

The Sweet Science of Glycobiology

Complex carbohydrates, molecules that are particularly important for communication among cells, are coming under systematic study

Ram Sasisekharan and James R. Myette

Carbohydrates are indispensable to life on Earth. In their simplest form, they serve as a primary energy source for sustaining life. For the most part, however, carbohydrates exist not as simple sugars but as complex molecular conjugates, or *glycans*. Glycans come in many shapes and sizes, from linear chains (polysaccharides) to highly branched molecules bristling with antennae-like arms. And although proteins and nucleic acids such as DNA have traditionally attracted far more scientific attention, glycans are also key to life. They are ubiquitous in nature, forming the intricate sugar coat that surrounds the cells of virtually every organism and occupying the spaces between these cells. As part of this so-called extracellular matrix, glycans, with their diverse chemical structures, play a crucial role in transmitting important biochemical signals into and between cells. In this way, these sugars guide the cellular communication that is essential for normal cell and tissue development and physiological function.

In recent years, our laboratory and others have studied a class of linear glycans known as *glycosaminoglycans* (or GAGs for short), and particularly a sub-

set known as HSGAGs, which are made up of heparan sulfate and its relative heparin. We are beginning to understand and appreciate the integral roles these glycans play in numerous biological processes relevant to health and disease. HSGAGs are structurally diverse and differ in their chemical composition, especially with regard to the number and position of sulfate groups and the biological form in which they exist. The highly sulfated heparin is produced by connective tissue mast cells. Heparan sulfate, which has more variation in its sulfate groups, is made by practically all animal cells.

Recent technical advances have shed a good deal of light on the fine structure of HSGAGs as well as on their functions. The growing ability to determine the structure and biological context of glycans is throwing open the field of glycobiology, leading to a broader understanding of the crucial roles that these glycans play in normal physiology and in disease processes ranging from cancer to microbial infection. This new understanding opens the door to novel carbohydrate-based drug therapies.

Building the Chains

An HSGAG chain may be generically described as a linear repeat of approximately 10 to 100 disaccharide building blocks that, when linked together, make up the backbone of each sugar molecule. In its most fundamental form, each disaccharide unit consists of two chemically distinct monosaccharides (a uronic acid and a glucosamine) linked by a glycosidic bond. The chains can vary a great deal in their structural configuration because the disaccharide building blocks may be chemically modified at a number of positions. These modifica-

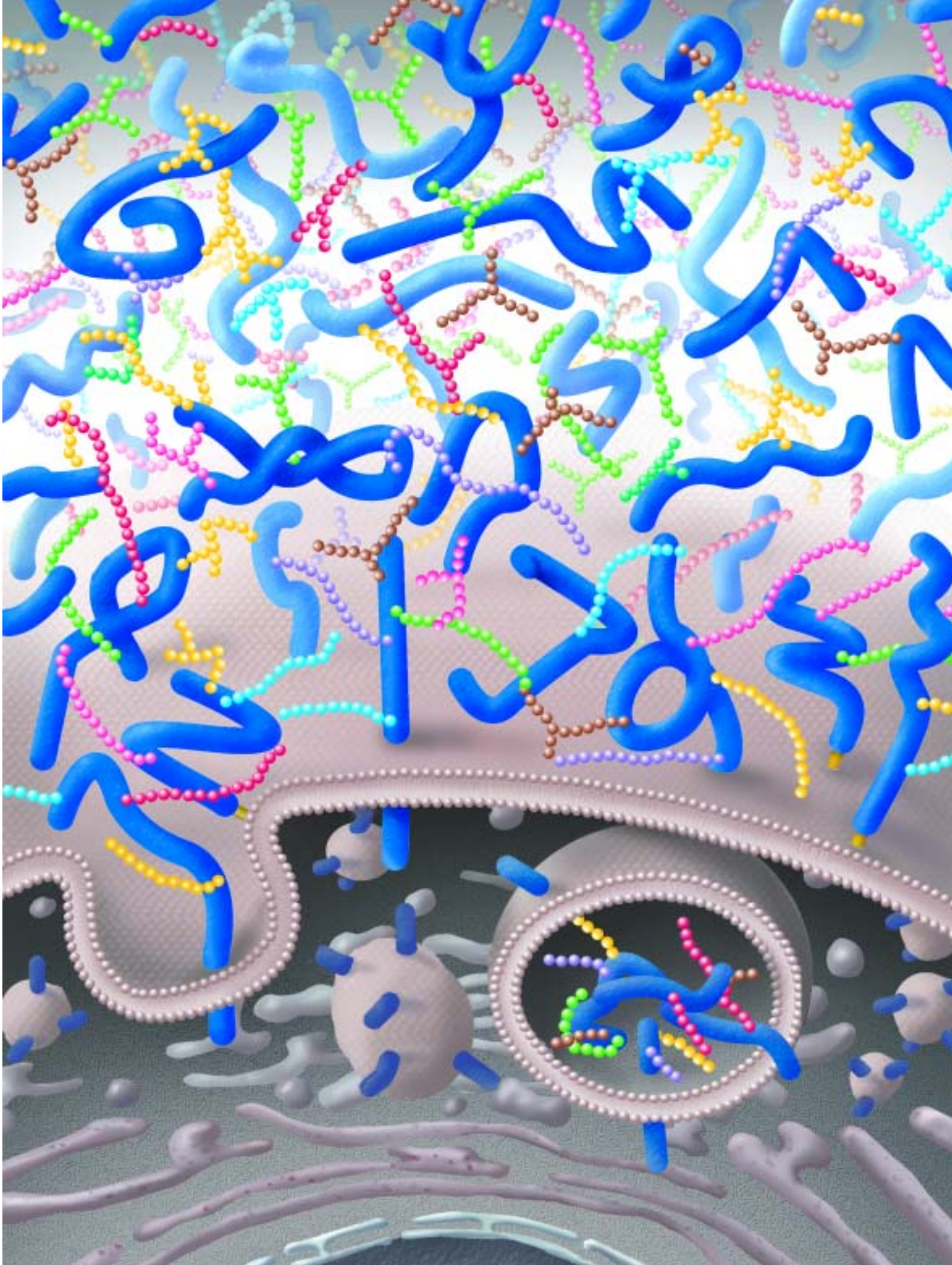
tions include the removal of the two-carbon acetyl groups at the amino position of the glucosamine portion and the addition of sulfate groups at several different positions, along with distinctions in the stereochemical arrangement of bonds around specific carbons. Different combinations of these various chemical modifications make it possible for even short chains to have an enormous number of structural permutations. In fact, the potential for an immense quantity of structural information to be embedded in a glycan exceeds that of nucleic acids or proteins.

Unlike the synthesis of DNA, RNA or proteins, however, glycan synthesis does not depend on a template that codes for the exact sequence of building blocks in a new chain, to be faithfully replicated over and over again as an identical copy. Instead, GAGs are synthesized through the concerted action of a large repertoire of enzymes whose existence and relative activities vary greatly. In short, HSGAG biosynthesis is a multi-step process with multiple enzyme players.

Most of the enzymes involved in HSGAG biosynthesis are now known, but exactly how the process of synthe-

Figure 1. Complex sugars, or glycans, are a structurally diverse class of molecules that critically shape the molecular landscape of cell surfaces and the extracellular matrix (ECM) between cells. Typically these polysaccharides, represented here as multicolored strings of beads, are covalently attached to proteins (shown here in blue), many of which also reside at the cell surface and throughout the ECM. Glycans may be linear chains (glycosaminoglycans) or branched sugars. Also represented here are some of the intracellular components of the cell, such as the Golgi (where these glycans are manufactured) and the lipid vesicles that export glycan-modified proteins to the cell surface.

Ram Sasisekharan is a professor of biological engineering at the Massachusetts Institute of Technology. He received his Ph.D. in 1992 from Harvard Medical School. James R. Myette is a postdoctoral fellow in Sasisekharan's laboratory. He received his doctorate in 1996 from the State University of New York at Buffalo. The Sasisekharan group's research objectives include understanding how the extracellular matrix regulates cell function, developing tools to study complex carbohydrates and developing drugs to treat related diseases. Address for Sasisekharan: Biological Engineering Division, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139. Internet: rams@mit.edu; <http://web.mit.edu/tox/sasisekharan/>



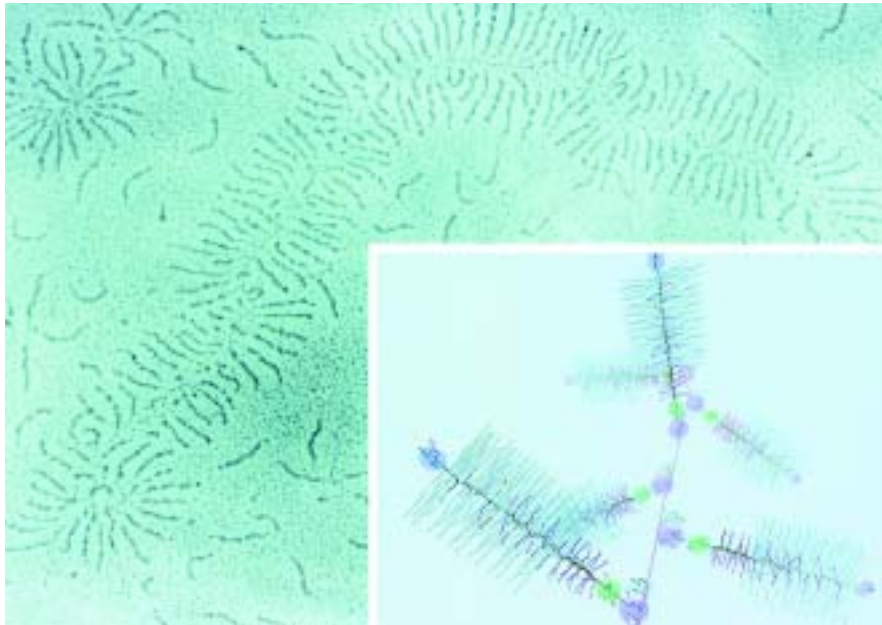
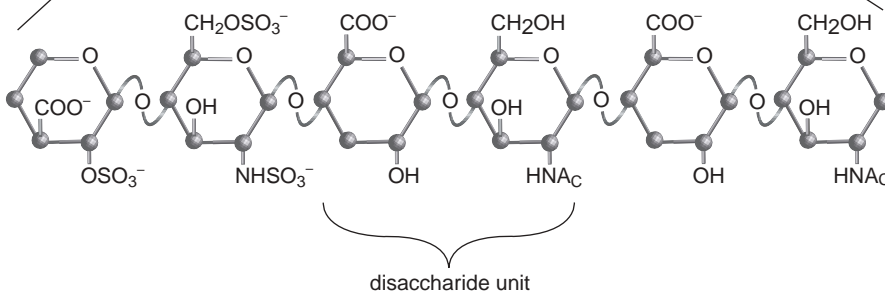
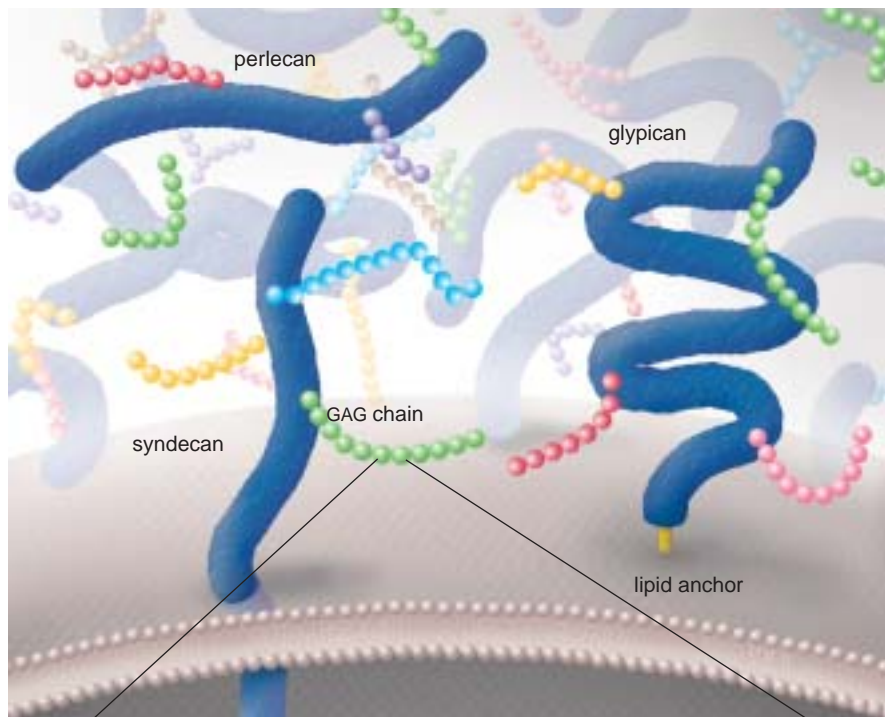


Figure 2. Proteoglycans from a cartilage cell are visible in this electron micrograph. Dozens of proteins can be seen emanating from an extended aggregate; glycans form feathery wisps around the protein columns. The inset models the structure of a few proteoglycans, each containing dozens of glycan chains. (Reproduced from Roseman, Saul, 2001, *Journal of Biological Chemistry* 276:41527–41542, by permission of the American Society for Biochemistry and Molecular Biology.)



sis plays out is still very much an open question. We know little about the ratio of enzymes or, even more basically, whether they act independently or cooperatively in a multienzyme complex. It is known that HSGAGs are made inside the cell in the membranes of the organelles known as the Golgi apparatus. Nearly all the enzymes involved with making HSGAGs either span the organelle's membranes or are at least peripherally associated with them. This arrangement essentially restricts the interaction of these enzymes to two dimensions within a lipid lattice. Although the complete biochemical picture is not yet known, it is likely that the enzymes for HSGAG biosynthesis come together within the Golgi membrane, perhaps as the chain is being assembled.

For the most part, glycans do not exist at the cell surface or in the extracellular matrix (ECM) as free-standing polymers. Rather, they are assembled onto specific proteins to form protein-glycan conjugates, or *proteoglycans*. With the exception of heparin, which is made as a free-standing sugar polymer, HSGAGs are generally found in three major classes of proteoglycans. A major distinction among these proteoglycans may be found in their particular arrangement relative to the cell surface. In *syndecans*, the core proteins cross the cell membrane. *Glypicans* are also inserted into membranes, but by a lipid anchor connected to the core protein. *Perlecans* reside in the ECM. There is much evidence that the particular composition of glycans attached to each core protein is not random.

Structure Determines Function

Proteoglycans are unique and structurally complex macromolecules. A clue to the function of HSGAG proteo-

Figure 3. Heparan sulfate glycosaminoglycans (HSGAGs) are linear, sulfated polysaccharides of varying length and chemical composition. At the most fundamental structural level, these sugars are comprised of repeating disaccharide units (each two simple sugars, made up of chemical groups attached to a six-membered ring made up of five carbon atoms and one oxygen). Added to this basic structure is a level of complexity brought about by numerous chemical modifications that can occur at discrete positions within a chain. HSGAGs generally exist at the cell surface or throughout the extracellular matrix as protein-sugar conjugates of three major classes, depending on where they are found: perlecans, syndecans and glypicans.

glycans comes from the list of important proteins with which they bind in discrete spatial and temporal interactions. These proteins include many key growth factors and growth-factor receptors, proteins involved in tissue and organ development, others involved in immune and inflammatory responses, some that mediate cell adhesion, and so on. Like proteoglycans, the proteins that associate with them generally reside outside cells, either near cell membranes or dispersed throughout the ECM. Many of these proteins circulate in the blood, where they are involved in processes such as blood coagulation, wound healing and tissue repair.

The interactions between glycans and the proteins they bind to reveal connections between structure and activity. These interactions have often been ascribed merely to the noncovalent electrostatic attraction between negatively charged sugars and positively charged proteins. A closer look, however, reveals that many protein-glycan interactions are in fact structurally selective. We offer three examples of such specific interactions—the binding of HSGAGs to antithrombin, to fibroblast growth factor and to herpes simplex virus gD glycoprotein.

In some cases, exquisite structural specificity guides the interaction between HSGAGs and proteins in a way reminiscent of the so-called lock-and-key complementarity between enzymes and their substrates. The binding of heparin to antithrombin III (or ATIII) is a classic example of such an interaction. ATIII is a protein that plays a key role in the cascade of steps that leads to blood coagulation. Clinicians have appreciated the influence of heparin on this process since the early 1930s, when heparin was first used as an anticoagulant during surgery. We now know that when heparin binds to ATIII, this binding induces an important change in the conformation of the protein. In turn, this change greatly increases the inhibitory action that ATIII exerts on certain other proteins that normally promote blood coagulation.

A series of experiments have shown that only a small segment within heparin (which exists as a mixed population of molecules) actually binds to ATIII and induces its conformational change. The minimal active binding sequence is a distinct pentasaccharide (that is, two-and-a-half disaccharide units). However, to trigger as much

anticoagulation as would a full-length heparin molecule, a longer polysaccharide is required, one that can simultaneously bind to the protein thrombin as well as to antithrombin III. Although the HSGAG region that binds to thrombin does not appear to require a precise sequence, its spacing relative to the ATIII-binding region is very important.

This example illustrates two important points about the interaction of proteins with HSGAGs, and probably other glycans. First, the protein-binding region within the polysaccharide is not randomly distributed throughout the chain; rather, it is generally restricted to a limited number of contiguous disaccharides out of the more than 100 that may make up its linear sequence. Second, a single glycan chain often has two or more sites for protein binding. One can think of the glycan, therefore, as a molecular scaffold that promotes the favorable interaction of two or more protein partners.

The example of fibroblast growth factor signaling also elegantly illustrates the concept of HSGAGs bringing proteins together. In particular, the gly-

cans facilitate the interaction of fibroblast growth factor with its receptor at the cell surface. The binding of growth factor to its receptor sets in motion a signaling cascade that ends up in the cell's nucleus, turning on genes that modulate cellular proliferation. To trigger this cascade, a receptor embedded in the cell membrane needs to undergo a structural change, a change that occurs when one receptor interacts simultaneously with a second receptor. It seems that the FGF molecules outside the cell (at least in the case of the growth factor known as FGF-2) must themselves form a *dimer*, or pair, to bring two receptors together on the cell surface. Certain studies have shown that FGF signaling may not absolutely require the presence of the glycan; yet in this convergence of molecules glycans do serve as a sort of glue, holding the entire complex together in the proper configuration necessary for maximal signal transduction.

The minimal HSGAG sequence needed to bind to FGF-2 can be reduced essentially to a region of three contiguous disaccharides (a hexasaccharide) containing at least one iduronate with a

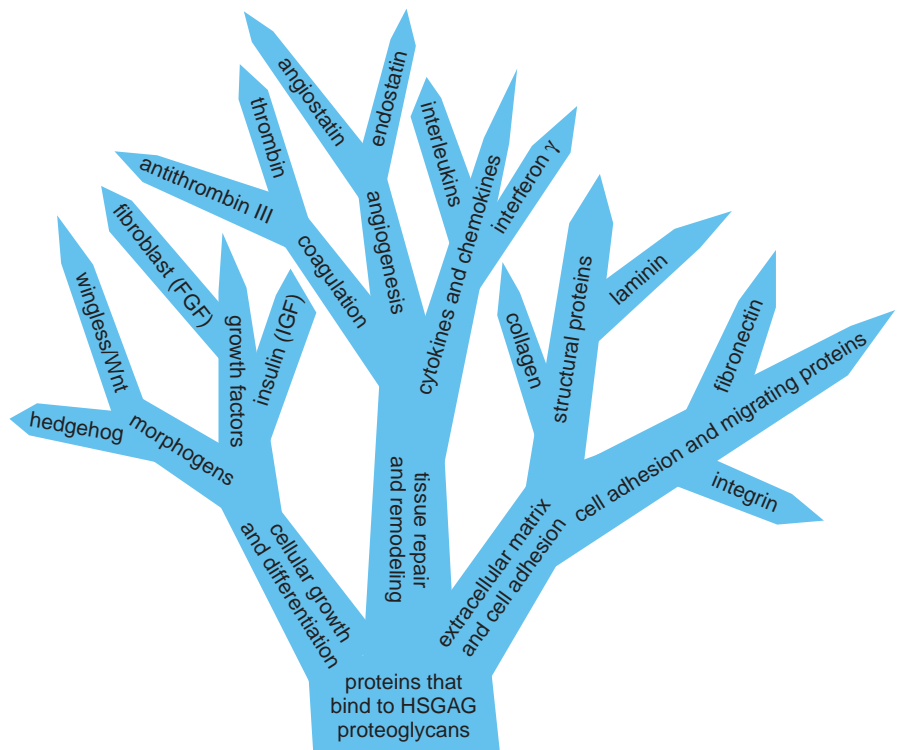


Figure 4. HSGAGs affect the growth and function of cells and tissues largely by interacting with specific proteins in a spatially and temporally orchestrated manner. Among other physiological processes, these protein-glycan interactions affect cellular growth and differentiation, tissue repair and remodeling and the physical communication among cells that takes place across the extracellular matrix. HSGAG-binding proteins involved in some of these critical functions are shown here as branches on a tree.

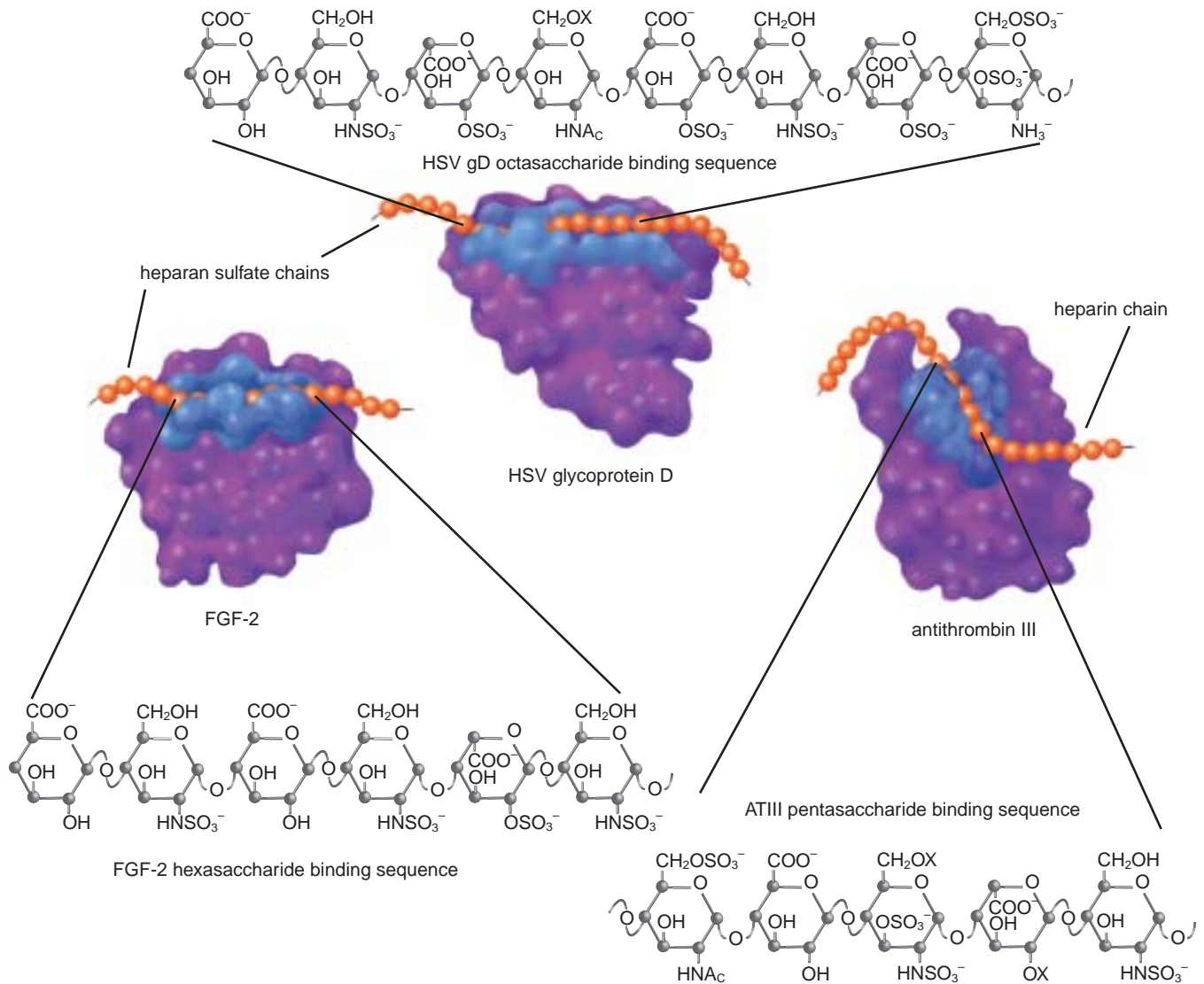


Figure 5. Proteins bind to structurally defined regions within HSGAGs. Three well-studied examples of such specific sugar-protein interactions include fibroblast growth factor-2 (FGF-2), which promotes cell proliferation; the herpes simplex type-1 virus glycoprotein D (HSV gD), which allows the virus to connect to and enter its host cell; and antithrombin III (ATIII), whose binding to heparin triggers blood coagulation. The proteins are shown here using a schematic representation of their molecular structure based loosely on a space-filling model, and the HSGAGs are depicted as beads on a string. The structurally characterized GAG sequence known to bind with each protein is presented in greater detail.

sulfate at the 2-hydroxy position. To facilitate the formation of FGF dimers, this binding sequence must be present in duplicate, perhaps on a single chain. To fully activate receptor signaling, an additional region in the chain is required on the HSGAG, namely one that binds to the FGF-2 receptor as well. The precise ratio of glycan-receptor binding sequences is still not clear.

Experiments do show that, as in FGF binding, a glycan that binds to the FGF receptor requires the presence of a particular sulfate group most likely located in the same hexasaccharide. In this case, the sulfate is located at the 6-hydroxy position of a glucosamine. We like to view this sulfate as a possible

structural switch by which HSGAGs may regulate the growth factor's signaling. Controlling the extent of this chemical modification might be a mechanism by which cells fine-tune this important biochemical pathway. In fact, some research groups have found that the extent of sulfation at the site in question changes as a function of age and in cancerous cells.

Our third example, that of herpes simplex virus gD glycoprotein, shows that the glycan can critically mediate the interaction between people and pathogens. Microbiologists have known for several years that herpes simplex type-1 virus (HSV-1), the cause of cold sores, requires heparan sulfate to infect cells;

cells that have been engineered in the lab to lack heparan sulfate on their surfaces resist viral infection.

Now we know a great deal more about the process. HSV infection occurs in two steps: The virus binds to the cell surface and then enters the cell. For both steps, unique glycans outside the cell interact with specific glycoproteins on the viral envelope. For the binding step, viral glycoproteins B and C interact with cell-surface HSGAGs. For the entry step, viral glycoprotein D (gD) binds to a specific HSGAG sequence possessing a unique sulfate found relatively rarely in nature. The interaction between the human glycans and the viral gD glycoprotein induces

a conformational change in gD that allows the virus envelope to fuse with the host cell's membrane. As it turns out, a cell lacking the enzyme that adds the specific and relatively rare sulfate to the glycan chain will not be infected. In addition to this sulfate, a minimal chain of four disaccharides is needed to bind to gD.

In all three examples above, it has been possible to isolate small portions of glycans that bind to particular proteins. An additional challenging property of glycans in general and HSGAGs in particular is that they mediate their protein-binding activity through multi-valency—that is, many copies or multiple numbers of glycans bind to a given protein in order to mediate biological function. Furthermore, at least some glycosaminoglycans can bend; they are not rigid polymers. This flexibility may help the chains fit around the proteins to which they bind in an interactive way.

Although the structure of a glycan is clearly central to its functionality, recent evidence suggests that an HSGAG's structure is not locked in once a chain has been manufactured in the Golgi membranes. Instead, it can be modified even long after the chain has been transported outside the cell; select enzymes secreted into this extracellular environment can remodel the glycans' architecture. Some of these enzymes degrade HSGAGs by actually cleaving them into smaller fragments. More subtle remodeling enzymes include the extracellular sulfatases. Rather than chopping up the chain, such enzymes remove a single specific sulfate group or, at most, a small number of sulfates. These highly selective enzymes have been shown clearly to affect at least one cell-signaling pathway, that involving the Wnt proteins, a family of HSGAG-binding glycoproteins that are critical as signaling molecules during vertebrate development. The FGF signaling pathway may also be affected by the action of these sulfatases. It is as if HSGAGs can serve as molecular rheostats, their programmed structural alteration serving to fine-tune the strength and timing of discrete extracellular signaling pathways.

In summary, the structural diversity of HSGAGs is not arbitrary or static; rather it is finely regulated. This dynamic design affects the flow of biochemical information across cell surfaces and throughout the extracellular

proteins	normal function	effects of dysregulation				cancer examples
		tumor growth	metastasis	dedifferentiation	inhibition of host defense	
fibroblast growth factor, hepatocyte growth factor, bone morphogenetic protein	cellular proliferation	✓	✓			glioma, breast, lung, liver, prostate, kidney
Wnt, hedgehog	cellular differentiation	✓		✓		basal cell carcinoma, breast, brain, colon
interleukin-8, interferon-γ	immune response			✓		breast, ovarian, prostate, lung
endostatin, vascular endothelial growth factor, thrombospondin	angiogenesis	✓				ovarian, colorectal, bladder, prostate
P-selectin, paminin, fibronectin	cellular adhesion	✓	✓	✓		breast, ovarian, prostate, gastric
collagen, elastase	extracellular matrix	✓	✓			breast, lung, prostate, gastric, colon
heparanase	HSGAG degradation	✓	✓			melanoma, pancreas
glycosyltransferases EXT1, EXT2	HSGAG biosynthesis	✓				bone, liver

Figure 6. HSGAGs often play a critical role in the development or progression of cancers. At its core, this relation goes back to the enzymes that make these sugars or degrade them. The regulation of proteins that bind these glycans is perturbed as a consequence.

matrix. The binding of these glycans to particular proteins fundamentally alters their fate—where they go and what they do. Thus these glycans act as key regulators of important extracellular signaling molecules and thereby fundamentally modulate cellular function.

Glycans in Biology and Disease

Genetics research has been central to our appreciation of the importance of HSGAGs to life. In some cases, this understanding comes from research on human diseases, where the molecules' importance can be painfully evident; more often, it comes from laboratory studies using model organisms such as the nematode (*Caenorhabditis elegans*), the fruit fly (*Drosophila melanogaster*) and the common lab mouse (*Mus musculus*), all of which can be manipulated genetically. These organisms, particularly the fruit fly, have revealed a central connection between HSGAGs and development. Many of the genes important to the production

of certain HSGAGs have been characterized in great detail.

Some human diseases involving HSGAGs result from having an excess of the glycans. Among these are various lysosomal storage diseases (known as mucopolysaccharidoses), where a lack of active HSGAG-degrading enzymes leads to a toxic buildup of glycans that damages tissues. These inherited diseases are disabling and sometimes fatal to children. Other pathologies arise from HSGAG deficiency. One well-studied disease is the bone disorder clinically known as hereditary multiple exostoses (HME). People with this skeletal disorder have many bony protuberances, or exostoses. The genetic cause of the disease is a mutation that disables the glycosyltransferase EXT1 (or, less frequently, EXT2), which is responsible for the extension of HSGAG chains. HME is inherited as an autosomal dominant disorder; as such, having one good copy of EXT1 is not enough to prevent abnormal cartilage growth.

It seems paradoxical that a disease in-

volving a ubiquitous and clearly indispensable function—one of the earliest steps in HSGAG synthesis—has such restricted symptoms, namely cartilaginous bone tumors. Probably the answer lies in the single functional copy of the gene that remains. Having half the normal amount of EXT is not fatal, but apparently some tissues, such as cartilage, are more sensitive to diminished production of HSGAGs. Alternatively, it is possible that other tissues have another, yet-to-be-discovered form (or forms) of the enzyme that compensates for the decline in EXT1 levels.

In between these two extreme disease categories—having an excess of HSGAGs or having a restricted amount—there is a third and more common scenario. HSGAGs may simply be formed in an inappropriate amount or structure. Indeed, the inappropriate or dysregulated expression of glycans is implicated in many common human ailments, from neurodegenerative diseases such as Alzheimer's and possibly Parkinson's disease, to growth dysplasias, cardiovascular disease and even cancer. Among these relevant pathologies, our laboratory has been particularly interested in the connection between HSGAGs and cancer.

Cancer and Infection

Cancer is a complex disease, with diverse origins and clinical outcomes. Nevertheless, nearly all cancers involving solid tumors progress through similar stages. These include oncogenesis, where cancerous cells originate; tumor growth; and, finally, metastasis. The microarchitecture of the extracellular matrix is perturbed at each of these stages. For this reason, it is important to view cancer not merely at the genetic level, but from the perspective of the ECM.

As we have pointed out, HSGAGs bind to a multitude of proteins involved in cell proliferation and differentiation. Without their careful orchestration, uncontrolled cell growth is likely. HSGAGs may regulate cancer by acting either as tumor promoters or as suppressors. Our laboratory made the observation that the glycan coat on cancer cells indeed contains different glycan structures that exert opposing effects on cancer progression. It is likely that this balance can be tipped, potentially serving as a switch to regulate this progression.

Cancer progression also requires the development of a rich blood supply to sustain tumor growth; inhibiting this

Studying Structure

Knowing the fine structure of complex carbohydrates such as the heparan sulfate glycosaminoglycans (or HSGAGs) is clearly critical to understanding their unique biological functions. Until recently, arriving at this knowledge has been arduous and technically challenging. There are several reasons.

First there is the problem of amount. Unlike heparin, which is commercially available in large quantities for medical use, the HSGAGs located outside the cell, biologically more diverse, are present at much lower levels in biological sources. This paucity of material makes them difficult to detect by common analytical methods such as nuclear magnetic resonance (NMR) or liquid chromatography. The issue of limited starting material is not a trivial matter. Although scientists dealing with DNA or proteins can amplify the limited starting material using laboratory shortcuts such as the polymerase chain reaction (PCR) and recombinant gene expression, no such shortcuts exist for glycans.

A second problem is purity. Heparan sulfate derived from cells and tissues is actually a mixture of many structural variants. Purifying these sugars is not straightforward. Polysulfated glycans tend to co-purify with other likewise negatively charged biomolecules (such as proteins and nucleic acids) when extracted from tissue. Multiple separation steps are typically required to obtain pure samples; this process, however, often undercuts yields.

A third problem is that of the molecules' structural stability. The sulfate esters in HSGAG chains are particularly susceptible to fragmentation when high-energy methods, such as mass spectrometry, are used to examine their structures.

The holy grail of HSGAG structure studies remains the ability to determine the linear sequence of disaccharides in an isolated glycan ultimately with a facility comparable to the methods now used to sequence amino acids in proteins and, even more easily, the nucleotides in DNA. Although extending this capacity to glycans is a continuing challenge, our lab and others have made considerable technical strides toward that goal.

The successes thus far in HSGAG sequencing are largely the result of three practical achievements. The first has been in place for several years—the use of chemicals and enzymes that cleave glycans in specific ways. In our lab the enzymes of choice are the heparanases. A “cocktail” of these enzymes can cleave any HSGAG into its disaccharide constituents, from which we can ascertain the overall composition of any HSGAG. Add to this enzyme tool kit a number of select GAG-desulfating enzymes, and the compositional analysis can be refined even further.

The second advance has been the use of highly sensitive instruments to detect and identify these various oligosaccharides. In our lab, for example, we are able to use capillary electrophoresis to separate enzymatically cleaved oligosaccharide fragments on the basis of the number of sulfates per

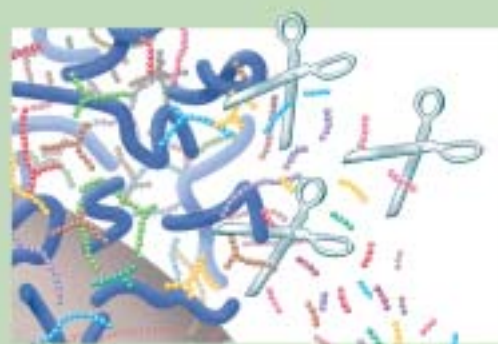
disaccharide unit. We use mass spectrometry (specifically, matrix-assisted laser desorption MS, or MALDI-MS) to further characterize these sugars based on their molecular mass. When used together with our GAG-related enzyme tools, these powerful analytical methods allow us to generate a master list of all structural possibilities for a chain. The next step in this process is to narrow down these possibilities—which brings us to our third, invaluable tool.

The third advance has been the development of a computational approach to annotate all this structural information while ultimately converging on an unequivocal structural solution that is compatible with a set of experimental results. At the core of this bioinformatics approach is a binary coding system, defining the presence (1) or absence (0) of a sulfate or an acetate at a given position in the disaccharide. This information is then incorporated into a higher-base coding system that assigns a single alphanumeric character to describe each disaccharide structural permutation. We believe that such a flexible, *property-encoded nomenclature* can serve as a versatile system for analyzing the structures of a wide range of glycans, including branched ones.

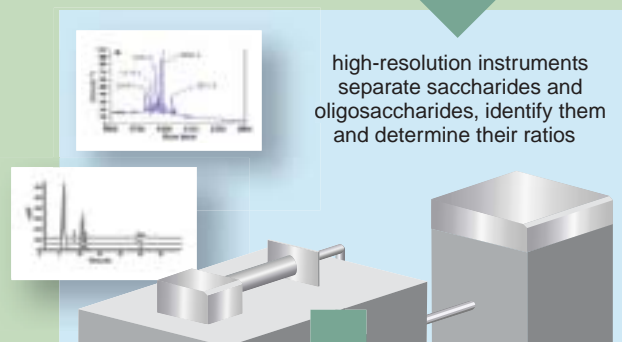
Although mass spectrometry and capillary electrophoresis have been great tools for separating and sequencing complex polysaccharides such as heparin and heparan sulfate, other useful sequencing tools are also being developed. Among these are gel electrophoresis, liquid chromatography and NMR. NMR is extremely useful, especially when used along with one or both of the existing techniques we have described. In principle, NMR allows a more comprehensive analysis by providing information on nearly all structural parameters, including the spatial arrangement of specific chemical bonds within a glycan chain.

Given some of the the intrinsic limitations of the different techniques, it is unlikely that any single method will be sufficient for the comprehensive determination of glycan structure. We have come to appreciate the need for an integrative approach to sequencing complex carbohydrates, an approach that includes high-resolution instrumentation, access to chemicals and enzymes that cleave carbohydrates in structurally specific ways and, finally, an effective computational framework for analyzing experimental results.

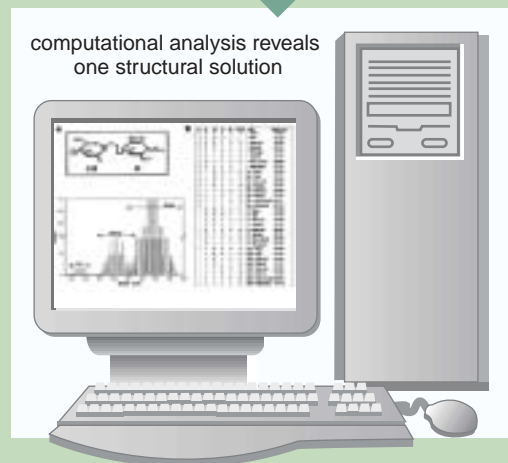
Such an approach could serve as the basis for development of a high-volume screening of sugars that could ultimately become a fully automated process analogous to current DNA microarray technology, which is used to study a large number of genes in parallel. In this not-so-futuristic “era of the glyco-chip” we would be able to effectively link glycan structure with function on a whole-cell and tissue level by monitoring the unique interactions of an array of proteins and polysaccharides. With such a formidable technology in place, there is the exciting prospect of meaningfully connecting the protein world with the more nascent glyco-world research.



chemicals and enzymes
cleave glycans specifically



high-resolution instruments
separate saccharides and
oligosaccharides, identify them
and determine their ratios



computational analysis reveals
one structural solution

Several important technical advances are helping scientists meet the challenge of sequencing heparan sulfate glycosaminoglycans. These include the combinatorial use of chemicals and enzymes able to cleave carbohydrates in chemically specific ways to generate a structurally defined pool of glycan fragments. Sensitive analytical techniques such as capillary electrophoresis, mass spectrometry and nuclear magnetic resonance can be used in tandem to analyze and identify glycan fragments based on their unique chemical properties. A computational framework brings the information together to permit the systematic evaluation of a large volume of data to resolve the molecules' fine structure.

drug name	function	mechanism	clinical status
Lovenox	anticoagulant	binds to ATIII	market
Fragmin	anticoagulant		market
GH-9001	oral anticoagulant (anti-thrombotic)		phase 1
Deligoparin (OP-2000)	treat ulcerated colitis (irritated bowel disease)	unclear	phase 3
Seprigel	anti-adhesive/surgery	restored hyaluronic acid polymer absorbs shock and protects tissues	market
Healon	treat cataracts		
Vevesca	treat Gaucher's disease	restores enzyme that degrades glycolipids, β -glucosceramidase	phase 3 (Europe)
GD0039 (swainsonine)	anticancer (metastasis inhibitor)	inhibits carbohydrate processing by blocking Golgi enzyme α -mannosidase	phase 2
PI-88	anticancer (angiogenesis and metastases)	interferes with FGF-2 signaling by inhibiting heparanase	phase 2
UT231B	antiviral (hepatitis C)	inhibition of α -glucosidase	phase 1
GCS-100 (pectin derivative)	anticancer (angiogenesis and cell death)	inhibition of carbohydrate-binding protein galectin-3	phase 2

Figure 7. Carbohydrate-based drugs are increasingly being developed for the treatment of human disease and for medical procedures. Many of these drugs are designed with specific biochemical targets in mind, along with strategies to affect these targets. The strategies include replacing enzymes involved in carbohydrate biosynthesis or degradation. Other drugs, such as the low-molecular-weight heparins, serve as derivatives of naturally occurring glycans that retain the ability to bind to specific proteins. Still other drugs mimic natural glycans, most often as a means to inhibit unfavorable molecular interactions that lead to disease.

process of *angiogenesis* has been a popular clinical strategy for a few years now. HSGAGs are involved in angiogenesis in complex ways. Some bind to growth factors that promote the proliferation and differentiation of the cells that line blood-vessel walls; others bind to factors that inhibit blood-vessel formation.

The frequently deadly last step in cancer progression is metastasis, where cancer cells break loose, enter the circulatory system and invade tissues far from the site of the tumor's origin. To do this, the cells must overcome at least two restrictions. First, cancerous cells must unstick from surrounding cells and adhere to foreign tissue. Cell adhesion generally requires the interaction of glycoproteins present on opposing cell surfaces or within the intervening ECM; two important classes of

these glycoproteins are selectins and integrins. This molecular handshake is mediated by specific surface glycans, and these interactions are perturbed in cancerous cells.

The second restriction to metastasis is that layers of tissue typically intervene between a tumor and a nearby blood vessel. To reach the interior of the vessel, tumor cells have to push through the basement membrane, a GAG-rich barrier. Doing so requires several different types of enzymes, including ones that degrade glycans. Among these are the heparanases, which cleave polysaccharide chains at specific, internal positions. It has been observed that heparanases are typically more active in cancer. Digging deeper reveals interactions between HSGAGs and cancer progression that

are not always straightforward. This is a very active area of investigation with much promise.

Glycans factor heavily in the etiology of many other diseases as well. In microbial infections, glycans frequently serve as gateways to invaders. Many of these pathogens besides herpes simplex virus type-1 have evolved specific surface molecules that facilitate their attachment and, sometimes, provide the means for the germs' entry into tissue-specific cells. These pathogens include the yellow fever virus and those that cause malaria (*Plasmodium*), whooping cough (*Bordetella pertussis*), certain sexually transmitted diseases (*Neisseria gonorrhoeae* and *Chlamydia*), gastric ulcers (*Helicobacter pylori*) and tuberculosis (*Mycobacterium tuberculosis*). Needless to say, the payoff for understanding the structural basis for these kinds of microbial-glycan interactions may be immense.

Glycomics

Now that the genomes of several species, including human beings, have been sequenced, the focus in molecular biology is increasingly turning from DNA to proteins. Proteins are considered, by convention, the end product of the "central dogma," which states that DNA leads to RNA leads to protein. But many important details are missing from this biochemical information flow. Consider the following observation. A recent annotation of the Human Genome Project has identified approximately 30,000 genes that presumably encode proteins. A strict interpretation of the paradigm that one gene codes for one protein means that the human body, in its almost unbounded complexity, is governed exclusively by the protein products of these 30,000 genes. From this limited perspective, it would seem as though there just are not enough proteins to execute the daunting task of forming and maintaining our bodies.

Of course, few biologists would embrace this simplistic view of protein function. In reality, proteins are dynamic polymers whose function is regulated on many levels. Many of the additional levels of protein regulation—for example, modification of their structure after translation from the DNA code or their distribution in complex ways in space and time—largely depend on glycans. Through regulated glycan structure and function, complex organisms can over-

lay an additional level of complexity onto the function of the finite number of proteins our genomes encode. We must now embrace a modified view of the central dogma in which glycans serve to bridge the gap between genes and the proteins they encode.

We have described how one set of glycans—the glycosaminoglycans—interact with proteins and regulate them. We have, however, only scratched the surface. There are many other glycan classes that are also complex in structure and diverse in their biological function. As with the GAGs, the biochemical pathways by which these other glycans are made are also complicated and regulated intricately among different tissues. Taken together, there are an enormous number of empirical parameters needed to describe the structure, distribution and function of these various carbohydrates. The comprehensive study of these molecules—from glycan structure and biochemical function to their genetics, physiology, medicine and, ultimately pharmaceuticals—requires a highly integrative approach with effective organization among diverse areas of glycan research and various research groups. From these needs, the field of functional glycomics has emerged.

This network has grown into a national consortium that operates under the auspices of the National Institute of General Medicine at the National Institutes of Health. To reach its full potential, this project requires a highly integrated way to annotate, process and disseminate a vast amount of information. A primary challenge facing the glycomics project, therefore, is developing the ability to link and integrate effectively this great volume of multidimensional experimental data in a format that possesses an accessible and flexible user interface. At MIT we are developing a comprehensive or “systems biology” interface for the consortium.

Carbohydrate Drugs

Given the indispensable roles glycans play in human physiology, we naturally believe this family of biomolecules has great therapeutic potential, both as drug targets and as drugs in their own right. In fact, several carbohydrate-based drugs are already on the market, and many others are in various phases of clinical trials. The majority of these drugs affect specific cell processes that are known to be mediated by glycans.

The strategies for their development essentially fall into two broad categories—restoration and intervention. Restorative therapies seek to replace carbohydrates that the body fails to synthesize, at least at adequate levels. Typically this restoration centers on replacing an enzyme deficiency. Alternatively these drugs can be derived from discrete bioactive regions of glycans or a structurally related chemical analogue (or *mimetic*), which can bind to the same target proteins as the native glycans. This strategy is a challenging one. Pinpointing the bioactive region of a glycan is a formidable task because of the large number of structurally related glycans that are also present within a given tissue. Furthermore, as stated earlier, the multivalent nature of their interaction with proteins imposes an additional level of complexity.

Currently, drugs that intervene in glycan-dependent disease processes represent a more practical strategy, especially as these drugs are used to treat highly relevant pathologies such as cancer, cardiovascular disease and microbial infection, to name a few. Carbohydrates and mimetics could be used to inhibit unfavorable molecular interactions that, for example, lead to cancer cell metastases or the docking of pathogens onto host cells.

For very practical reasons, sugars hold enormous promise as drugs compared with protein-based therapies. Sugars are smaller than proteins, and they are more stable and more easily formulated for delivery into the body. They are also probably less likely to trigger an immune response. On the other hand, the manufacture of carbohydrate-based drugs does pose technical challenges. To begin with, specific glycans cannot be readily manufactured like DNA or proteins, at least not yet. Chemical synthesis is arduous, is currently practical only for smaller carbohydrates and is difficult to mount on a large scale.

To circumvent some of these technical obstacles, a second manufacturing strategy involves modifying pre-existing polysaccharides. This strategy affords the obvious advantage of bypassing the synthesis of the drug essentially from scratch. The quintessential example of this approach is heparin, which, as we mentioned earlier, is used as an anticoagulant, especially after surgery. It is generally administered as an unfractionated mixture of GAGs with var-

ious molecular structures. In this form, heparin is not optimal as a drug. But we now know a great deal more about the parts of heparin that give it its potency. By separating the mixture of GAGs and characterizing their structure-activity relations, we can use this information to rationally design new drugs. Ultimately, one should also be able to manicure these saccharides at specific chemical positions using a combination of chemical and enzymatic clipping tools. Using structure-specific HSGAG-desulfating enzymes, for example, scientists may in time be able to tailor-make sugars.

Carbohydrate drugs may in many cases supersede more traditional approaches to rational drug design such as gene therapy, which have been attractive in theory but have frequently proved impracticable and risky. Moreover, we have already noted just how critical carbohydrates are to so many different biochemical pathways and disease processes—among them, cancer, angiogenesis, tissue repair, skeletal development, cardiovascular disease and microbial infection. As such, they are logical candidates for rational drug design. A thorough understanding of carbohydrate structure and function will lead to a new generation of precise and effective medications to treat many of these ailments.

Bibliography

- Bernfield, M., M. Gotte, P. W. Park, O. Reizes, M. L. Fitzgerald, J. Lincecum and M. Zako. 1999. Functions of cell surface heparan sulfate proteoglycans. *Annual Review of Biochemistry* 68:729–777.
- Conrad, H. E. 1998. *Heparin-Binding Proteins*. San Diego: Academic Press.
- Maeder, T. 2002. Sweet medicines. *Scientific American* 287(July):40–47.
- Renato, V. 2001. Heparan sulfate proteoglycans: intricate molecules with intriguing functions. *Journal of Clinical Investigation* 108:165–167.
- Schmidt, K. 2002. Sugar rush. *New Scientist* 176(October 26):34.
- Travis, J. 2002. The true sweet science. *Science News* 161(April 13). <http://www.sciencenews.org/20020413/bob10.asp>
- Varki, A., R. Cummings, J. Esko, H. Freeze, G. Hart and J. Marth. 1999. *Essentials of Glycobiology*. New York: Cold Spring Harbor Laboratory Press.

For relevant Web links, consult this issue of
American Scientist Online:

[http://www.americanscientist.org/
template/IssueTOC/issue/400](http://www.americanscientist.org/template/IssueTOC/issue/400)