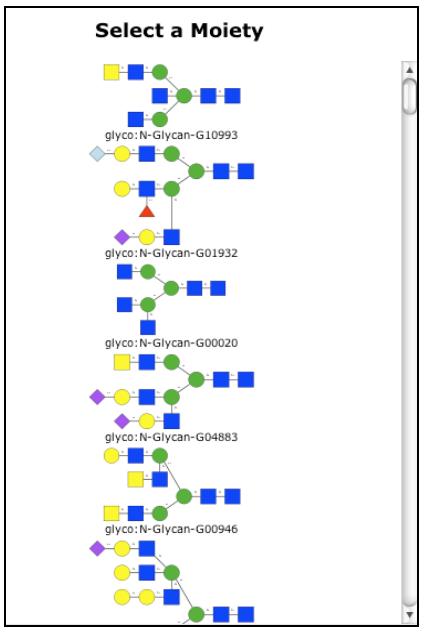

Fig. **3.** All matching glycans

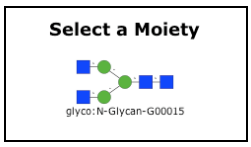

Fig. 4. Exact matching glycan

As a sample scenario (Figure 2), the user is first presented with a list of available root residues of the glyco tree to choose from. Once a root is chosen, the residue is rendered into the main viewport, with any connected residues branching off from it to the left (step 1). The user can then construct a glycan representation by expanding or collapsing residues based on the glyco tree until a desired configuration is achieved (steps 2-4). Residues are expanded or collapsed through the use of a simple "plus/minus" toggle that appears when mousing over the residue, as depicted in step 2. Once the final molecule is constructed (step 5), the user can then search for the structure in the knowledge-base. The "Match Glycans" button is used to search for all glycans which contain the given structure, with the results arranged in a list as presented in Figure 3. However, if a user wants to search for only that particular configuration, then "Exact Match Glycans" can be used, as shown in figure 4. Selecting a glycan from the results initiates pathway exploration starting from the selected glycan.

5.5 Pathway Browser

The Pathway Browser allows exploration of a biological pathway beginning from a user-selected glycan. In keeping with our desire for maximum usability, the interface layout has been kept as simple and intuitive as possible. The pathway is created dynamically, with the data that forms the pathway structure itself coming from the SPARQL Server, and any related experimental data coming from the GlycoVault.

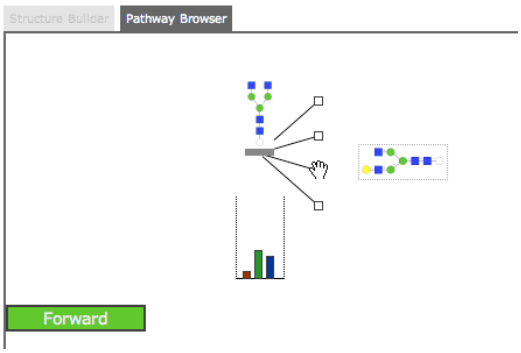

Fig. 5. Start of pathway exploration

The pathway is rendered as a series of nodes and edges. Molecules are rendered by the Image Server (discussed previously), which makes use of GlycO, the same ontology used for constructing the pathway. Reactions which have not yet been expanded are rendered as small squares, while already expanded reactions are rendered as arrows between molecules. Positioning the mouse pointer over a reaction that has not been expanded previews the molecules resulting from the reaction beside it, as depicted in Fig. 5. Exploring a pathway consists of expanding one potential branch point at a time. Potential branch points within a pathway consist of the molecules or reactions themselves, as a molecule can be used as a substrate in multiple reactions, and a reaction can have multiple products.

Related experimental data is rendered into small bar graphs located directly beneath molecules and reactions. The data under molecules represent glycan abundances, while the data under reactions represent transcriptome expression levels (corresponding to enzyme abundance levels). If experimental data is unavailable, then the corresponding graph is left empty. However, if there is more than one set of data for a particular node, the graph will present controls allowing the user to cycle through them.

A sample pathway fragment illustrating the described features is depicted in Fig. 6, starting from the constructed glycan, as described in the previous section. This figure illustrates the answer to the aforementioned question, “Are the changes in the abundances of these glycans due solely to an increase in sialyl transferase expression, or do changes in the expression of other enzymes contribute significantly to this effect?” In fact, a glycobiologist will have the relevant pathway with associated overlaid transcriptomic and glycomic data at each step of the pathway.

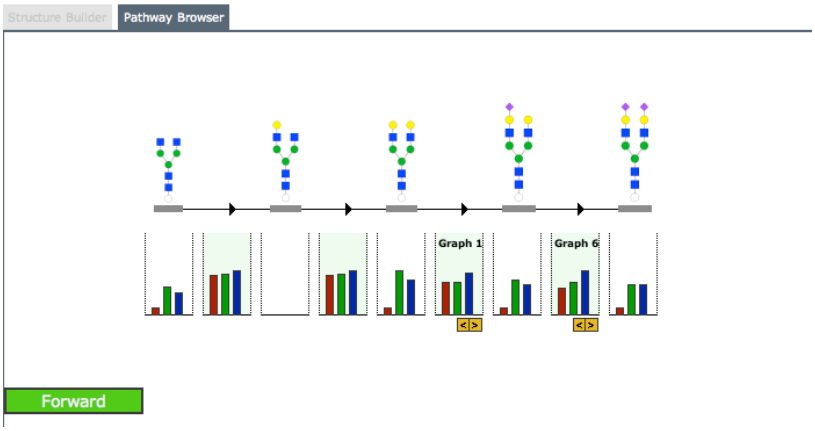

Fig. 6. Pathway fragment with corresponding graphs

Additionally, the pathway traversal direction can be toggled by clicking on the Forward/Backward button, with the currently displayed label dictating which way the pathway will be traversed. A glycobiologist may navigate forward or backward on a pathway to pinpoint where abundance levels significantly change. Backtracking to any point in the pathway is also possible by clicking on an already expanded node. This allows revisiting prior points along the pathway, as well as exploring other branches.

5.6 Implementation choices

We are utilizing both Web 2.0 technology as well as Semantic Web technology to provide information at a higher level than offered by most on-line resources. The user interfaces for both the Canonical Structure Builder and the Pathway Browser are implemented using Javascript. We chose Javascript because a web accessible pathway browsing tool will be most useful for glycobiologists, as it can be used from almost any location, and provides platform independence. The decision to favor Javascript as opposed to a Java applet was also made to keep the interface lightweight, as downloading an applet is time and bandwidth consuming as well as problematic on certain web browsers.

Moreover, the decision to use Javascript allows us to utilize the AJAX framework to incrementally download new data as needed. AJAX allows small HTTP requests to be sent in the background, thus never requiring a refresh of the webpage which makes the interface more responsive, and generally provides a better user experience. Using AJAX also helps keep memory demand on the web browser low because it does not necessitate the download of an entire data-source at once. The AJAX requests carry SPARQL queries to the aforementioned SPARQL server, which then returns relatively small chunks of data to the interface for rendering. Future improvements to the Pathway Browser will employ AJAX to incorporate precaching of likely future pathway selections.

Query results are returned in JSON to more closely mesh with the Javascript user interface. We chose JSON over SPARQL XML because returning results in XML required a Javascript based parsing algorithm to make use of them. JSON, on the other hand, offers the benefit of not requiring any parsing algorithms when being used within Javascript. Javascript can evaluate JSON as a native object and use it as a nested array-like structure, which is far simpler to work with.

As a result of these decisions, the GlycoBrowser client interface is only loosely tied with the data services it relies on, thus allowing them to be easily swapped out if necessary. Also, the browser retains platform independence by utilizing web technologies, and achieves greater efficiency through the use of a lightweight client application.

6 Conclusions and Future Work

Although there are many pathway visualization tools available, GlycoBrowser, a metabolic pathway exploration tool, is particularly suitable for glycobiologists. While knowledge, structures, and pathways are represented in the form of OWL ontologies, the graphical presentation is automatically constructed using molecule representations widely accepted by glycobiologists. Experimental data is overlaid in a contextually meaningful way. For example, as shown in Fig. 6, GlycoBrowser positions the glycomic and transcriptomic data in a way that facilitates the identification of correlations between these datasets as embryonic stem cells differentiate.

In the near future we plan to expand our toolset to include curation tools, allowing us to directly modify the GlycO and EnzyO ontologies to dynamically add new structure and pathway representations. We plan to further enhance the capabilities of GlycoBrowser to simultaneously explore multiple pathway branches for comparison purposes. We also plan to include search capability for entire pathway fragments, in addition to currently available incremental exploration. Furthermore, we plan to add comprehensive filtering of overlaid experimental data.

References

1. Joseki. http://www.joseki.org
2. Open Biomedical Ontology
3. Aurrecoechea, C., Heiges, M., Wang, H., Wang, Z., Fischer, S., Rhodes, P., Miller, J., Kraemer, E., Stoeckert, C.J., Roos, D.S., Kissinger, J.C.: ApiDB: integrated resources for the apicomplexan bioinformatics resource center. Nucleic Acids Research **35** (2007) D427-D430
4. Brown, M.R.: FastCGI: A High-Performance Gateway Interface. Programming the Web - a search for APIs workshop. 5th Int. World Wide Web Conference, Paris, France (1996)
5. Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K.: Jena: Implementing the Semantic Web Recommendations. Proc. of the 13th Int. World Wide Web Conference (WWW), New York (May 2004) 74-83
6. Ceroni, A., Dell, A., Haslam, S.M.: The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. Source Code for Biology and Medicine **2:3** (2007)

7. Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C., Conklin, B.R.: GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. Nature Genetics **31** (2002) 19-20
8. Ellson, J., Gansner, E., Koutsofios, L., North, S., GordonWoodhull: Graphviz - Open Source Graph Drawing Tools. Lucent Technologies/AT&T Labs Research Tech. Report (2000)
9. Goldberg, D., Sutton-Smith, M., Paulson, J., Dell, A.: Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. PROTEOMICS **5** (2005) 865-875
10. Harris, M.A., Clark, J.I., Ireland, A.: The Gene Ontology (GO) project in 2006. Nucleic Acids Research **34** (2006) D322-D326
11. Janik, M., Kochut, K.: BRAHMS: A WorkBench RDF Store And High Performance Memory System for Semantic Association Discovery. Proc. of the 4th Int. Semantic Web Conference (ISWC), Galway, Ireland (November 2005)
12. Karp, P.D., Riley, M., Paley, S., Pellegrini-Toole, A.: The MetaCyc Database. Nucleic Acids Research **30** (November 2002) 59-61
13. Kelder, T., Pico, A.R., Iersel, M.v., Conklin, B., Evelo, C.: WikiPathways: Pathway Editing for the People. 11Th Annual International Conference on Research in Computational Biology, SanFrancisco (April 2007)
14. Klyne, G., Carroll, J.: Resource Description Framework (RDF). W3C recommendation (February 2004)
15. Kochut, K., Janik, M.: SPARQLeR: Extended SPARQL for Semantic Association Discovery. Proc. of the 4th European Semantic Web Conference (ESWC), Innsbruck, Austria (Jun 2007) 145-159
16. McGuinness, D.L., Frank van Harmelen, e.: Web Ontology Language (OWL). W3C recommendation (February 2004)
17. Nikitin, A., Egorov, S., Daraselia, N., Mazo, I.: Pathway studio—the analysis and navigation of molecular networks. Bioinformatics **19** (2003) 2155-2157
18. Nimmagadda, S., Basu, A., Eavenson, M., Han, J., Janik, M., Narra, R., Nimmagadda, K., Sharma, A., Kochut, K.J., Miller, J.A., York, W.S.: GlycoVault: A Bioinformatics Infrastructure for Glycan Pathway Visualization, Analysis and Modeling. Proceedings of the 5th International Conference on Information Technology: New Generations (ITNG'08), Las Vegas, Nevada (April 2008)
19. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M.: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research **27** (1999) 29-34
20. Prud'hommeaux, E., Seaborne, A.: SPARQL: A Query Language for RDF. W3C recommendation (June 2007)
21. Rubio, D.: An Introduction to JSON. http://dev2dev.bea.com/pub/a/2007/02/introduction-json.html (2007)
22. Sahoo, S.S., Thomas, C., Sheth, A., York, W.S., Tartir, S.: Knowledge Modeling and its Application in Life Sciences: A Tale of two Ontologies. World Wide Web Conference [WWW2006], Edinburgh, Scotland (2006)
23. Saraiya, P., North, C., Duca, K.: Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. Information Visualization (2005) 1-15
24. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. **13** (November 2003) 2498 – 2504
25. Thomas, C., Sheth, A.P., York, W.S.: Modular Ontology Design Using Canonical Building Blocks in the Biochemistry Domain. Proc. of the 4th Int. Conference on Formal Ontology in Information Systems (FOIS), Baltimore (November 2006)
26. York, W.S., Kochut, K.J., Miller, J.A., Sahoo, S., Thomas, C., Henson, C.: GLYDE- II – GLYcan structural Data Exchange using Connection Tables. University of Georgia Technical Report (2007)