Rajini Haraksingh

Genomics and Bioinformatics Fall 2005

Final Paper

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

Rahul Raman *et.al.* (Nature 2005) V.2 pp.817

Introduction

Glycans are multimeric biological sugars composed of monosaccharide subunits held together by glycosidic linkages. Glycans are a subset of the essential and most abundant class of biological molecules called carbohydrates. In contrast to proteins and nucleic acids which only form linear structures, glycans form both linear and branched polymers. This stems from the ability of ringed monosaccharides to form glycosidic linkages at any of their hydroxyl groups. However, most branching is well-defined rather than haphazard and this allows glycans to be classified by their branching. In addition to forming carbohydrates, glycans can be covalently conjugated to proteins or lipids to form glycoproteins or glycolipids. Branched glycans are attached to proteins and lipids by a process called glycosylation. This process has two forms, N-linked (or addition of a glycan to an Asn residue in a polypeptide) or O-linked glycosylation (addition of a glycan to a Ser/Thr residue in a polypeptide).

The relatively limited number of genes comprising the genome of an organism, even one as complex as human, and the extent of diversity of biological molecules indicates that there is not a one-to-one correspondence of genes to molecules. Rather, there must be extensive post translational modification of proteins in order to achieve the magnitude of biological diversity in a cell. Glycosylation which occurs in the endoplasmic reticulum and Golgi apparatus is perhaps the most extensive form of post translational modification of proteins. Glycans are ubiquitously present at the cell surface where they are thought to modulate protein interactions at the cell-extra-cellular interface. The current belief is that glycans play essential roles in cell growth, development, tumour growth, metastasis, anti-coagulation, recognition in the immune response, cell-cell communication and microbial pathogenesis.

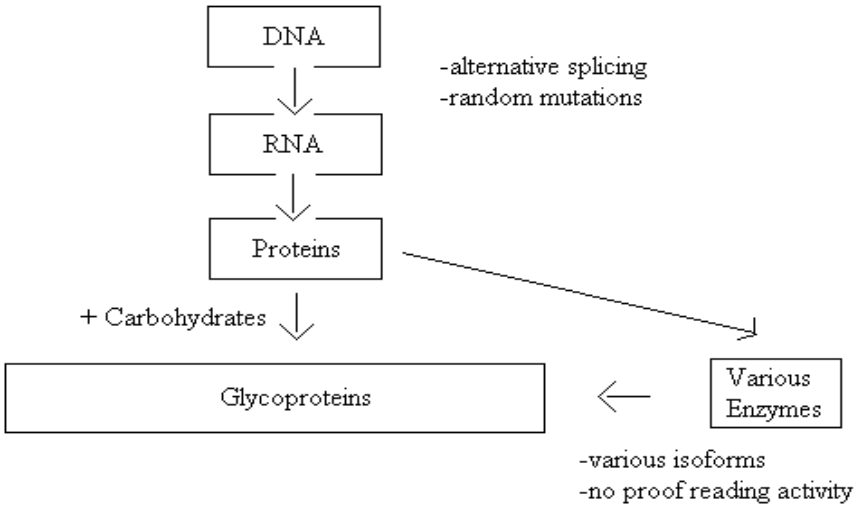

Fig. 1. The levels of diversity of molecules in a cell. This figure shows the flow of biological molecules in the biosynthesis of macromolecules in the cell. The sizes of the boxes indicate the relative amounts of diversity within a group of macromolecules. The main determinants of diversity and where they occur are written next to the arrows.

The study of glycans and their interactions with other molecules has been challenging due to three fundamental properties of these molecules. Firstly, the biosynthesis of glycans, unlike that of proteins or nucleic acids, is non-template driven, requires the action of several glycosyltransferases which may have several cellular isoforms and lacks proof reading activity. These aspects of the biosynthesis give rise to the inherent heterogeneity of this class of molecules (Fig.1). Thus the use of functional genetics to study these molecules is very limited as most of the diversity occurs at the post translational level. Secondly, analytical techniques based on the physical and chemical properties of these molecules have been limited due to the chemical heterogeneity of glycans. Thus, isolating homogenous populations of glycans for structural studies has been very difficult. Finally, it has been a major challenge to understand the biochemistry of the protein-glycan interactions owing to the multivalency and graded affinity of this binding.

The study of this class of biological molecules has presented unique challenges that revealed the need for a systems approach to assimilate information from the molecular to organismal level. Thus, glycomics or a systems approach to the study of glycans was born. It is not known how long this concept has been around. However the word glycomics was first used

about five years ago and follows the naming convention of genomics (the study of the genome) and proteomics (the study of the proteome). Several international collaborative efforts have been established with the aim of studying glycan structure-function relationships in an integrated manner. Some of the larger ones are the Consortium for Functional Glycomis (CFG), EuroCarb and the Japanese Consortium for Glycomics.

Summary

This paper uses CFG as a model to review the current technologies for structural characterization of glycans and biochemical analysis of glycan-protein interactions. This review focuses mainly on the datasets involved and how they are interrelated and integrated to allow glycomics. This paper also discusses briefly the development of a bioinformatics platform to integrate the diverse datasets generated and thus provide a systems framework for glycomics. Transgenic mice strains containing knock outs of several glycan biosynthetic enzymes have been established. These strains are being used to study how genetics affects the phenotype of the whole organism. The chemical diversity of glycans and the complexities of the whole organism phenotypes of knock out mice highlight the need to couple functional genetic studies with quantitative measurements of gene expression of glycan biosynthesis enzymes and their binding proteins. The issues discussed above provided the motivation for the development of glyco-gene microarrays. However, there have been challenges in using these genome-wide arrays to investigate the dynamics of the glycan-protein interaction. The most important of these is the limited sensitivity of gene expression data relative to downstream post translational events. However, an Affymetrix array-based glyco-gene microarray has been designed to overcome some of the challenges. These have been used to provide information on glycan synthesis

enzyme expression in knockout mice strains and this information has been correlated to the actual glycan structures known to be present in a given sample.

Fine structure characterization of glycans and glycan-glycan binding protein (GBP) interactions is another aim of glycomics. However, isolation of homogeneous samples of glycans and of glycan-GBP complexes is difficult. Nevertheless, Mass spectrometric (MS) methods have been used to obtain high throughput (albeit less sensitive) mass profiles of glycans from entire cell and tissue samples. In addition Matrix-assisted laser desorption/ionization (MALDI)-MS and its recent automation and improvements are able to provide a snapshot as to the most likely structures of glycans present in a given tissue. High performance liquid chromatography has also been used to characterize the glycans present in a mixture. One of the most sensitive advances in fine structure techniques has been the use of Fourier transform ion cyclotron mass spectrometry. Nuclear Magnetic Resonance (NMR) is also a powerful tool that can provide quantitative data on distinct monosaccharides in a mixture.

Hundreds of glycan structures have been artificially synthesized in order to capture the diversity of biological glycans. These are used mainly to assess the relative binding affinities of various glycan binding proteins in chip based arrays and competitive assays.

The major motivation for a bioinformatics platform for glycomics is the need for integration of information in multiple datasets (glycan binding protein databases, glycan biosynthetic enzyme databases, glycan structure databases etc.) to facilitate glycomics. These datasets are further organized into various levels of complex inter-relationships. This complexity must be hidden from the user during data acquisition and dissemination. It was found that a three-tier software architecture that consists of a back-end relational database which stores data and its relations, a middle layer that relays information between the database and the user

interface and a front layer comprising the user interface, is the most appropriate for this function. Several other aspects of the bioinformatics platform are discussed in less detail such as the ability to link orthogonal data sets derived from similar samples, the molecule page interface and the ability of the platform to support computational tools needed to mind the data.

This paper gives a good overview of the current state of glycomics for someone with limited knowledge of the field. The background is relevant and emphasizes the motivation for the advancements in this field. I think this is an appropriate time in the field for such a review to be published as the establishment of several international consortiums to investigate glycomics puts us on the brink of a rapid expansion of the knowledge of this field.

While the level of detail and length are appropriate for a review there are some points I felt that were not well developed. Firstly, the discussion on assessing whole organism phenotypes on knocking out one particular glycan biosynthetic gene was unconvincing. The authors listed a variety of unrelated phenotypes that occurred on knocking out a particular gene. These seem too diverse to make any conclusions. While some of the phenotypes can be sorted though by knocking out later-stage enzymes as discussed, I felt that perhaps the use of a simpler organism, such as yeast would give more insight. Yeast have 50% of their genome conserved with humans and are already a well established organism for studying eukaryotic glycosylation. A unicellular organism may provide a less complicated variety of phenotypes that may help to sort through the many different phenotypes observed in a system as complicated as a mouse.

Another issue that worried me was the that MALDI-MS was able to provide a set of the most likely glycan structures when annotated by an expert. This statement makes me question the objectivity and reproducibility of the results of such an analysis and makes me doubt the usefulness of this method for determining the glycan composition of a sample. However, the

authors did imply that while each method for fine structure determination was not very sensitive, a combination of several methods did provide a reasonably accurate glycan profile of a sample.

Finally, this paper promises much hope for this field. I got the impression that we are on the brink of a biological revolution as we now have the tools to begin to understand the glycome.

Proposal

Three of the main future directions for the field of glycomics are as follows: to explain the regulation of glycan diversity bases on its biosynthesis, to understand the basis for specificity in glycan-protein interactions and to discern how glycans on the cell surface mediate cell-cell interactions and extracellular signalling via multivalent interactions with proteins. I would like to extend the work presented in this paper in the direction of understanding the basis for specificity in glycan-protein interactions. I believe that such interactions are the core of the functional significance of glycans, more so than the generation of diversity or the larger scale cell-cell interactions. A thorough understanding of the biochemical basis for specific glycan-protein interactions can be directly extended into areas such as rational drug design, vaccine development and understanding host-pathogen interactions.

The first step in understanding the specificity of glycan-protein interactions is to determine the parts of these molecules that are necessary for the interaction and to be able to distinguish these regions from those that have no bearing on the interaction. This will be the focus of my proposal.

The main aim of my proposal is to develop a screening method to identify similar motifs in otherwise distinct glycans present in humans, to which a particular human glycan binding protein (GBP) binds. By identifying such regions of conserved binding in otherwise dissimilar glycans, one should be able to determine the structural or biochemical basis for the binding

interaction. This experiment is not concerned with nor does it depend on the type of "similarity" in the putative conserved binding motifs. That is to say, these conserved binding motifs can be conserved in terms of the actual atomic composition of the molecule at that region or in terms of the three dimensional shape of the molecule at that region. This experiment does not seek to distinguish between these two cases, at least not initially.

In order to test for such putative conserved GBP binding motifs, a large scale comparison of the interaction of a particular GBP with many different glycans is essential. Thus, an array style test is best suited for such a high-throughput screen. I will make use of the new developments in glycan synthesis techniques and glycan chip arrays. I plan to design a solid phase, NHS-activated glass slide printed with a wide variety of glycans of equal concentrations at each locus. Individual chips will be washed with a homogenous sample of a particular glycan binding protein. After extensive washing to decrease non-specific binding, the chip will be incubated with an anti-GBP antibody that is conjugated to a fluorophore. The positions on the chip where the GBP is bound will then be determined by the positions where there is fluorescence. The use of a fluorescent antibody against the GBP instead of simply a fluorescent GBP is for two reasons. Firstly, since these GBPs are unlikely to have been studied extensively, it is unknown whether conjugation of a fluorophore will interfere with the binding. Secondly, the use of a fluorescent antibody provides an amplified signal than the signal that would be obtained by just the GBP as more than one antibody molecule can potentially bind to the GBP.

I plan to proceed with the analysis of the results of such an experiment as follows. The biochemical and three dimensional structures of all glycans to which a particular GBP binds, will be determined using a combination of already established methods such as NMR spectroscopy and FT-ICR-MS. A combination of methods will be used to increase the sensitivity and accuracy

of the fine structure information determined. Comparisons of these structures will be carried out to determine areas of conserved three dimensional structure or areas of conserved atomic structure. Such areas, if found, would be hypothesized as the site at which the GBP binds. In order to test this hypothesis, such areas would be "mutated" by synthesizing similar glycans that differ from the putative targets in only the hypothesized site of GBP interaction. GBP binding assays would then be repeated to determine the extent of binding in the 'mutant' version of the glycan compared to the 'normal' version. Differential binding to these two versions of a particular glycan by a particular GBP will be good evidence that we have identified the binding site of the protein.

This work can further be extended by doing the converse experiment. That is, one can take an already established GBP with a known binding target such as the mammalian Toll-like receptor 4 (TLR4) which binds to lipopolysaccharide (LPS) on the surface of gram negative bacteria. By making a mutant library of TLR4, and this time attaching the mutant proteins to the chip and washing different glycans over, we can determine which mutants bind which particular glycans. We can choose to analyse mutants that bind to different glycans and no longer bind to LPS. By analyzing the effects of the mutation on the primary, secondary and tertiary structure of the GBP (and comparing to normal TLR4) we can hope to find a basis for the part of the GBP that is important in binding to the glycan.

I believe that determining the regions of glycans and GBPs that are essential for the interaction between the two is of utmost importance in understanding the specificity of the interaction. I think that a screen such as the one I describe above provides a good way to make an educated start in such a daunting project.

References:

1. Raman, R. *et.al.* Glycomis: an integrated systems approach to structure-function relationships of glycans. *Nature Methods* Vol2.No.11 (2005).

2. Varki, A. *et.al. Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, New York, 1999).

3. Raman, R. *et.al.* Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem. Biol.* 12, 267-277 (2005).

4. Rudd, P.M. *et.al.* Glycosylation and the immune system. *Science* 291, 2370-2376 (2001).

5. Voet, D. and Voet, J. *Biochemisty* (Wiley, 1995).

6. Medzhitov, R. Toll-like receptors and innate immunity. *Nature Reviews Immunology* 1:135-145 (2001).

Websites:

1. Consortium for Functional Genomics

http://www.functionalglycomics.org/static/consortium/main.shtml

2. World Wide Words

http://www.worldwidewords.org/turnsofphrase/tp-gly1.htm