A Web-Based Resource for Glycome Informatics

Kiyoko F. Aoki-Kinoshita         Masao Ichikawa              Shuichi Ikeda
 kkiyoko@t.soka.ac.jp               e0356101@soka.ac.jp               e0356201@soka.ac.jp

Kouichi Yamada               Takako Yamaguchi
 e0356245@soka.ac.jp                 e0356244@soka.ac.jp

Department of Bioinformatics, Faculty of Engineering, Soka University, 1-236 Tangi-cho, Hachioji, Tokyo 193-8577, Japan

Keywords: glycome informatics, glycan database, web resource

1  Introduction

In recent years, many algorithms and computational models for analyzing the glycome have appeared in the literature, theoretically providing insights into glycan structure and function either for specific diseases or on a comprehensive scale. However, to our knowledge, these methods are not freely available at any particular site for use by biologists. Furthermore, the application of these methods to a particular research project would require the knowledge of an informatician to implement it. Knowing that this is infeasible for many, we present a new resource that centralizes these methods for free use by the glycobiology community, called RINGS (Resource for Informatics of Glycomes at Soka). By putting these tools into practical use, and in turn receiving community feedback, these tools can be further enhanced to make informatics techniques useful for meaningful glycome research. We are dedicated to providing the community with pertinent and valuable tools to advance glycobiology. RINGS is currently available at http://rings.t.soka.ac.jp.

2  RINGS Architecture

The architecture of the RINGS resource is three-tiered. The database is implemented using MySQL on a Linux workstation. A middleware platform provides an extra layer of abstraction such that the data can be uniformly accessed via a library. The Apache web server provides the web service. Web interfaces are implemented such that the data is accessed via the middleware layer.

The RINGS data is organized as illustrated in the very simplified entity-relationship diagram of Figure 1. At the time of this writing, the glycan structure data is based on data from KEGG GLYCAN [1], but RINGS also provides links between glycans and gene/protein sequence data based on reaction information in KEGG. The glycan structures are thus linked with gene/protein data via their relationships that are stored in the Interaction table. Although at the current time the data is limited to simple types of relationships, such as biosynthesis and degradation, one of our goals is to provide more detailed information regarding the relationships between specific proteins and their related glycans and to incorporate data regarding recognition and binding, such as with lectins. We are actively involved in the standardization of an XML format for carbohydrate structures such that such data can be easily transferred from other external databases and incorporated into RINGS to provide relevant information in a concise manner.

Figure 1: Simplified entity-relationship diagram of RINGS.
3 Materials and Methods

The main RINGS page provides entry points to four tools: (1) DrawRINGS, a glycan structure drawing and querying tool, (2) a BLAST server, (3) a prototype of a tool for analyzing microarray expression data of glycosyltransferase genes to predict possible glycan structures in the experiment [2], and (3) a simple interface to translate KEGG Glycan IDs to KCF format. The genes in the BLAST server results are linked to the relevant glycans to which they are related based on the Interaction data. The corresponding 3D protein structure and reaction information can be viewed as well. DrawRINGS allows 2D glycan structures to be inputted using the mouse through a Java applet. The inputted structures can be saved in KCF format and/or used as queries to the RINGS database to retrieve the most similar structures and related protein interaction information. The microarray tool takes as input a list of glycosidic linkages that correspond to that catalyzed by a particular glycosyltransferase and its corresponding expression value. This data is used to compute from the database those glycan structures that consist of those linkages weighted according to expression value. A penalty value can be inputted to give those glycans containing unlisted linkages a lower score in the results. Currently, we are working on an interface to generate and use glycan-structure-based score matrices for improved glycan structure comparisons [3], in addition to more advanced informatics tools such as a sibling-dependent tree Markov model for capturing patterns in the labeled ordered tree structures of glycans [4] and a kernel for classifying glycan structures [5].

4 Discussion

With the advent of a number of publicly available carbohydrate databases such as KEGG GLYCAN, that provided by the Consortium for Functional Glycomics [6] and glycosciences.de [7], there has been an increase in the number of informatics methods for the analysis of glycomes. In order to provide a publicly available web resource for biologists to be able to effectively utilize these methods, we have developed the RINGS resource. Eventually, an XML standard will be developed to allow these database resources to transfer glycan structure data efficiently. RINGS will be able to fully take advantage of this technology to provide glycobiologists with the latest data to analyze and to help spur on the field of glycome informatics.

References