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

Raman *et al*., *Nature Methods*, 2, 817-821 (2005).

Part I

There are three main classes of repeating biopolymers: nucleic acids, proteins and carbohydrates (glycans). DNA is the blueprint of life. It contains the information that is transferred from one generation to the next and acts as a template for the synthesis of RNA. RNA in turn mediates the production of proteins, molecules that form structural components and catalyse most of the reactions in living cells. The linearity of protein and nucleic-acid polymers and the regularity of the bonds joining each monomeric unit have allowed the development of efficient and reliable tools for their analysis and synthesis.

Obtaining similar tools for glycans has been hampered by their structural and chemical complexity. The synthesis of carbohydrates is non-template based and it is dependent on the expression of one class of proteins, glycosyltransferases, which carry out the synthesis of carbohydrates. Each monosaccharide has several free hydroxyl groups that can be used to link the unit to the next monosaccharide. This allows branching and increases the number of possible polysaccharide structures. Because of their enourmous heterogeneity, the classical genetical and biochemical tools to assess phenotypic variations are of little use. Synthetically, glycans are challenging to work with

because many more functional groups have to be protected to get one specific group to react, and the stereochemistry of every new glycosidic linkage needs to be controlled.

The protein-linked glycans represent the most common form of post-translational modification. In that form, complex glycans are involved in biochemical pathways and regulate biological processes ranging from development, to coagulation, infection by bacterial and viral agents. Complex glycans act at many levels: from a multicellular level, at the interface between cells, tissues and organs, to coordinate biological processes.

Due to the glycan's importance (pharmacological as well) in many biological processes and to the fact that they are so challenging to work with, an integrated, systematic approach to understanding the glycan structure-function relationships was deemed necessary. This initiative was pursued through the establishment of several international collaborative efforts for functional glycomics development: CFG, Eurocarb, Japanese Consortium for Glycomics etc.

The starting point was to gain insight into the biosynthesis, structure and function of complex glycans. Through cloning the enzymes that are involved in the biosynthesis of complex glycans, and developing an understanding of their mechanism of action and substrate specificity, an understanding of the range of complex glycan structures that are possible was obtained. In addition, numerous gene-knockout experiments were performed. Loss-of-function studies using genetic knockouts have provided important structure–function correlates for both branched glycans and linear polysaccharides. The knock-out experiments revealed interesting phenotypes that subsequently shed insight into the specific glycan sequences mediating biological processes at the cell surface. On

the other hand, in order to really understand the structure-function relationship for specific glycans, it was necessary to correlate the whole-organism phenotype experiments with gene expression high-throughput assays for glycan biosynthetic genes and their binding partners (and to apply these measurements to the glycan structures present on a specific cell/ tissue type).

Along these lines glycogene microarrays were designed in order to analyze the expression patterns for biosynthetic and catabolic enzymes that are involved in glycan construction and elaboration; furthermore, printed glycan arrays were also developed for the detection of the interaction of carbohydrate-binding proteins with certain glycans. With respect to these interactions, there was a practical need to map the plea of glycan structures present on a certain tissue/cell surface and also to characterize the glycan-GBP interactions at the primary structure level i.e. to achieve a fine-structure characterization of the multivalent binding of glycans with the GBP on the cell surface.

There has been a large increase in the number of techniques that have been successfully applied to the analysis of complex glycans and glycoconjugates, including mass spectrometric (MS) and capillary electrophoretic (CE) techniques. Many of these technologies have distinct advantages compared with traditional analytical methodologies, including the ability to analyze minute amounts of biologically based material. Furthermore, HPLC-based techniques have been successfully used to obtain a profile of glycans in a mixture based on their elution profile.

To handle the structural complexity and heterogeneity of complex glycans, analytical techniques that combine aspects of separation, analytical measurement and

data integration through bioinformatics were developed. Examples of such technologies include capillary liquid chromatography (LC)-MS, CE-MS and LC-NMR techniques. Integration into a bioinformatics framework was performed through chip-based separation and analysis technology; recently, a Fourier based MS approach (FT-ICR-MS) was used to characterize glycans including those which are a part of glycolipids.

In conclusion what should the approach be to characterize such complex biological mixtures? The solution seems to lie in the integration of experimentally derived data sets using a bioinformatics framework. In other words, the development of strategies that would bring together multiple datasets may shed light onto the structure-function relationships of glycans.

A single analytical tool, no matter how sophisticated, is insufficient to completely characterize biological-relevant glycan mixtures. Several types of measurement that yield complementary information must be integrated to completely characterize complex glycan mixtures and yield important structural correlates to biological functions. These methods generate different types of data sets, which often prove complementary to one another. Therefore, to complete a detailed characterization of complex glycans, it is necessary to integrate the diverse and orthogonal experimental measurements that are generated by many different methods. The implementation of data-integration methodologies may provide a more complete description of the structure–function relationships of complex glycans, as well as clarifying the biochemical pathways that are required to elicit specific responses. As the authors point out, a current undertaking by the

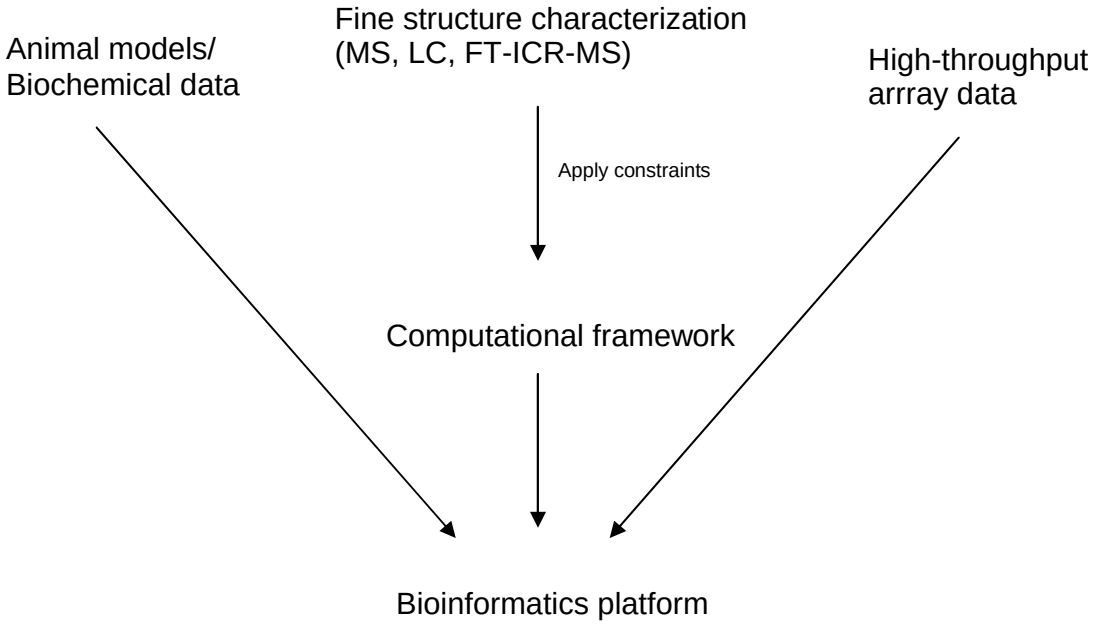
Consortium for Functional Glycomics is to understand the role of carbohydrate–protein interactions using such a strategy.

Comments/Questions:

Integrating the animal model's data (obtained using genetic approaches e.g. knock-ins, knock-outs, conditional mutants) and biochemical/cell biological data (localization of particular glycan-GBP interactions, RNAi knockdowns of glucosyltransferase genes and their effect on the cellular phenotype in question e.g. cell adhesion, trafficking, etc) with the high-throughput data obtained from glyco-gene microarrays and the fine-structure characterization of glycans (MS, LC, CE, etc) will conceivably help develop a structure–function relationship model for complex glycans This can be very useful when we think of the huge potential of these molecules as therapeutics and diagnostic agents (a plethora of glycan-based drugs are already being used to treat various afflictions ranging from Gaucher's disease to thrombosis and anemia). Unquestionably, a thorough understanding of glycan structure and function will lead to a new generation of highly effective therapeutics.

On the other hand, the time-scale for all these developments might be quite long. The high-throughput data generated through lectin and glycan arrays is very important because of the large number of biologically-relevant glycan-GDP interactions and so far, the number of glycans that can be printed onto the NHS-activated glass slides is very small. Furthermore, all these interaction need to be confirmed biochemically through *in*

vivo/ *in vitro* assays and compared to their proteomic counterparts, these interaction are more time consuming and harder to assess. Furthermore, the heterogeneity of glycans and the potential for multiple functions for a particular glycan isoform and vice-versa will make it harder to pinpoint pathologically relevant interactions. All in all though, the perspectives for glycomics to become a valuable part of the other "omes" are bright.



Part II

One avenue that could be explored through the methods presented in the Raman *et al*. study involves the use of glycan and lectin arrays to look for therapeutic targets in glycan-related genetic disorders.

The glycosphingolipid lysosomal storage diseases are a family of human metabolic diseases that, in their severest forms, cause death in early infancy as a result of progressive neurodegeneration. They are caused by mutations in the genes encoding the glycohydrolases or the activator proteins that catabolise glycosphingolipids (GSLs) within the lysosomes. In these diseases the GSL substrate of the defective enzyme accumulates in the lysosome, where it is stored and leads to cellular dysfunction and disease. The therapeutic options for treating these diseases have relied on strategies for augmenting enzyme concentrations to compensate for the underlying defect. These strategies include bone-marrow transplantation, enzyme-replacement therapy and gene therapy (1-3).

Gaucher disease is the commonest among the GSL lysososmal storage diseases, and type 1 occurs at a high frequency in Ashkenazi Jews (4). Type 1 Gaucher disease is caused by mutations in the gene coding for the enzyme glucocerebrosidase and the resultant storage of GlcCer. The cell type that is most affected by the defect is the macrophage because these cells accumulate GlcCer from the cells that they ingest as part of their normal phagocytic activity. A series of enzymes — including imiglucerase (Cerezyme; Genzyme) and larodinase (Aldurazyme; Genzyme) — that degrade or alter

complex glycan structures have been successful in replacement therapy for individuals with glycan-related genetic disorders (5-7). Importantly, the development of the available drugs in this class required a detailed structural/functional understanding of their mechanism of action, their effect on glycan structure and the role of specific glycan structures in the disease of interest.

Theoretically, an alternative strategy that could be applied to treat type 1 Gaucher's disease is substrate deprivation. In this way, there would be a balance between the rate of GSL biosynthesis and the (impaired) rate of GSL catabolism, thus preventing the accumulation of GSL (8,9). To date, two main classes of compounds that inhibit the ceramide-specific glucosyltransferase have been identified. This transferase catalyzes the transfer of glucose from the nucleotide sugar UDP-Glc to ceramide. The first class of compounds act as reversible, mixed-type inhibitors of ceramide (9). The second class comprises the N-alkylated derivatives of deoxynojirimycin which act as glycosyl competitors (blocking the binding of the transferase to the UDP-Glc) (10). Both classes of inhibitors have a Ki in the micro to milimolar range which is far from satisfactory to make them clinically relevant compounds. This clearly shows the need for more potent binding partners/inhibitors of the ceramide-specific Glc-transferase.

This is where a high-throughput glycomics approach could prove fruitful. Complex glycans that are isolated from natural sources or are chemically synthesized could prove themselves useful both biologically and pharmaceutically. The first approach involves the *de novo* synthesis of complex glycans by chemical means. This tactic has successfully been used to generate a synthetic version of a truncated heparin

oligosaccharide for thrombotic indications (11). Alternatively, solid-phase synthetic procedures could also be used. Another strategy could use the isolation of natural glycans, chemical modification and/or degradation of the backbone structure, followed by purification. This approach is in wider use owing to the ease of scale-up, as well as its ability to take advantage of the natural structural diversity of known complex glycans (12).

There are several other ways to create different glycan populations. One of them involves the incorporation of non-natural amino acids in the peptide backbone; these amino acids contain either a reactive chemical handle or a *N*-acetylglucosamine-modified amino acid that allows the creation of N-linked polysaccharide chains through chemical and/or enzymatic routes (13). An alternative metabolic-engineering approach that has the potential to work both *in vitro* and *in vivo* has been reported, in which defined non-natural monosaccharides are introduced into a cell, where they are incorporated into natural glycans through the action of endogenous polysaccharide synthetases (14, 15). After incorporation, these monosaccharides can be detected through unique chemical handles.

The point here is to maximize the glycan ligand probes for a high-affinity binding assay. In other words, the utility for a large number of glyconjugates would be to find and design inhibitors to the physiological interaction between D-glucosyltransferase and UDP-glucose N-acylsphingosine. There may be other glycan targets that the glucosyltransferase can bind to and depending on the affinity, those interactions might be very relevant.

After the synthesis step, one could use a printed glycan array for the detection of the specificity of the glucosyltransferase to various glycans. The procedure would follow the one described by Raman *et al*. where the glycan binding protein is introduced into the array upon treatment with a primary Ab and then detection would be achieved through a fluorophore-linked secondary Ab. The molarity of the GBP would be also have to be varied to promote high-affinity binding.

After the identification of the potential binding partners of the ceramide specific glucosyltransferase, the interaction would have to be tested through several *in vitro* and *in vivo* assays e.g. an *in vitro* assay for the glycosylation of ceramide in the presence of a fluorigenic substrate of the Cer-specific glucosyltransferase, Cer-specific Glctransferase and the newly found GBP. *In vivo* assays can also be used since antibodies raised against GlcCer have already been characterized; cells grown in the presence of the "inhibitory" GBP would be tested for the presence of the glycosylated form of ceramide.

Glycans, either alone or as glycoconjugates, have great potential for use as drugs, for several practical reasons. Firstly, complex glycans are relatively small and intrinsically more stable than protein-based drugs. Second, they are more easily formulated for drug delivery and lastly, sugars are highly specific and potentially less immunogenic than other natural products, such as proteins or RNA-based strategies (16). As Raman *et al*. point out, the large-scale production of specific biologically active glycans is quite demanding because of their intrinsic non-template biosynthesis. However, once the manufacturing methods are established, a wide range of complex polysaccharide biotherapeutics could be readily available.

References

1. Lloyd, K.O. and Furukawa, K. (1998) Biosynthesis and functions of gangliosides: recent advances. *Glycoconj J* 15, 627-636.
2. Van Echten, G. and Sandhoff, K. (1993) Ganglioside metabolism. Enzymology, Topology, and regulation. *J Biol Chem* 268, 5341-5344.
3. Ichikawa, S. and Hirabayashi, Y. (1998) Glucosylceramide synthase and glycosphingolipid synthesis. *Trends Cell Biol* 8, 198-202.
4. Beutler, E. and Grabowski, G. (1995) Gaucher Disease, McGraw Hill, New York, USA
5. Barton, N.W. et al. (1990) Therapeutic response to intravenous infusions of glucocerebrosidase in a patient with Gaucher disease. *Proc Natl Acad Sci U. S. A* 87, 1913-1916.
6. Cox, T.M. (1994) Therapeutic advances in Gaucher's disease: a model for the treatment of lysosomal storage diseases. *Trends Exp Clin Med* 4, 144-160.
7. Grabowski, G.A. et al. (1995) Enzyme therapy in type 1 Gaucher disease: comparative efficacy of mannose-terminated glucocerebrosidase from natural and recombinant sources. *Ann Intern Med* 122, 33-39.
8. Radin, N.S. (1996) Treatment of Gaucher disease with an enzyme inhibitor. *Glycoconj J* 13, 153-157.
9. Vunnam, R.R. and Radin, N.S. (1980) Analogs of ceramide that inhibit glucocerebroside synthetase in mouse brain. *Chem Phys Lipids* 26, 265-278.

10. Platt, F.M. and Butters, T.D. (1998) New therapeutic prospects for the glycosphingolipid lysosomal storage diseases. *Biochem Pharmacol* 56, 421-430.
11. Schofield, L., Hewitt, M. C., Evans, K., Siomos, M. A. & Seeberger, P. H. Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. *Nature* 418, 785−789 (2002).
12. Linhardt, R. J. & Gunay, N. S. Production and chemical processing of low molecular weight heparins. *Semin. Thromb. Hemost.* 25 (Suppl. 3), 5−16 (1999).
13. Zhang, Z. *et al*. A new strategy for the synthesis of glycoproteins. *Science*, 303, 371−373 (2004).
14. Wieser, J. R., Heisner, A., Stehling, P., Oesch, F. & Reutter, W. *In vivo* modulated *N*-acyl side chain of *N*-acetylneuraminic acid modulates the cell contact-dependent inhibition of growth. *FEBS Lett.* 395, 170−173 (1996).
15. Kayser, H. *et al*. Biosynthesis of a nonphysiological sialic acid in different rat organs, using *N*-propanoyl-D-hexosamines as precursors. *J. Biol. Chem.* 267, 16934−16938 (1992).
16. Sioud, M. Therapeutic siRNAs. *Trends Pharmacol. Sci.* 25, 22−28 (2004).