

GLYCOMICS: A PATHWAY TO A CLASS OF NEW AND IMPROVED THERAPEUTICS

Zachary Shriver*, S. Raguram[‡] and Ram Sasisekharan[§]

Abstract | Complex glycans that are located at the surface of cells, deposited in the extracellular matrix and attached to soluble signalling molecules have a crucial role in the phenotypic expression of cellular genotypes. However, owing to their structural complexity and some redundancy in terms of structures that elicit a function, the therapeutic potential of complex glycans has not been well exploited, with a few notable exceptions. This review outlines recent advances that promise to increase our ability to use complex glycans as therapeutics. Opportunities for the development of further structure–function relationships for these complex molecules are also discussed.

GLYCANS

Endogenous biomolecules that consist of monosaccharides that are *O*-linked to one another. Glycans can either be branched, in which several glycosidic linkages extend from a single monosaccharide, or linear, in which monosaccharides are linked to one another end-to-end.

**Momenta Pharmaceuticals, 43 Moulton Street, Cambridge, Massachusetts 02138, USA.*

[‡]*Parivid Corporation, One Broadway Suite 600, Cambridge, Massachusetts 02139, USA.*

[§]*Biological Engineering Division, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.*

Correspondence to R.S. e-mail: rams@mit.edu
doi:10.1038/nrd1521

Complex GLYCANS that are attached to proteins and lipids coat all eukaryotic cells, and are also found in the extracellular space between cells¹. In simple terms, glycans can be divided into two categories: the linear sugars or polysaccharides, which consist of repeating pyranose monosaccharide rings; and the branched sugars, which are saccharide structures that are based on multiple linkages between monosaccharide rings (FIG. 1). Through the post-translational modification of proteins and lipids, and non-covalent interactions with proteins, complex glycans regulate biochemical pathways and, accordingly, impinge on diverse biological processes ranging from development^{2,3} to coagulation⁴ to infection by bacterial and viral agents⁵. Complex glycans act at a multicellular level, at the interface between cells, tissues and organs, to coordinate biological processes⁶. Therefore, from a biological perspective, complex glycans represent a promising, but relatively untapped, source for the development of new pharmaceutical agents.

The discovery of detailed structure–activity relationships for DNA and proteins has led to an in-depth understanding of how genotype, in part, controls phenotype. Furthermore, this biological and biochemical knowledge has been successfully translated into drug discovery efforts. Advances in our understanding of the biosynthesis, structure and function of complex glycans

promise to allow us access to this important class of molecules to modulate disease processes. Several reviews and scientific investigations have been published in recent years describing important findings in the basic biology of complex glycans^{2,6–8}. Here, we begin by outlining some of the most notable findings, to illustrate the advances that have been made, before moving on to discuss the potential of complex glycans as therapeutics.

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^{1,9,10}, we now have an understanding of the range of complex glycan structures that are possible. In addition, numerous gene-knockout experiments have increased our understanding by examining, at an organismal level, the effect of impinging on the biosynthesis and/or catabolism of complex glycans. Loss-of-function studies using genetic knockouts have provided important structure–function correlates for both branched glycans and linear polysaccharides^{9,11}. For example, the conversion of branched sugars from one type (high mannose) to hybrid and complex structures is initiated by the action of the enzyme GlcNAcT-I glycosyltransferase, which is encoded by the *MGAT1* gene^{12,13}. The conversion of hybrid to complex structures is controlled subsequently by the *MGAT2* gene, which encodes GlcNAcT-II glycosyltransferase¹⁴. Mutation of *MGAT1*

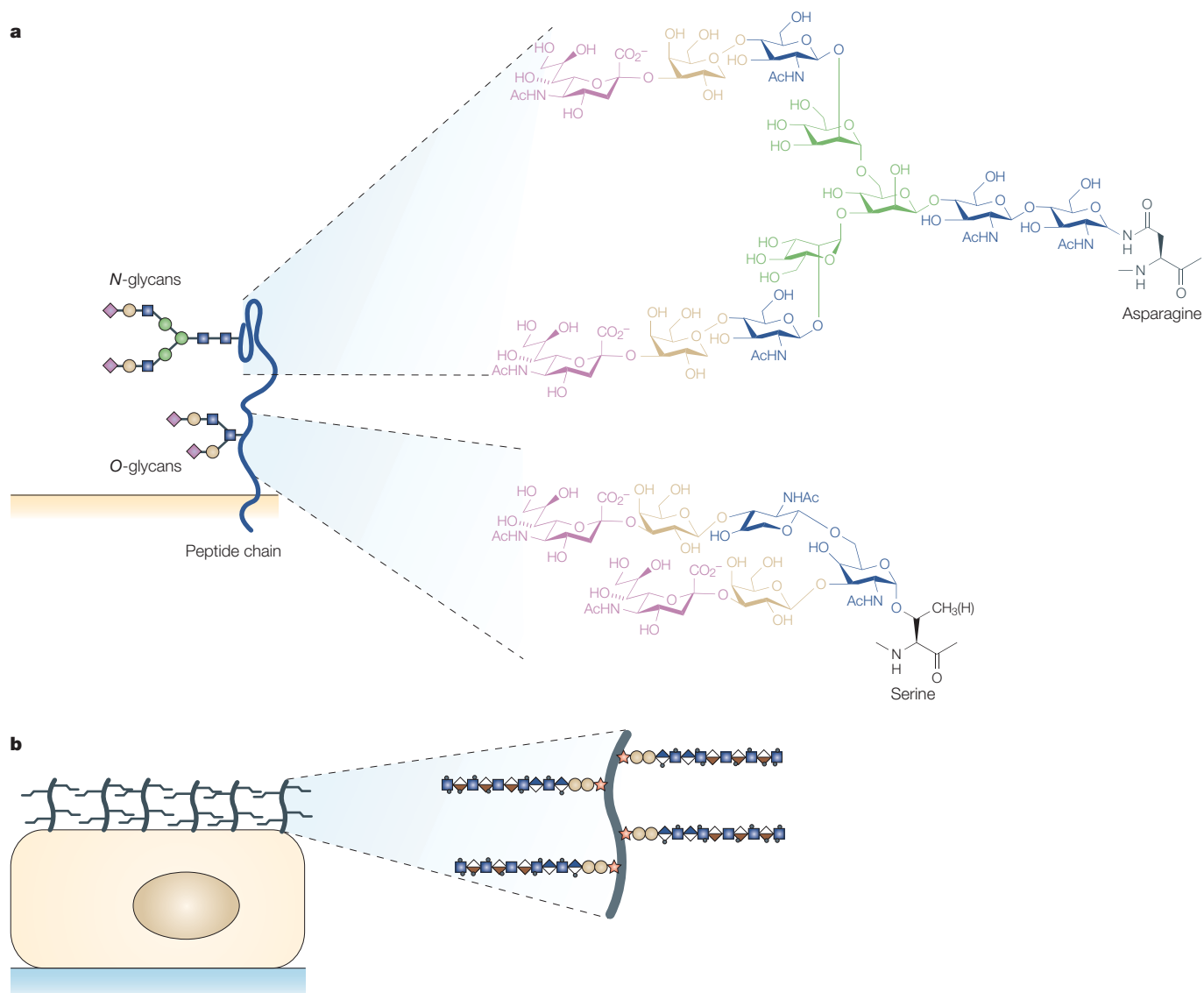


Figure 1 | Biological roles of branched and linear glycans. a | Branched glycans that are attached to proteins, lipids and other biomolecules as glycoconjugates represent the most extensive and diverse form of post-translational modification. Using proteins as an example, glycans are either *N*-linked through the amino acid asparagine, or *O*-linked through the amino acids serine or threonine, to the peptide backbone. The structure and conformation of branched glycans are important in their function; they affect protein binding, targeting, folding and stability, and specificity. **b** | Linear polysaccharides are found on the cell surface and the extracellular matrix. Many linear polysaccharides are attached to a protein core, although some are free. Depending on their location and sequence, linear polysaccharides have several biological functions, including storing signalling molecules, mediating cell-surface binding and signalling events, and directing intracellular-transport events. Key: purple diamond, sialic acid; beige circle, galactose; blue square, *N*-acetylglucosamine; green circle, mannose; orange star, xylose; half-coloured diamonds, acidic sugars; black dots, *O*- or *N*-sulphation.

SIALYLATED

Sialic acid is an important 'capping' monosaccharide that imparts essential structural and biological information. For recombinant proteins, the presence of sialic-acid capping is important in the half-life and stability of the protein. Sialic-acid capping also has a role in a range of biological processes, including immune-system functioning.

results in embryonic lethality³. Conversely, mutation of *MGAT2* results in a range of abnormalities that are illustrative of the role of complex branched sugars in various processes¹⁵. These and other studies have led to the realization that complex glycan structure is modulated by the expression and levels of specific glycosyltransferases acting in conjunction with one another.

Advances have also been made in terms of identifying the protein partners that bind to complex glycans at the surface of cells and in the extracellular matrix. Examples include the binding of galectins with β -galactosides, the binding of sialyl Lewis X structures with selectins,

the interaction of sialylated structures with the sialic-acid-binding immunoglobulin-like lectin (siglec) family of proteins, and the binding of specific heparin glycosaminoglycan sequences to proteins, such as antithrombin III (*ATIII*) and fibroblast growth factor. In many cases, knowledge of the co-crystal structures of glycans with their protein partners has led to an in-depth understanding of the specific contacts that are responsible for high-affinity binding^{4,16–18}. However, although our understanding of the structure of complex glycans has markedly increased, our ability to harness them in drug development has yet to reach fruition.

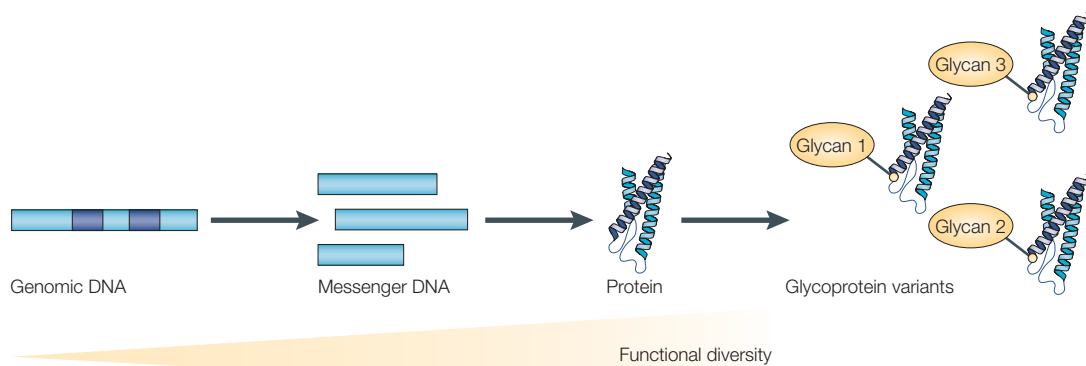


Figure 2 | **Structural complexity and information content of glycans.** Glycans amplify the genomic information content of cells. The recent sequencing of the human genome yielded the surprising result that it contained a smaller number of genes than originally thought. After the translation of RNA to protein, a huge amplification of function occurs through post-translational modification by glycan addition. In this case, a single polypeptide chain can yield many glycoforms with differing properties, thereby providing functional diversity even in a single gene (see the main text for an example of how glycosylation can drastically affect function).

Drug development challenges with glycans

The problem of how to exploit the vast potential of complex glycans in drug development involves overcoming fundamental challenges to clarify important structure–activity relationships. First, experimental procedures are needed that can accurately and reproducibly measure the structure or sequence of complex glycans. Second, there needs to be a framework for describing glycan sequences in terms of the properties that lead to their functions (FIG. 2). The integration of these two approaches is especially important in the study of complex glycans as, in many cases, there seems to be functional and structural overlap — a single structure might have many divergent functions depending on its spatial and temporal expression or, conversely, a given function might be elicited by several closely related glycan structures.

Although there are several systems with well-defined structure–function relationships that have facilitated drug discovery — such as the ATIII/pentasaccharide system^{4,19}, the mannose 6-phosphate pathway and lysosomal storage²⁰, and the sialic-acid content of *N*- and *O*-linked glycoproteins controlling the clearance of glycoproteins from the blood²¹ — integration of a BIOINFORMATIC framework with structural investigations in many other systems is lacking. Beyond this important limitation, from a drug development standpoint, several factors have hampered the development of glycan-based pharmaceuticals.

First, glycans have a remarkably high information content owing to diversity in their primary chemical structures, which requires the application of various analytical techniques to discern the glycan fine structure and sequence (see later). This information density derives from the fact that complex glycans not only contain information from diverse monosaccharides but also have linkage variability. There are several different families of complex glycans that are differentiated according to the types of linkage and the exocyclic substitutions of the monosaccharides. Therefore, unlike chemical polymers — which often contain homogeneous repeating units in which the diversity of the mixture primarily arises from the degree of

polymerization or cross linking between these units — linear glycan mixtures are complex owing to chemical diversity in the repeating units.

Second, glycans are not synthesized *in vivo* by reading from a template, unlike proteins or DNA. Rather, complex glycans are created through the concerted action of several enzymes. This, together with a lack of proof-reading machinery, results in complex glycans being heterogeneous and often polydisperse entities, particularly when compared with other biological polymers, such as proteins and nucleic acids.

Third, there is no mechanism, at present, to amplify glycan structures. The advent of amplification techniques for proteins and nucleic acids allowed numerous detailed structure–function studies to proceed for these biopolymers. Unfortunately, owing to the complexity of their biosynthesis, as well as the fact that they are created not in a template-based manner but rather by the concerted action of biosynthetic enzymes, there are no such strategies for complex glycans. In addition, our ability to manipulate the structure of complex glycans through chemical modification is limited, although it is increasing (see later).

Fourth, glycans predominantly act at an extracellular and/or multicellular level, and, as such, the screening of glycan activities using simple *in vitro* cell-model systems does not always provide an avenue to determine the *in vivo* function accurately. To add to the problem, many glycans interact only weakly as 1:1 complexes with targets; in fact, the strength of the interaction is typically provided through the multivalent interactions of the glycan with the targets²². As a result, screening for strong glycan binders assuming a 1:1 stoichiometry — in a similar manner to screens that are used by the pharmaceutical industry to uncover small-molecule lead compounds — might overlook information that is crucial to understanding the role of complex glycans in a given biological process. An example of this in the design of compounds to inhibit leukocyte homing and extravasation, in which multivalency and structural complexity are both important^{23,24}.

HEPARIN

A linear polysaccharide consisting of a disaccharide repeat unit, in which each individual unit can be differentially sulphated, which leads to structural complexity. Heparin is used as an anticoagulant.

ANTITHROMBIN III

(ATIII). A serine protease inhibitor that interrupts the coagulation cascade to provide an important feedback loop. Heparin binds to ATIII, which causes a conformational change in the protein and promotes its anticoagulant function.

BIOINFORMATICS

Data integration, mining and comparison tools that are used to combine data sets.

Table 1 | **Representative examples of glycan-based therapeutics**

Drug	Disease	Clinical status	Manufacturer
Lovenox	Thrombosis	Market	Aventis
Fragmin	Thrombosis	Market	Pfizer
Aranesp	Anaemia	Market	Amgen
Seprigel	Anti-adhesive	Market	Genzyme
Healon	Cataracts	Market	Pfizer
Arixtra	Thrombosis	Market	Sanofi
Cerezyme	Gaucher's disease	Market	Genzyme
Aldurazyme	MPS I	Market	Genzyme
PI-88	Cancer	Phase II	Progen

MPS I, mucopolysaccharidosis type I.

Glycans and biotechnological applications

Historically, a series of enzymes — including imiglucerase (Cerezyme; Genzyme) and larodinas (Aldurazyme; Genzyme) — that degrade or alter complex glycan structures have been successful in replacement therapy for individuals with glycan-related genetic disorders. 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 various diseases, such as Gaucher's disease and **mucopolysaccharidosis type I** (TABLE 1).

Glycans also have an important role in two distinct but related areas of drug development: either as part of glycoprotein therapeutics, modulating protein activity and stability, or alone as saccharide-based therapeutics. Given the importance of protein-based therapies — including antibodies and drugs such as growth hormone and ERYTHROPOIETIN, which together constitute a multi-billion-dollar-per-year industry — most of the initial focus has been on understanding and manipulating the complex glycan structures that are attached to glycoproteins. However, the development of structure–activity relationships for free glycans, specifically polysaccharides, has also allowed the development of new classes of therapeutics.

Complex glycan structures of glycoproteins. Most therapeutic proteins, including antibodies, growth factors and cytokines, are derived from endogenous glycoproteins. As a component of therapeutic glycoproteins, glycans modulate the activity, stability, serum half-life and immunogenicity of biotechnologically derived proteins. The fact that complex glycans are important in protein function has been appreciated on a general level for many years, since the early days of the development of protein-based therapeutics. It has long been noted that different isoforms of a protein therapeutic, which carry distinct glycoforms, possess varied physicochemical and biological attributes. Therefore, it was concluded that the quality of a protein product could be influenced by the degree and differences in glycan or oligosaccharide structure^{25–27}. Historically, the lack of analytical tools for understanding glycan structure and the dearth of structure–function relationships for specific therapeutic

glycoproteins created a situation in which the focus was not on the optimization of glycosylation or on understanding how particular glycoforms impinge on biological functioning; rather, the focus was on making protein glycosylation as homogenous as possible.

Given the seeming importance of glycosylation in the serum half-life and *in vivo* biological activity of therapeutic glycoproteins, this endeavour has sought to impinge on the glycosylation machinery in three ways: by the addition or subtraction of chemicals that influence the synthesis of complex polysaccharide structures; by the *in vitro* elaboration of glycan structures through the addition of exogenous glycosyltransferases and activated sugar substrates; and by the genetic manipulation of cells that produce therapeutic glycoproteins. In the first case, the effect of cell-culture conditions on glycan structure has been investigated^{28,29}, as has the introduction of specific monosaccharides to cells, such as *N*-acetylmannosamine or glucosamine^{26,27}. In the second case, it has been reported that specific saccharide structures can be introduced into recombinant glycoproteins using exogenous glycosyltransferases. A recent report illustrated that the addition of a specific saccharide structure (sialyl Lewis X) to a soluble human complement receptor through the *in vitro* use of specific glycosyltransferases resulted in the effective targeting of the complement receptor to activated endothelial cells that expressed E-selectin³⁰. Finally, in the third case, the genetic manipulation of both mammalian and insect cells has been reported to result in the formation of fully elaborated, relatively homogenous glycoforms^{31–33}. Recently, a new procedure has been developed in which a therapeutic glycoprotein (for example, epoetin (Epoen; Amgen)) is manipulated at a genetic level to create further sites of potential asparagine or *N*-linked glycosylation³⁴. As shown in TABLE 1, modifying the glycan coat of a protein therapeutic led to the creation of the second-generation anti-anaemia drug, darbepoetin (Aranesp; Amgen), with new and improved therapeutic qualities (see later)³⁵.

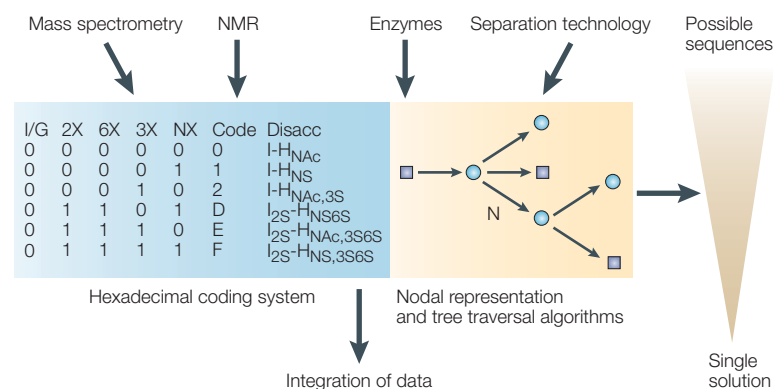
Complex glycans as therapeutics. In addition to their role as conjugates of protein therapeutics, complex glycans that are isolated from natural sources or are chemically synthesized are themselves active biological and pharmaceutical agents. In this respect, several glycan drugs are now available clinically and many others are in various stages of clinical trials (TABLE 1). The development of glycan-based therapeutics has traditionally been accomplished through two separate approaches, as discussed below.

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: fondaparinux sodium (Arixtra; Sanofi). Solid-phase synthetic procedures have been recently developed for glycans, which are analogous to the available peptide and nucleic-acid synthetic routes, and hold promise for potential scale-up in the near future. Solid-phase synthetic routes have already been developed for some drugs, such as carbohydrate-based vaccines³⁶.

ERYTHROPOIETIN

An endogenous glycoprotein, the main function of which is to stimulate the proliferation and differentiation of erythroid precursors in the bone marrow. As a pharmaceutical agent, this protein is used for the treatment of anaemia owing to the effect of concomitantly administered chemotherapy agents in cancer patients.

Box 1 | Bioinformatics-based integration of data sets



Recently, a systematic and numerical-based nomenclature, which is also known as property-encoded nomenclature or PEN, has been introduced to encode the dense information that is inherent in complex glycans^{46,47}. The strengths of this approach are its ability to deal with diverse sets of data and to move from a set of all possible theoretical sequences to a single solution using a minimum number of experimental constraints. This bioinformatics platform has been successfully integrated with a wide range of analytical techniques to facilitate the rapid and precise sequencing of minute amounts of glycan samples. PEN has been successfully applied to information from mass-spectrometric techniques to sequence the linear glycans that bind to antithrombin III (ATIII)⁸¹. This same bioinformatics framework can combine orthogonal data sets to arrive at solutions, including integrating diverse techniques such as NMR and capillary electrophoresis⁴⁷. Therefore, PEN allows several different techniques to be used in conjunction to provide extra confidence, as well as eliminating possible sequences until convergence on a single sequence is achieved. This approach, and others like it that seek to integrate several experimental data sets, have the potential to deconvolute complex glycan mixtures and yield important structure–function relationships.

The second manufacturing strategy uses 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. The quintessential example of this approach is the family of LOW MOLECULAR-WEIGHT HEPARINS (LMWHs)³⁷, which are described in more detail below.

Glycans, either alone or as GLYCOCONJUGATES, have great potential for use as drugs, for several practical reasons. First, complex glycans are relatively small and intrinsically more stable than protein-based drugs. Second, they are more easily formulated for drug delivery. Third, and finally, sugars are highly specific and potentially less immunogenic than other natural products, such as proteins or RNA-based strategies³⁸. At the same time, the large-scale production of specific biologically active glycans is still a technically demanding task. Owing to their non-template biosynthesis, glycans cannot be readily manufactured, unlike protein therapeutics in which cells can be used as ‘biotechnological factories’. However, important advances have been made in several orthogonal approaches for their structural optimization and manufacture; these developments will enhance our ability to manufacture a wide range of complex polysaccharide biotherapeutics.

LOW MOLECULAR-WEIGHT HEPARIN

This is produced when heparin is cleaved to produce lower molecular-weight species, which are then purified.

GLYCOCONJUGATE

Glycans that are covalently attached to other biomolecules, such as proteins or lipids.

New technologies to study complex glycans

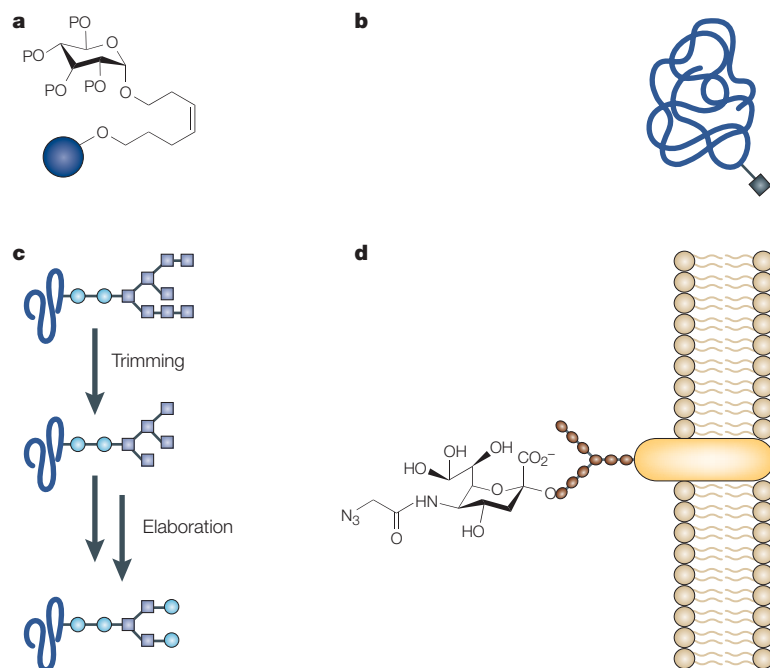
As our understanding of the structure–function relationships of complex glycans increases, so will our ability to harness their power in various disease processes. Glycan-based drugs have generated much excitement and provided important insight into the power of glycan-based therapeutics. However, the ultimate promise of glycans as drugs is only beginning to be exploited. In an effort to harness the promise of glycans as therapeutics, advances have been made in analysing glycan structures in a rapid manner using a minimum of material, in synthesizing glycan structures *in vitro* and in harnessing endogenous glycosylation pathways *in vivo* to create new reproducible glycan structures. Each of these is discussed in more detail below.

Developments in analytical procedures. Recently, there has been a marked increase in the reporting of techniques that have been successfully applied to the analysis of complex glycans and glycoconjugates, including mass spectrometric (MS)^{39–41} and capillary electrophoretic (CE)⁴² techniques. Many of these technologies have distinct advantages compared with traditional analytical methodologies, including the ability to analyse minute amounts of biologically based material. In addition, owing to their high resolving power, these state-of-the-art techniques have the ability to differentiate between closely related glycan structures.

To handle the structural complexity and heterogeneity of complex glycans, there has been an increased recognition that hyphenated techniques — that is, analytical techniques that combine aspects of separation, analytical measurement and, in some instances, data integration through bioinformatics — are essential for glycan structural characterization; examples of such technologies include capillary liquid chromatography (LC)-MS⁴³, CE-MS⁴⁴ and LC-NMR⁴⁵ techniques. Integration into a bioinformatics framework has been successfully completed for a range of analytical methods, including NMR, CE and MS^{46,47}, and is discussed in more detail in BOX 1. Several investigators have also sought to develop chip-based separation and analysis technology; various strategies have been reported, including fluorescent labelling of ligands coupled to imaging analysis, and combining chip-based separation with MS analysis^{48–52}.

Advances in synthetic strategies. In addition to advances in analytical methods, new techniques are also beginning to be designed for the synthesis of homogenous glycans and glycoconjugates; such approaches are crucial to develop and test structure–activity relationships (BOX 2). For example, solution and solid-phase synthetic procedures are continuing to be evaluated for both linear and branched glycans^{53,54}. The development of *in vitro* chemoselective and chemoenzymatic ligation procedures promises to create new glycoconjugates. This has most recently been illustrated by the creation of a long-lasting insulin through the enzymatic introduction of a trisaccharide to the amino terminus of the polypeptide chain.

Box 2 | The creation of defined glycan structures



Several distinct methodologies have recently been developed that, together, hold much promise towards facilitating structure–function studies of complex glycans, both alone and as glycoconjugates. A solid-phase-synthesis strategy for glycan structures has recently been reported, which allows for the rapid synthesis of complex glycans and requires minimal clean-up⁵³; in the example illustrated (a), a monosaccharide is attached to an octenediol-functionalized resin and elaborated through the addition of further monosaccharides (one per cycle). Similarly, the introduction of a single non-natural amino acid into the amino-acid sequence of a protein, in which the modified amino acid is genetically encoded, has been reported to produce defined glycoconjugates⁵⁵; in the example illustrated (b), the modified structure is a β -*N*-acetylglucosamine (GlcNAc)-serine (represented as a black square attached to the peptide backbone), which allows the further elaboration of complex structures through either chemical or enzymatic means. Two recent reports have shown that yeast can be engineered to produce complex glycans, such as those in human glycoproteins (c). This process involves the systematic elimination of endogenous yeast glycosylation pathways, in conjunction with the correct localization of five eukaryotic glycan synthesis and metabolism proteins. Targeted localization of the enzymes allows the generation of a synthetic *in vitro* glycosylation pathway, which produces complex glycan structures. Finally, the addition of small non-natural monosaccharides to cells both *in vitro* and *in vivo* results in their uptake and integration into the glycan-synthetic machinery (d)^{59,60}. Monosaccharides can be engineered with a chemical handle, which allows selective chemistry to occur, such as the targeting of cancer cells with reactive therapeutics or with a diagnostic-imaging agent.

Two further techniques have been developed that can create homogenous protein/glycan conjugates in the cell in a precisely defined manner. The first 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 elaboration of a polysaccharide chain through chemical and/or enzymatic routes⁵⁵. The second approach involves the Herculean effort of recreating the mammalian glycosylation machinery in yeast⁵⁶. Both have been shown to provide homogenous glycoforms that

can be further elaborated and that hold promise for the routine creation of defined glycosylation patterns for structure–function relationships.

Finally, 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^{57–60} (BOX 2). After incorporation, these monosaccharides can be detected through unique chemical handles. Such an approach has shown some promise in terms of analysing signalling pathways that are regulated by particular glycan post-translational modifications^{61,62}, as well as in tailoring glycosylation *in vivo* for drug development efforts.

Improvements through glycan redesign

Advances in our understanding of the structure–function relationships of complex glycans have allowed three important developments: first, the creation of new therapeutic modalities for a wide range of diseases; second, the complete description of existing complex glycan therapeutics to potentially develop generics; and third, the redesign of known drugs to create new second-generation molecules with tailored biological characteristics. This last innovation, in particular, effectively illustrates the power of structure–function studies of complex glycans. Two examples of this approach are detailed below. The first involves the polysaccharide-based LMWHs; recent advances in our understanding of these molecules have created opportunities for the development of redesigned LMWHs as improved anticoagulants and, potentially, as effective agents in other diseases, such as cancer or inflammation. The second example is the anaemia drug erythropoietin; understanding and systematically altering the glycan structure of this drug has allowed the development of second-generation molecules with optimized activity. This is of particular importance because understanding the glycan structure of a glycoprotein therapeutic, such as erythropoietin, and altering its glycan structure without affecting the protein backbone, makes it possible to create molecules with divergent properties and raises the possibility of the creation of tailored drugs.

LMWHs. Unfractionated heparin, which is the starting material for all LMWHs, has been used as an anticoagulant for more than 60 years⁶³. The mechanism of action for heparin was identified through numerous structural and biological studies. This research revealed that the anticoagulant activity of heparin predominantly occurs through the inhibition of two components of the coagulation cascade: factor Xa and factor IIa. In both cases, a specific pentasaccharide sequence in heparin binds with high affinity to the serine protease inhibitor ATIII (REF. 4). This binding induces a conformational change in the protein that allows ATIII to inhibit factor Xa⁴. In addition, heparin chains that are longer than 18 saccharide units, which contain the ATIII-binding pentasaccharide motif, bridge factor IIa and ATIII — this allows ATIII to

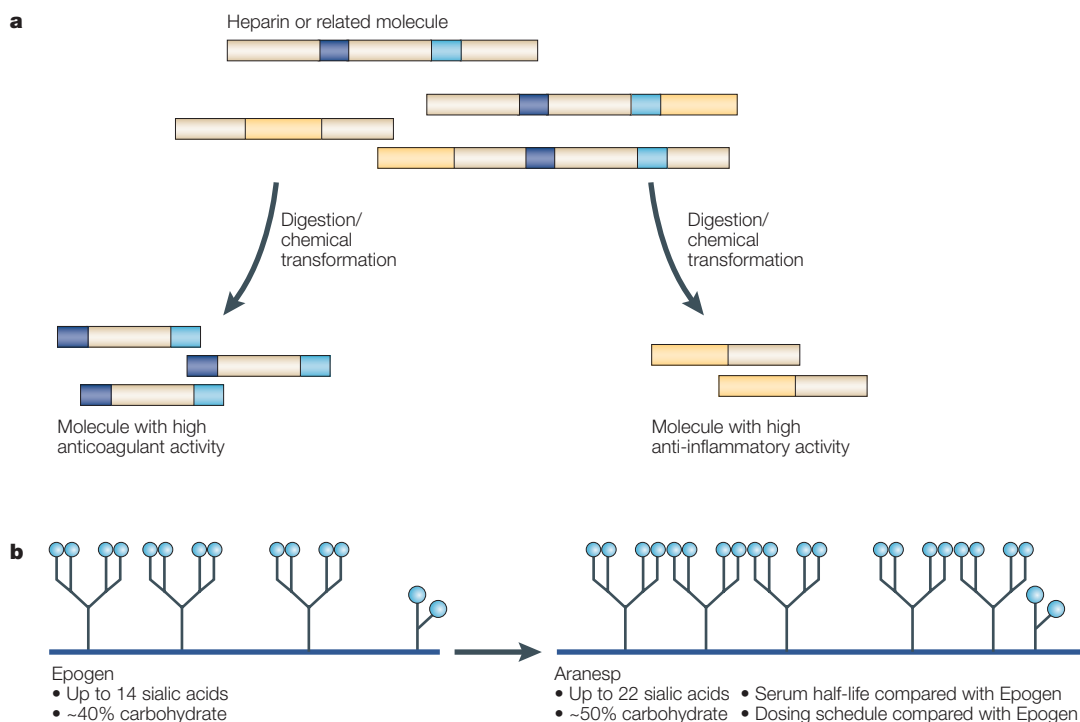


Figure 3 | New second-generation therapeutics based on glycan structures. a | Heparin is a mixture of linear polysaccharides that differ both in terms of sequence and chain length. The heparin chain contains sequences that bind with high affinity to antithrombin III (ATIII) and thrombin (factor IIa). By binding these factors, heparin effectively inhibits blood coagulation. However, much of the heparin chain is not involved in its anticoagulant activity; in fact, a part is implicated in the side effects that are associated with heparin therapy. Through understanding the structure of heparin and how it impinges on function, it has been shown that the deleterious and inactive components can be removed, which yields a low molecular-weight heparin with high activity and a tailored clinical profile^{64,65}. **b** | Epogen is a 30-kDa protein with three *N*-linked branched sugars. Through site-directed mutagenesis, two further *N*-linked glycosylation sites were introduced into the protein to produce the next generation anti-anaemia drug Aranesp³⁴. This protein has higher *in vivo* activity and a longer serum half-life than the parent molecule, which result in a substantial clinical benefit.

inhibit factor IIa⁶⁴. LMWHs, which are important antithrombotic drugs, are derived from unfractionated heparin by a range of chemical and/or enzymatic techniques³⁷. Their structural complexity is therefore a consequence of the heterogeneity of the parent compound as well as chemical differences in their preparation. Three main LMWHs are in clinical use at present: dalteparin (Fragmin; Pfizer), tinzaparin (Innohep; DuPont Pharmaceuticals) and enoxaparin (Lovenox; Aventis).

Given the understanding of the mechanisms by which heparin inhibits factor Xa and factor IIa, several strategies have been devised to design LMWHs (and mimetics) with tailored activity. *De novo* synthesis has generated molecules that have an ATIII-binding site separated from a thrombin-binding site by spacers with different properties. A series of related molecules that were created by this route were used to probe *in vivo* anticoagulant function, and were optimized for anti-Xa and anti-IIa activity⁶⁴. To circumvent the challenges that are associated with chemical synthetic strategies for heparin-based polysaccharides, a chemo-enzymatic synthesis has also been used for the synthesis of Arixtra. Finally, a practical methodology was recently developed to quantify the molar composition of a structural correlate (a tetrasaccharide, which is part of the pentasaccharide motif) in a given LMWH sample⁶⁵. The abundance

of this structural motif was found to strongly correlate with anticoagulant activity. Using this structural correlate, which is referred to as $\pm 4-7$, a rational approach was applied to enhance the abundance of the correlate and to generate a new family of rationally designed LMWHs with increased anticoagulant activities⁶⁵ (FIG. 3a). These molecules were found to circumvent many of the shortcomings of known LMWHs, including the possession of an increased efficacy in arterial thrombotic settings.

Finally, heparin, as a complex polysaccharide, has been shown to possess several activities that extend beyond anticoagulation, including anti-inflammatory and anticancer properties. Numerous scientific studies have identified a role for heparin oligosaccharide structures in mediating diseases, such as viral infection⁶⁶, cancer⁶⁷, Alzheimer's disease⁶⁸ and inflammation⁶⁹. There is great potential for the design of new treatments for these diseases through chemical, chemo-enzymatic and enzymatic routes, using the heparin scaffold as a starting point.

Erythropoietin. One of the most important biotechnology products is the blockbuster anti-anaemia drug erythropoietin. This 30-kDa protein contains three different sites of *N*-linked glycosylation and stimulates

Box 3 | **The Consortium for Functional Glycomics**

The goal of the Consortium for Functional Glycomics is to understand the role of carbohydrate–protein interactions in cell–cell communication. As such, the consortium has set out to build an infrastructure that will facilitate research in the field of glycomics.

The infrastructure consists of scientific cores that are dedicated to generating reagents and developing technologies, and then making them available to investigators who are both internal and external to the consortium (FIG. 4). The unique reagents and methodologies that have been developed so far include: a glycogene microarray to analyse expression patterns for biosynthetic and catabolic enzymes that are involved in glycan construction and elaboration; expression analysis using sophisticated algorithms for the automated analysis of complex data sets; and printed glycan arrays for the detection of the specificities of carbohydrate-binding proteins.

The different cores of the consortium generate various types of data pertaining to carbohydrate–protein interactions, which range from genetic knockouts to analytical measurements by mass spectrometry and array technologies. For example, profiling has been conducted for *N*- and *O*-linked glycans from different mouse organs and human tissues, which has led to the detection and identification of new glycans. The genomics-, proteomics- and glycomics-related information about carbohydrate-binding proteins, glyco-enzymes and carbohydrates is integrated into a single bioinformatics platform.

To this end, a central database is being created, which should prove as useful to investigators in glycomics as **GenBank** has been to investigators in molecular biology. This central database is a composite of relational databases, into which the results from each of the cores is being entered to allow the seamless viewing of all scientific data so as to facilitate future discoveries (see **Updates on the Consortium** in the online links box). The integrated data are made available through different user interfaces, such as a molecule page on which all the data pertaining to a specific molecule are presented (see **Galectin 3 (human)** in the online links box for an example of a molecule page). In addition, through the use of statistical analysis, data mining and other bioinformatics tools, predictive models can be constructed on the basis of diverse and complementary data sets.

the production of red blood cells *in vivo*⁷⁰. Investigations into the glycosylation of erythropoietin have highlighted the importance of specific capped structures in terms of the serum half-life of the protein⁷¹. Worldwide, there are two main versions of the drug: Procrit, which is marketed by Johnson and Johnson; and Epogen, which is marketed by Amgen. Recently, however, a new form of erythropoietin has been introduced that has generated much excitement in the nephrology and oncology fields. This new erythropoiesis-stimulating protein, Aranesp, differs from erythropoietin only in terms of the number of *N*-linked glycans that are attached to the peptide backbone³⁴ (FIG. 3b). In biological and clinical studies, Aranesp was found to possess higher activity and a substantially increased serum half-life compared to erythropoietin³⁵. This has allowed the creation of a dosing schedule for Aranesp that supports higher patient compliance and so increases the overall effectiveness of the anti-anaemia drug.

Importantly, however, this is not the entire story. In marked contrast to sialylated erythropoietin, it was recently reported that erythropoietin that lacks the capping sialic acid (asialoEPO) possesses potent and broad neuroprotective activities in diseases ranging from cerebral ischaemia to mechanical damage of the central nervous system⁷² — in other words, it seems that the erythropoietic and tissue-protective effects of erythropoietin can be dissociated from one another. Taken together, these findings show that by understanding the glycosylation of a biomolecule therapeutic and rationally changing its glycosylation, it is possible to alter the efficacy and *in vivo* half-life of a protein, or, alternatively, to change the *in vivo* effect of a drug. Therefore, it seems that, as with linear glycans, an array of sugar structures might be important for distinct functions and it is not

simply a matter of homogenizing to a single glycoform. In fact, the development of complete integrated systems for understanding the structure–function activities of complex glycans is necessary to maximize the therapeutic potential of complex glycans.

Challenges and opportunities ahead

Given the structural and informational complexity of glycans, what are the main challenges ahead and how will they influence drug development in this field? We have at our disposal diverse techniques to characterize and synthesize glycans, yet it remains a challenge to accurately translate this into a description that reflects the function of a complex glycan pool, to allow rapid drug development. To add to the problem, although the arrangement of the monomer units or sequence is clearly an important characteristic of all polymers, including complex glycan mixtures, this is insufficient to completely describe a complex glycan pool. This is in contrast to other biopolymers, such as DNA, in which the sequence alone is sufficient to describe the molecule. Therefore, the descriptions of glycan structure–function relationships that are yielded by the burgeoning field of glycomics will undoubtedly differ from those produced by proteomics and genomics, which are used to describe protein and nucleic-acid structure–function relationships. In the case of complex glycans, which exist as mixtures, there are several types of molecule present; therefore, beyond their identities, the relative abundance of each sequence in the mixture is required to accurately compare glycan pools to one another.

Traditionally, summary properties have been used to describe complex glycan mixtures. Although these are useful in describing certain attributes of the mixture, they often overlook other crucial properties. For instance,

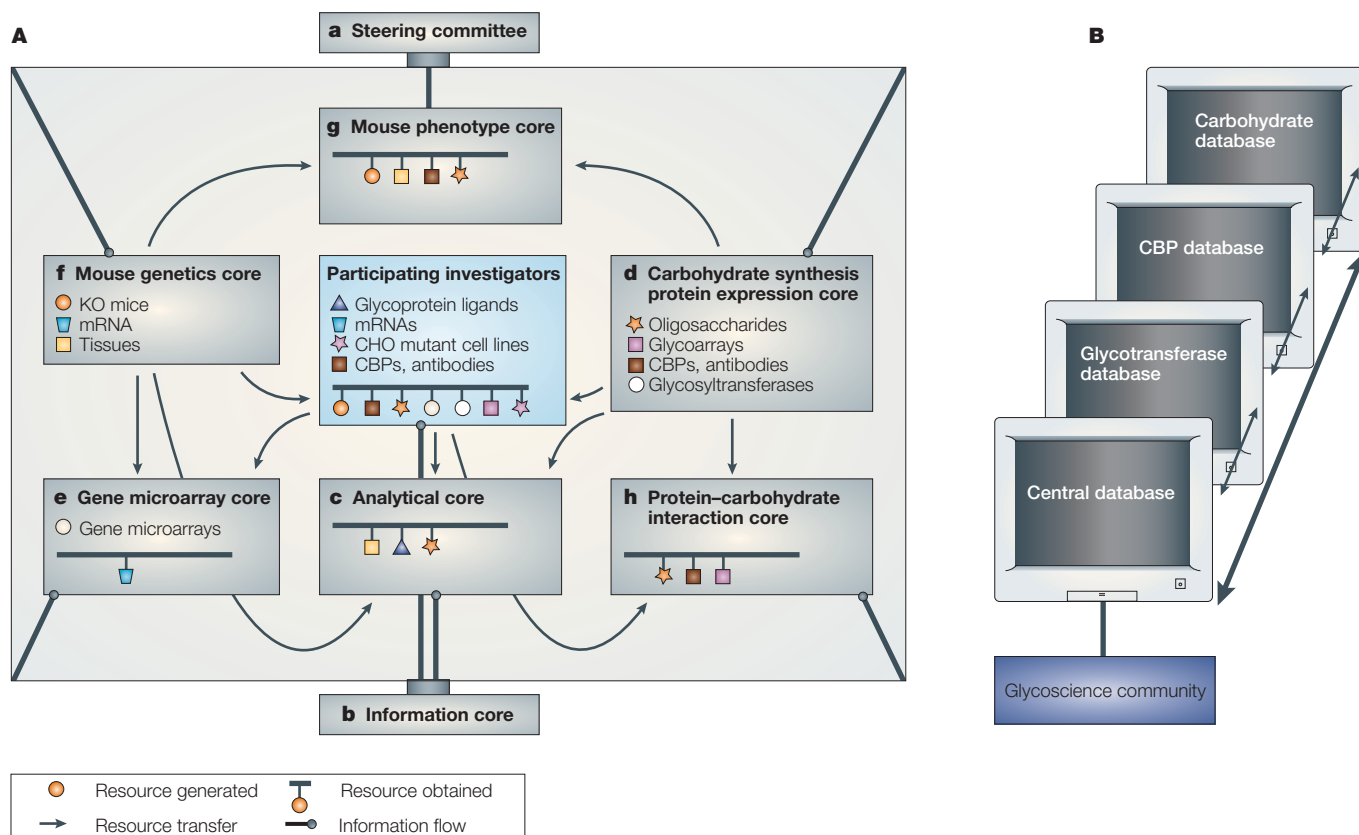


Figure 4 | The Consortium for Functional Glycomics. A | The infrastructure of the consortium consists of scientific cores (labelled **b–g**) that are dedicated to generating reagents and developing technologies for glycan analysis, and then making them available to investigators who are both internal and external to the consortium. **B** | A central database is being constructed that is a composite of relational databases, into which experimental results that are derived from each of the cores are being entered; this will allow the seamless viewing of all scientific data to facilitate future discoveries. Through the integration of many data sets, important relational and pattern data will emerge about glycan structure–function relationships. CBP, carbohydrate-binding protein; CHO, Chinese hamster ovary cells; KO, gene knockout.

molecular-weight averages and distributions, and monosaccharide composition, have been used to describe complex glycan mixtures, such as LMWHs. Molecular-weight distributions are sufficient to describe homopolymers and block polymers; however, they are not adequate to describe complex glycan mixtures. Clearly, molecular-weight distributions and averages are important and necessary properties, but because of the structural complexity and heterogeneity of complex glycans, they do not capture many important aspects of the mixture.

How should the task of characterizing such complex biological mixtures and applying this information to drug development be approached? The answer lies in the integration of experimentally derived data sets using a bioinformatics framework; that is, the development of strategies that are similar to those outlined in BOX 1. It must be recognized that a single analytical tool, no matter how sophisticated, is insufficient to completely characterize the glycan mixtures that are currently used as drugs or that might be developed in the future. Instead of relying on one technique, 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 set, 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. Data-integration methodologies hold much promise for the analysis of complex glycans. The implementation of these techniques is expected to 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. An effort currently being undertaken by the Consortium for Functional Glycomics seeks to understand the role of carbohydrate–protein interactions using such a strategy (BOX 3; FIG. 4).

In these ways, we might be able to develop a structure–function relationship model for complex glycans that will unlock the potential of these molecules as therapeutics and diagnostic agents. Because of their multifaceted responses and the fact that many

structures can 'code' for a single function, such a model will capture the essence of the 'glycome'. This will allow the development of an '-omics' approach to the study of glycan structure that is similar to, but in many ways distinct from, the way in which we have developed genomics for the analysis of the genetic programming of the cell, or proteomics for the analysis of the protein repertoire of the cell.

Summary

In view of the various biochemical pathways and disease processes in which glycans are crucially engaged — angiogenesis⁷³, cancer^{74–76}, tissue repair, cardiovascular disease⁷⁷, immune-system function^{78,79}, and microbial and viral pathogenesis⁸, to name a few — the possibilities for glycans as therapeutics and diagnostics⁸⁰ are numerous and exciting. As we further improve our ability to study glycans, we will gain a better understanding of existing drugs, as well as developing new specific therapeutics. Important challenges lie ahead as we make our way in a post-genomic world, in which an understanding of glycan structure and function will undoubtedly lead

to a new generation of highly effective therapeutics. Importantly, to overcome the challenges and to harness the excitement that this subject area has recently generated, the US National Institutes of Health has funded the development of the Consortium for Functional Glycomics to advance the burgeoning glycomics field. From our perspective, the future is exciting and the possibilities are numerous.

At present, complex glycan drugs are marketed in two main areas: the treatment of thrombosis (for example, Lovenox and Fragmin) and as important structural components to promote healing (for example, Seprigel (Genzyme) and Healon; (Pfizer)). The introduction of Aranesp, which is a hyperglycosylated form of an existing anti-anaemia drug, has provided an intriguing hint as to the potential of glycan-based drug-improvement strategies. Furthermore, the potential of glycan-based strategies in multicellular diseases, such as cancer, is now beginning to be exploited. Finally, enzymes that act on glycoconjugates, such as Aldurazyme and Cerezyme, are now being used in enzyme-replacement strategies.

- Varki, A. *et al.* *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, New York, 1999).
- Perrimon, N. & Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725–728 (2000).
- A review of the importance of linear polysaccharides in developmental processes.**
- Ioffe, E. & Stanley, P. Mice lacking *N*-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid *N*-linked carbohydrates. *Proc. Natl Acad. Sci. USA* **91**, 728–732 (1994).
- Jin, L. *et al.* The anticoagulant activation of antithrombin by heparin. *Proc. Natl Acad. Sci. USA* **94**, 14683–14688 (1997).
- Fu, X. *et al.* Antibiotic optimization via *in vitro* glyco-randomization. *Nature Biotechnol.* **21**, 1467–1469 (2003).
- Iozzo, R. V. Heparan sulfate proteoglycans: intricate molecules with intriguing functions. *J. Clin. Invest.* **108**, 165–167 (2001).
- Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370–2376 (2001).
- Alper, J. Glycobiology. Turning sweet on cancer. *Science* **301**, 159–160 (2003).
- Lowe, J. B. & Marth, J. D. A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* **72**, 643–691 (2003).
- Kusche-Gullberg, M. & Kjellen, L. Sulfotransferases in glycosaminoglycan biosynthesis. *Curr. Opin. Struct. Biol.* **13**, 605–611 (2003).
- Ye, Z. & Marth, J. D. *N*-glycan branching requirement in neuronal and postnatal viability. *Glycobiology* **14**, 547–558 (2004).
- Kumar, R., Yang, J., Larsen, R. D. & Stanley, P. Cloning and expression of *N*-acetylglucosaminyltransferase I, the medial Golgi transferase that initiates complex *N*-linked carbohydrate formation. *Proc. Natl Acad. Sci. USA* **87**, 9948–9952 (1990).
- Kumar, R. & Stanley, P. Transfection of a human gene that corrects the Lec1 glycosylation defect: evidence for transfer of the structural gene for *N*-acetylglucosaminyltransferase I. *Mol. Cell Biol.* **9**, 5713–5717 (1989).
- D'Agostaro, G. A. *et al.* Molecular cloning and expression of cDNA encoding the rat UDP-*N*-acetylglucosamine: α -6-*c*-mannoside β -1,2-*N*-acetylglucosaminyltransferase II. *J. Biol. Chem.* **270**, 15211–15221 (1995).
- Wang, Y. *et al.* Modeling human congenital disorder of glycosylation type IIa in the mouse: conservation of asparagine-linked glycan-dependent functions in mammalian physiology and insights into disease pathogenesis. *Glycobiology* **11**, 1051–1070 (2001).
- Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **407**, 1029–1034 (2000).
- Schlessinger, J. *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* **6**, 743–750 (2000).
- Rini, J. M. & Lobsanov, Y. D. New animal lectin structures. *Curr. Opin. Struct. Biol.* **9**, 578–584 (1999).
- Desai, U. R., Petitou, M., Bjork, I. & Olson, S. T. Mechanism of heparin activation of antithrombin. Role of individual residues of the pentasaccharide activating sequence in the recognition of native and activated states of antithrombin. *J. Biol. Chem.* **273**, 7478–7487 (1998).
- Kornfeld, S. Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. *Annu. Rev. Biochem.* **61**, 307–330 (1992).
- Stockert, R. J. The asialoglycoprotein receptor: relationships between structure, function, and expression. *Physiol. Rev.* **75**, 591–609 (1995).
- Yang, Z. Q., Puffer, E. B., Pontrello, J. K. & Kiessling, L. L. Synthesis of a multivalent display of a CD22-binding trisaccharide. *Carbohydr. Res.* **337**, 1605–1613 (2002).
- Gotte, M. Syndecans in inflammation. *FASEB J.* **17**, 575–591 (2003).
- van Zante, A. & Rosen, S. D. Sulphated endothelial ligands for L-selectin in lymphocyte homing and inflammation. *Biochem. Soc. Trans.* **31**, 313–317 (2003).
- Baker, K. N. *et al.* Metabolic control of recombinant protein *N*-glycan processing in NS0 and CHO cells. *Biotechnol. Bioeng.* **73**, 188–202 (2001).
- Yang, M. & Butler, M. Effects of ammonia and glucosamine on the heterogeneity of erythropoietin glycoforms. *Biotechnol. Prog.* **18**, 129–138 (2002).
- Yang, M. & Butler, M. Effect of ammonia on the glycosylation of human recombinant erythropoietin in culture. *Biotechnol. Prog.* **16**, 751–759 (2000).
- Senger, R. S. & Karim, M. N. Effect of shear stress on intrinsic CHO culture state and glycosylation of recombinant tissue-type plasminogen activator protein. *Biotechnol. Prog.* **19**, 1199–1209 (2003).
- Thomas, L. J. *et al.* Production of a complement inhibitor possessing sialyl Lewis X moieties by *in vitro* glycosylation technology. *Glycobiology* 9 June 2004 [pub ahead of print].
- Weikert, S. *et al.* Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. *Nature Biotechnol.* **17**, 1116–1121 (1999).
- Zhang, X., Lok, S. H. & Kon, O. L. Stable expression of human α -2,6-sialyltransferase in Chinese hamster ovary cells: functional consequences for human erythropoietin expression and bioactivity. *Biochim. Biophys. Acta* **1425**, 441–452 (1998).
- Tomiya, N. *et al.* Complex-type biantennary *N*-glycans of recombinant human transferrin from *Trichoplusia ni* insect cells expressing mammalian β -1,4-galactosyltransferase and β -1,2-*N*-acetylglucosaminyltransferase II. *Glycobiology* **13**, 23–34 (2003).
- Egrie, J. C. & Browne, J. K. Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br. J. Cancer* **84** (Suppl. 1), 3–10 (2001).
- Egrie, J. C., Dwyer, E., Browne, J. K., Hiltz, A. & Lykos, M. A. Darbepoetin α has a longer circulating half-life and greater *in vivo* potency than recombinant human erythropoietin. *Exp. Hematol.* **31**, 290–299 (2003).
- 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).
- Linhardt, R. J. & Gunay, N. S. Production and chemical processing of low molecular weight heparins. *Semin. Thromb. Hemost.* **25** (Suppl. 3), 5–16 (1999).
- Sioud, M. Therapeutic siRNAs. *Trends Pharmacol. Sci.* **25**, 22–28 (2004).
- Kaji, H. *et al.* Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins. *Nature Biotechnol.* **21**, 667–672 (2003).
- Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnol.* **21**, 660–666 (2003).
- Dell, A. & Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356 (2001).
- Que, A. H. *et al.* Coupling capillary electrophoresis with electrospray Fourier transform mass spectrometry for characterizing complex oligosaccharide pools. *Anal. Chem.* **75**, 1684–1690 (2003).
- Kawasaki, N., Itoh, S., Ohta, M. & Hayakawa, T. Micro-analysis of *N*-linked oligosaccharides in a glycoprotein by capillary liquid chromatography/mass spectrometry and liquid chromatography/tandem mass spectrometry. *Anal. Biochem.* **316**, 192–202 (2003).
- Gennaro, L. A., Delaney, J., Vouros, P., Harvey, D. J. & Dornon, B. Capillary electrophoresis/electrospray ion trap mass spectrometry for the analysis of negatively charged derivatized and underivatized glycans. *Rapid Commun. Mass Spectrom.* **16**, 192–200 (2002).
- Manzi, A. E. *et al.* Exploring the glycan repertoire of genetically modified mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. *Glycobiology* **10**, 669–689 (2000).
- Venkataraman, G., Shriver, Z., Raman, R. & Sasisekharan, R. Sequencing complex polysaccharides. *Science* **286**, 537–542 (1999).

This report illustrates the way in which data integration allows the rapid and precise determination of polysaccharide sequence.

47. Guerrini, M. *et al.* A novel computational approach to integrate NMR spectroscopy and capillary electrophoresis for structure assignment of heparin and heparan sulfate oligosaccharides. *Glycobiology* **12**, 713–719 (2002).
48. Wang, D., Liu, S., Trummer, B. J., Deng, C. & Wang, A. Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature Biotechnol.* **20**, 275–281 (2002).
49. Fukui, S., Feizi, T., Galustian, C., Lawson, A. M. & Chai, W. Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate–protein interactions. *Nature Biotechnol.* **20**, 1011–1017 (2002).
50. Keiser, N., Venkataraman, G., Shriver, Z. & Sasisekharan, R. Direct isolation and sequencing of specific protein-binding glycosaminoglycans. *Nature Med.* **7**, 123–128 (2001).
51. Nimrichter, L. *et al.* Intact cell adhesion to glycan microarrays. *Glycobiology* **14**, 197–203 (2004).
52. Fazio, F., Bryan, M. C., Blikt, O., Paulson, J. C. & Wong, C. H. Synthesis of sugar arrays in microtiter plate. *J. Am. Chem. Soc.* **124**, 14397–14402 (2002).
53. Plante, O. J., Palmacci, E. R. & Seeberger, P. H. Automated solid-phase synthesis of oligosaccharides. *Science* **291**, 1523–1527 (2001).
- An outline of the strategy for the solid-phase synthesis of complex polysaccharides.**
54. Ritter, T. K., Mong, K. K., Liu, H., Nakatani, T. & Wong, C. H. A programmable one-pot oligosaccharide synthesis for diversifying the sugar domains of natural products: a case study of vancomycin. *Angew Chem. Int. Ed. Engl.* **42**, 4657–4660 (2003).
55. Zhang, Z. *et al.* A new strategy for the synthesis of glycoproteins. *Science* **303**, 371–373 (2004).
56. Hamilton, S. R. *et al.* Production of complex human glycoproteins in yeast. *Science* **301**, 1244–1246 (2003).
- This report highlights the complex N-linked glycoprotein structures that are produced in yeast.**
57. Wieser, J. R., Heisner, A., Stehling, P., Oesch, F. & Reutter, W. *In vivo* modulated N-acetyl side chain of N-acetylneuraminic acid modulates the cell contact-dependent inhibition of growth. *FEBS Lett.* **395**, 170–173 (1996).
58. Kayser, H. *et al.* Biosynthesis of a nonphysiological sialic acid in different rat organs, using N-propanoyl- α -hexosamines as precursors. *J. Biol. Chem.* **267**, 16934–16938 (1992).
59. Kick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl Acad. Sci. USA* **99**, 19–24 (2002).
60. Luchansky, S. J., Goon, S. & Bertozzi, C. R. Expanding the diversity of unnatural cell-surface sialic acids. *ChemBiochem* **5**, 371–374 (2004).
61. Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl Acad. Sci. USA* **100**, 9116–9121 (2003).
62. Hang, H. C., Yu, C., Kato, D. L. & Bertozzi, C. R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl Acad. Sci. USA* **100**, 14846–14851 (2003).
- A strategy for the rapid profiling of glycosylation changes in vivo.**
63. Petitou, M., Casu, B. & Lindahl, U. 1976–1983, a critical period in the history of heparin: the discovery of the antithrombin binding site. *Biochimie* **85**, 83–89 (2003).
64. Petitou, M. *et al.* Synthesis of thrombin-inhibiting heparin mimetics without side effects. *Nature* **398**, 417–422 (1999).
- An excellent example of a structure–function study of an important class of antithrombotics.**
65. Sundaram, M. *et al.* Rational design of low-molecular weight heparins with improved *in vivo* activity. *Proc. Natl Acad. Sci. USA* **100**, 651–656 (2003).
66. Tiwari, V. *et al.* A role for 3-O-sulfated heparan sulfate in cell fusion induced by herpes simplex virus type 1. *J. Gen. Virol.* **85**, 805–809 (2004).
67. Zacharski, L. R. & Loynes, J. T. The heparins and cancer. *Curr. Opin. Pulm. Med.* **8**, 379–382 (2002).
68. Scholefield, Z. *et al.* Heparan sulfate regulates amyloid precursor protein processing by BACE1, the Alzheimer's β -secretase. *J. Cell Biol.* **163**, 97–107 (2003).
69. Malhotra, S., Bhasin, D., Shafiq, N. & Pandhi, P. Drug treatment of ulcerative colitis: unfractionated heparin, low molecular weight heparins and beyond. *Expert Opin. Pharmacother.* **5**, 329–334 (2004).
70. Goldwasser, E. Erythropoietin and its mode of action. *Blood Cells* **10**, 147–162 (1984).
71. Fukuda, M. N., Sasaki, H., Lopez, L. & Fukuda, M. Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. *Blood* **73**, 84–89 (1989).
- This definitive study shows the importance of glycosylation in modulating the circulating half-life of proteins.**
72. Erbayraktar, S. *et al.* Asialoerythropoietin is a nonerythropoietic cytokine with broad neuroprotective activity *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 6741–6746 (2003).
73. Liu, D., Shriver, Z., Venkataraman, G., El Shabravi, Y. & Sasisekharan, R. Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis. *Proc. Natl Acad. Sci. USA* **99**, 568–573 (2002).
74. Liu, D., Shriver, Z., Qi, Y., Venkataraman, G. & Sasisekharan, R. Dynamic regulation of tumor growth and metastasis by heparan sulfate glycosaminoglycans. *Semin. Thromb. Hemost.* **28**, 67–78 (2002).
75. Smorenburg, S. M. & Van Noorden, C. J. The complex effects of heparins on cancer progression and metastasis in experimental studies. *Pharmacol. Rev.* **53**, 93–105 (2001).
76. Wahrenbrock, M., Borsig, L., Le, D., Varki, N. & Varki, A. Selectin–mucin interactions as a probable molecular explanation for the association of Trousseau syndrome with mucinous adenocarcinomas. *J. Clin. Invest.* **112**, 853–862 (2003).
77. Shriver, Z., Liu, D. & Sasisekharan, R. Emerging views of heparan sulfate glycosaminoglycan structure/activity relationships modulating dynamic biological functions. *Trends Cardiovasc. Med.* **12**, 71–77 (2002).
78. Baum, L. G. Developing a taste for sweets. *Immunity* **16**, 5–8 (2002).
79. Crocker, P. R. Sialic acid-binding immunoglobulin-like lectins in cell–cell interactions and signalling. *Curr. Opin. Struct. Biol.* **12**, 609–615 (2002).
80. Dudkin, V. Y., Miller, J. S. & Danishefsky, S. J. Chemical synthesis of normal and transformed PSA glycopeptides. *J. Am. Chem. Soc.* **126**, 736–738 (2004).
81. Shriver, Z. *et al.* Sequencing of 3-O sulfate containing heparin decasaccharides with a partial antithrombin III binding site. *Proc. Natl Acad. Sci. USA* **97**, 10359–10364 (2000).

Acknowledgements

We would like to thank G. Venkataraman for help with the manuscript. Financial support for this work was provided by the National Institutes of Health and a Glue Grant for the Consortium for Functional Glycomics.

Competing interests statement

The author declares **competing financial interests**: see Web version for details.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez Gene:

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>
ATIII | MGATT | MGAT2

Online Mendelian Inheritance in Man:

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>
Alzheimer's disease | mucopolysaccharidosis type I

FURTHER INFORMATION

Consortium for Functional Glycomics (CFG):

<http://web.mit.edu/glycomics/consortium>

Essentials of Glycobiology:

<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View.ShowTOC&rid=glyco.TOC&depth=2>

Galectin 3 (human): http://web.mit.edu/glycomics/moleculerpages/cbp/galectins/gal3_human

GenBank: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

Meetings Related to Glycosciences and Carbohydrates:

<http://plab.ku.dk/tcbh/glyco-meetings.htm>

Updates on the Consortium:

<http://web.mit.edu/glycomics/consortium/whatsnew.shtml>

Access to this interactive links box is free online.