Samir Sur
Genomics and Bioinformatics
Final Project
12/10/2005

Review

Glycomics: an integrated systems approach to structure-function relationships of glycans[1]

The rapid development of tools and technology in the fields of genomics and proteomics has allowed scientists to pursue a vast understanding of gene and protein structure and function. Similarly important advances have been made in the field of bioinformatics in order to annotate and organize the incredible amounts of data produced by genomics and proteomics. Such advances in these fields have also led to the realization that protein post-translational modifications have a greater role in controlling cell phenotypes than originally believed.

Of the various types of post-translational modifications, glycosylation – the attachment of glycans to proteins – is likely the most extensive and significant. There is strong evidence suggesting a role in numerous biological functions including cell growth and development, immune recognition and response, and cell-cell communication (2-4). The versatility of glycan-protein interactions is wide-ranging as they are able to regulate pathways in an analog fashion – fine-tuning biological responses due to the complexity of their interactions. This complexity arises from the fact that glycan interactions are characterized by graded affinity, avidity, and multivalency involving multiple protein contacts at various binding sites (3).

Consequently, glycomics – defined as the integrated approach towards investigating glycans – has been limited in its advancement due to these significant and unique challenges hindering the application of genomics and proteomics methods to the field. However, a concerted effort has been undertaken in this matter and there are currently viable systems for characterizing the structure-function relationships of glycans. Raman et al. have provided a

review of these methods and technologies with an emphasis on combining and organizing the growing sets of data provided by such methods through a bioinformatics platform. Their review is summarized here.

Raman et al. have used the Consortium for Functional Glycomics (CFG) as the basis for their review of the resources and technologies available in the field. It should be noted that CGF is one of several international collaborative efforts initiated to advance the study of glycan structure-function relationships. Their review encompasses four major developments in glycomics methods and discusses the possibility of a bioinformatics platform to integrate the various systems and make accessible the ever-increasing amounts of glycomics data. The technologies reviewed here include: a functional genetics approach to glycomics, the development of glycol-gene microarrays, the characterization of the primary chemical structures of glycans, and the biochemical analysis of glycan-protein interactions.

In order to characterize the biological roles of glycans in mammalian cells by functional genetics, knockouts of genes coding glycosyltransferases have been used extensively. The authors report that CFG has been able to expand the list of genes known to correspond to glycosyltransferases. Knowledge and characterization of these genes has enabled studies in which these genes are knocked out in somatic cells. Thus, this technology has allowed the study of alteration or inhibition of glycosylation on mammalian cellular phenotype.

However, since it is understood that glycans are widely present at the cell-extracellular interface, and are important factors in cell-cell communication, whole-organism functional genetics is required to characterize the full breadth of glycan function in the entire organism (2,3). The authors review specifically knockouts of GLcNAc and GalNAc transferases (enabled by advances in transgenetic technologies) which resulted in severe phenotypes including

embryonic lethality, immune dysfunction, and inflammation deficits. These transferases function directly in glycan biosynthesis and therefore these results are perhaps not unexpected when considering the wide-involvement of glycans in biological processes.

The authors also mention further studies using transgenic mice containing knockouts of later-stage enzymes which are responsible for modifying glycans. Although these studies are potentially very useful in characterizing glycan function, the authors point out the difficulties in assessing the subtle phenotypes caused by such inactivations.

In genomics, the use of gene expression arrays (such as Affymetrix chips) has been instrumental in furthering our understanding of genetic networks and pathways. Similarly, it follows that the expression profiles of glycan-related genes in various tissue and cell types. The authors here emphasize the development of special arrays (based on the Affymetrix chip model) designed specifically to study these genes controlling glycan biosynthesis. Previously, limited representation of these genes in commonly available mouse and human genome arrays and limited sensitivity in measuring subtle expression level changes hindered such techniques.

The authors also review recent advances in determining glycan chemical structure. High-throughput methods are generally less sensitive and this is especially true in characterizing fine structures in a glycan mixture. Here, the authors note, mass spectrometric methods have been used to obtain structural information, however, these methods have been limited to most likely structures present in the mixture. Ascertaining the explicit monosaccharide composition of glycans is made difficult by the heterogeneity of glycans stemming their non-template driven biosynthesis. In order to characterize this fine structure, fragmentation and further characterization of structure is required. These methods are dependent on significant computations in assigning structure based on MS fragmentation patterns and although

advancements have been made, difficulties continue to exist in this area. Raman et al. also mentioned a bioinformatics approach to sequencing glycans. In this approach, data from multiple complimentary techniques are incorporated as such methods use the best set of attributes afforded by the analytical methods in determining glycan sequence.

Finally, the authors reviewed biochemical techniques for analyzing the specificity of glycan-protein interactions. The major example noted here is the development of well-based and printed glycan arrays to which multivalent GBPs are introduced. Prior treatment with an antibody allows for detection and subsequent analysis. Important characteristics of the printed wells include the fact that they better mimic the physiological distribution of glycans on a cell surface – a factor which significantly affects binding to GBPs.

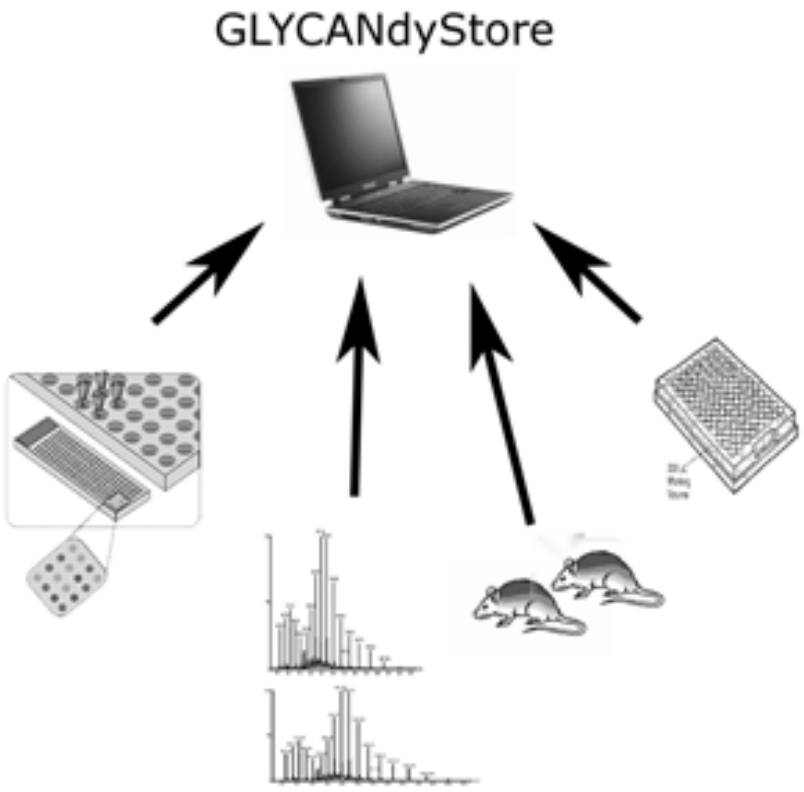

Figure 1| The future of glycomics: schematic of bioinformatics platform to integrate the various datasets generated by glycomics technologies including glyco-gene microarrays, mass spectrometric methods, transgene knockouts, and biochemical analysis. The hypothetical database (called GLYCANdyStore here) stores and integrates data to allow for efficient data mining and large-scale computations.

In the opinion of the authors, the advancement of these methods has created a clear need for a bioinformatics platform capable of integrating, storing, organizing, and

disseminating the various and growing glycomic datasets (Figure 1). Raman et al. support the continuing development of such computational tools and databases as an important future step in glycomics. They envision substantial gains from techniques such as data mining analysis and sequencing. This integration of systems is a vital component of glycomics in the authors' view and on a larger scale, glycomics itself fits into a systems approach to fully comprehending cellular phenotype and its gene and protein function.

Overall, this review provides a brief perspective on the various methodologies in the field of glycomics. Raman et al. present a thorough discussion on the challenges faced in glycomics and applying the technology of genomics and proteomics to the field. In general, this is a broad introduction to the field and its practice; however, it is probably not appropriate as a detailed account of the specific experiments mentioned. Furthermore, the authors make a strong case for the importance of glycans and glycomics in our understanding of biological processes and this review seems well placed as a bulletin for the scientific community at large.

Research Proposal

Although Raman et al. firmly emphasize the progression of glycomics through a bioinformatics platform, advancements in glycomics have created numerous potential applications beyond computational studies (1). As previously mentioned, glycans perform a variety of important functions at the cell-extracellular interface through their multivalent interactions with GBPs. The binding of glycans to proteins is especially interesting because it is due to multiple low affinity interactions which act together to form highly specific binding (3). Although these interactions have hindered attempts to fully characterize their mechanism and structure, the multivalent binding of glycans has shown an exceptional ability to specifically bind

antibodies even from crude human serum. This discovery has important, direct application to medicine in terms of diagnosis of microbial infections, cancer and autoimmune disease (4).

Based on these results, we propose two studies to further our understanding of glycans and their significance in medicine. First, we propose a study to characterize the specificity of glycan-antibody binding by means of a printed covalent glycan array and antibody binding assay. Second, we propose a profile study of glycan biosynthesis genes expression levels in distinct groups (such as cancer patients, or those suffering from autoimmune disease). These experiments would each have twofold importance as they apply glycomics directly to the practice of medicine as well as have the potential to elucidate two relatively underdeveloped aspects of glycomics: the biosynthesis of glycans, and the mechanism of their binding interactions.

Glyco-chips have been recently used in studies profiling the specificity of a diverse range of GBPs (4, 5). Many of these studies have indicated a high specificity in glycan-antibody interactions. In these studies, both specific antibodies in solution as well as more complex mixtures of antibodies (in the form of human serum) were incubated with the glycan array and identified after washing. Results from these experiments have shown specific interactions including the mouse anti-CD15 antibody binding exclusively to Lewis^x glycans (4). We propose a similar study encompassing a larger-scale testing of isolated antibodies and complex samples. This would allow for a more accurate view of glycan binding specificities and would perhaps lead to insights into binding mechanisms for well characterized antibodies.

The procedure for such a study involves creating a covalent glycan array and the subsequent antibody binding assay. The glycan array can be made using standard microarray printing technology currently available (4, 5). The array we propose is similar to the GlycoChip array (Glycominds, Lod, Israel) in which the reducing end of the glycan is bound covalently to

the plastic array by a linker. This flexible linker allows the attached glycan increased specificity and mimics the orientation of glycans in cell surfaces (5). Such arrays may contain upwards of 200 glycans (4).

The antibody binding assay is then performed to profile specificity of glycan-antibody interactions. We propose a simple sandwich procedure for this assay in which the sample antibodies (either from isolated solutions, or more complex mixtures) are incubated with the array. The detection of bound antibodies is then completed by an overlay with labeled specific secondary antibodies. These fluorescent labels are easily detected and their intensities quantified for analysis (4).

This procedure, when used to find and compare antiglycan antibodies from healthy human subjects vs. diseased groups, has the potential to identify important potential drug targets and diagnostic agents for clinical practice. While such studies have noted the potential of such techniques, few, if any, have taken an extensive approach as we have proposed here.

While the identification of human antiglycan antibodies represents an important application of glycomics to the field of medicine, we propose another approach emphasizing the biosynthesis of such glycans as well. In this procedure, a gene-expression profile is created by use of glyco-gene microarrays and profiles can be compared across populations and groups to identify glycan function.

The development of microarray technology has created to ability for researchers to study the expression of thousands of genes in a single experiment. Although these gene chips did not include many of the genes related to the biosynthesis of glycans, advances in identification and annotation of these genes has led to the ability to make glyco-gene chips appropriate for the

profiling in glycomics (6). In recent years, over 400 samples of various tissue and cell types have been analyzed on the CGF glyco-gene microarrays (1).

We propose an experiment in which expression profiles are observed for human cell samples and compared across different population groups (again, such as cancer patients etc.). Expression profiling has been previously used to infer function of uncharacterized ORFs by compiling large databases of such profiles and pattern matching (7). Here we propose a similar study focused on the genes controlling biosynthesis of glycans. This experiment would require a significant number of expression profiles to be generated using the glyco-gene chips; however, we believe the potential for functional identification is great.

Thus, we have proposed two experiments in which advances in glycomics technology are used to apply the growing data concerning glycan structure and function to medicine. Although there is a concerted effort in glycomics to create the types of bioinformatics platforms currently used in proteomics and genomics, we believe that our proposals provide important steps towards incorporating this rapidly growing field into medical science.

References

1. Raman, R. et al. “Glycomics: an integrated systems approach to structure-function relationships of glycans.” Nature Methods 2.11 (2005): 817-824.

2. Lowe, J.B., and Marth, J. D. “A genetic approach to mammalian glycan function.” Annu. Rev. Biochem. 72 (2003): 643-691.

3. Collins, B. E., and Paulson, J.C. “Cell surface biology mediated by low affinity multivalent protein-glycan interactions.” Curr. Opin. Chem. Biol. 8 (2004): 617-625.

4. Blixt, O. et al. “Printed covalent glycan array for ligand profiling of diverse glycan binding proteins.” Proc. Natl. Acad. Sci. 101 (2004): 17033-17038.

5. Schwarz, M. et al. “A new kind of carbohydrate array, its use for profiling antiglycan antibodies, and the discovery of a novel human cellulose-binding antibody.” Glycobiology 13.11 (2003): 749-754.

6. Comelli, E.M. et al. “Custom microarray for glycobiologists: considerations for glycosyltransferase gene expression profiling.” Biochem. Soc. Symp. 69 (2002): 135-142.

7. Hughes, T. R. et al. “Functional Discovery via a Compendium of Expression Profiles.” Cell 102 (2000): 109-126.