

# Metabolism: sweeter paths in glycoscience

Vivien Marx

Carbohydrates are tough molecules to study, but glycoscientists are developing and democratizing the needed tools.

Too much sugar is tough on the waistline and can lead to health problems, but researchers in glycobiology just can't get enough of these branched molecules. Fortunately, sugars, also called glycans, are everywhere.

Mammalian cells have a 10–100-nm thin sugary coat. Long thought to be a candied fortification, this coat appears to be an information-rich forest of sugar molecules that look like branched, swaying trees, as Stanford University chemist and biologist Carolyn Bertozzi describes them in her talks. The language of these trees fascinates Bertozzi and other glycoscientists.

The cell's sugar layer is akin to “a living sea” on the cell's exterior, says Pamela Marino, a biochemist at the US National Institutes of Health (NIH) who directs the National Institute of General Medical Sciences (NIGMS) biochemistry and bio-related chemistry branch. This living sea is where many processes such as cell–cell recognition and signaling take place. Besides their role in the body's energy metabolism, sugars have intracellular roles; *O*-GlcNAc is important in gene expression, for example. Sugars appear to be involved in development, infection, inflammation, cancer and neurodegenerative diseases.

Plenty has yet to be discovered about the structural diversity of glycans and the functional roles of glycosylation, the patterns of the different sugars attached to molecules such as proteins or lipids, says Douglas Sheeley, who heads the NIGMS biomedical technology branch. Many cellular proteins are glycosylated as they are expressed and packaged in the cell, with numerous enzymes taking part. The glycosylation patterns shape protein function in ways scientists are still learning about.

Methods for studying sugars have steadily matured but have not galloped ahead

as in genomics or proteomics. “Carbohydrates are hard molecules to study,” says Sheeley. Individual monosaccharides in a sugar polymer closely resemble one another; the monomers can link in different ways—attachments of different types can occur at each carbon on a monosaccharide ring; sugars are branched, not linear like DNA or RNA.

Glycobiology labs have choices of tools: biochemical methods, arrays, mass spectrometry and combinations thereof. But non-glycoscience labs are often stymied by how long it can take to learn available methods. Tools are needed for experts and non-experts alike, note Bertozzi and Krishnan Palaniappan of the Google-owned Verily Life Sciences, for the community to strive toward obtaining a “complete parts list” of the human glycoproteome and to learn which proteins are glycosylated at which sites and under which conditions<sup>1</sup>. Funders, too, want to advance tool development and democratize tools. “The hope is to create tools that are straightforward to use,” says Sheeley.

## Tough temptation

Motivating labs to tackle carb complexity are sugars' diverse biomedical roles. The body's mucosa, such as the lining of the lung in people, contains chain-like mucin molecules tipped with sugars such as  $\alpha$ 2,6-linked sialic acids; in birds it's  $\alpha$ 2,3-linked sialic acids. When the avian flu switches its specificity to



A. Doerr, Nature Research

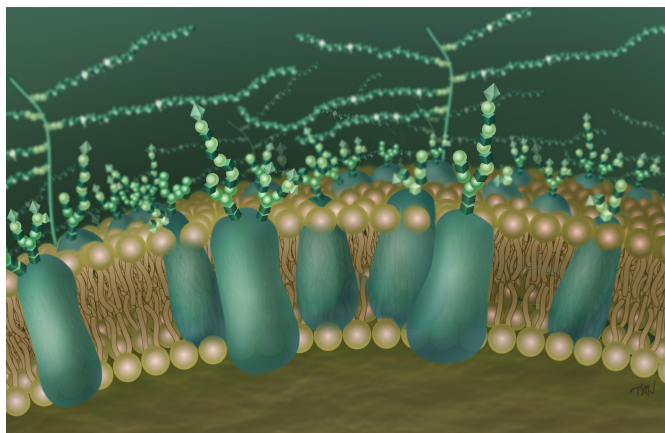
Life is sweet, especially in glycoscience. The human body could harbor as many as 10 to 100 million different glycoproteins.

$\alpha$ 2,6-linked sialic acids, it can infect humans, says Marino.

Viruses can grab on to sugars on a cell's surface, sometimes evolving ways to latch on to multiple sugar branches to secure their grip. Some bacteria, among them several human pathogens, latch on to sugars, too, as do bacterial toxins. Some viruses, including HIV, use sugars to shield their surface proteins from the immune system's antibodies.

When working on genes and proteins, labs can synthesize the molecules they need, sequence them, manipulate them, use model systems to study them and store results in databases. Glycoscience is not at that point yet, says Marino. “But we've been making good progress.”

Multiple funding programs at the NIH and the National Science Foundation have glycoscience components; the National Institute of Standards and Technology has a carbohydrate division, the Defense Advanced Research Projects Agency is developing a glycoscience program, and the [NIH Common Fund](#) that Sheeley and Marino co-lead with NIH colleagues finances glycoscience



T. José, New England Biolabs; A. Vartanian, Nature Research

Mammalian cells are covered in a forest of sugar molecules. Much is awaiting discovery about the structural and functional diversity of glycans.

projects. Glycoscience programs are also being funded in Canada, Europe, Australia and Japan.

Improved methods are emerging as outlined in roadmaps such as the one drawn up by the National Research Council of the National Academies of Science in collaboration with scientists around the world<sup>2,3</sup>. Some of those tools and methods include carbohydrate synthesis and sequencing, both of which have automation potential, says Marino. Labs look at individual glycans and glycoproteins in different ways: they might digest a glycoprotein into smaller pieces, deglycosylate it and use mass spec to see which glycans were detached; they might study deglycosylated peptides to investigate proteins, or they can look at the intact glycoprotein.

Glycan arrays are a high-throughput approach for exploring glycan-binding partners from proteins to microbes. Arrays have been enabling tools, says Marino, and new types are under development. A team at the University of California at San Diego is working on a large sialic acid array; at Emory University scientists are developing glycan arrays that barcode glycans with oligonucleotides so that labs can use genome sequencing to characterize how glycans are bound to proteins. These and other tools are developed also with nonspecialist labs in mind.

Bertozzi and her team are working on easier ways to track *O*-GlcNAcylation, which traditionally involves time-consuming and laborious immunoprecipitation and western blot interpretation. Separately, she is working on mass spec and labeling approaches to profile glycoproteins. The idea is to improve glycoprotein profiling when study-

ing tumor types and disease pathways. Researchers at Eastern Virginia Medical School have applied an approach from her lab that uses mass spectrometry and sugar analogs<sup>4</sup>. The researchers targeted sialoglycoproteins on the surface of prostate cancer cells and compared metastatic and nonmetastatic cells, identifying nearly

### Mass spec ways

Biochemical approaches to identify proteins attached to a sugar tend to involve manual steps, as Bertozzi and Palaniappan point out. NMR spectroscopy, liquid chromatography and electrophoresis can also be used to characterize glycans associated with proteins. But, they note, sample complexity, among other factors, can make it hard to use these approaches. Liquid chromatography with mass spec can be a powerful alternative. Metabolic labeling with sugar analogs does not perturb cell physiology; it allows researchers to pull out glycans and then use mass-spectrometry-based analysis to identify altered proteins.

Chemical tools can also help to address the complexity and heterogeneity of glycosylation, which still present formidable obstacles. Glycoproteins can be present as low-abundance, complex mixtures of glycosylated variants in a sample, making them hard to find. Different sugars can attach to one site on a protein, and each sugar monomer can attach in different ways.

Glycans can suppress ionization in the mass spectrometer. Bertozzi devised another labeling technique to work around that constraint based on inserting a dibromide motif. To make fragmented glycopeptides easier to identify, the researchers label cells with unnatural sugars and tag them with dibrominated probes. She and her team want to make this labeling technique easier for a wider community to enable its use in core facilities.

Mass spec is routine for work with proteins and peptides, but not in glycoscience,

says Julian Saba, a chemist, glycoscientist and workflow developer at Thermo Fisher Scientific. “I think we’re starting to get there,” he says, as tools from academic labs and vendors are inching their way out of specialist labs. One peptide can have 2 or even 100 glycans attached. These different glycoforms split the mass spec peptide signal. Cancer cells can be glycosylated in various ways and the pattern can change—for example, in metastasis—he says, so labs want to know both glycan and protein.

Out of curiosity, Saba looked at raw mass spec proteomics data in such public repositories as PRIDE, the proteomics identifications database. He saw that between 2% and 20% of data from cancer cells are glycosylation data, but the spectra are neither good nor easy to interpret. Even in targeted experiments, the ions from glycopeptides are too sparse to deliver good data and the signal intensity is low compared to that achieved with nonglycosylated peptides.

Most mass spectrometers generate ions by collision-activated dissociation (CAD). But, says Saba, CAD fragments the glycans more than the peptides and breaks the bond between glycan and protein, so a lab cannot be sure which protein the glycan was attached to. Glycoproteins require special sample prep such as enrichment strategies, and software for data analysis. “Unfortunately, a core lab might not be equipped for it,” he says. Tools are emerging as glycoscience grows. For his company’s mass spec instruments, glycoproteomics is the third most popular application, behind classic peptide identification and quantitative proteomics workflows.

Enzymes can be helpful for releasing glycans, but they can also hinder some workflows, says Saba. When a glycan is released from a glycoprotein with an enzyme such as PNGase F, the asparagine to which the glycan was attached is converted to aspartic acid. That’s a mass shift of around 0.98 that labs can watch for as a deamidation, indicating that the particular peptide is glycosylated. The challenge is that a buffer can also trigger this deamidation spontaneously in sample prep. That is why he recommends not releasing the glycan from the glycoprotein. “Don’t remove it,” he says. “You’ve lost information about the glycan.” Fragmentation approaches such as electron-transfer dissociation (ETD) and related methods fragment the peptide, leaving the glycan untouched.

Mass spec gets challenging when labs want to parse the glycan more closely, says Saba. When the glycan is attached to the protein, mass spec does not let labs determine its structure, because they can't discern the linkages between the sugar monosaccharides. This is when releasing the glycan from the peptide helps. Mass spec is not yet a broadly applicable technique for studying glycan complexity. Three different amino acids in a peptide can combine in six different ways, but with three monosaccharides, the number of combinations ramps up to nearly 20,000 possibilities, says Saba. He and his team want to make glycoscience workflows more accessible, in glycoproteomics and, eventually, as the field progresses, in glycan analysis.

### Glycome projects

The word "glycobiology" originated with University of Oxford researcher Raymond Dwek in the 1980s, and over time the community began using "glycomics" to distinguish itself from other 'omics areas, says Harvard Medical School researcher Richard Cummings, a biochemist and glycobiologist. It's been hard to synthesize glycans in the lab, which has hampered glycoscience, but he and his former team at Emory University found that household bleach releases glycans from tissue well<sup>5</sup>. The ability to release, isolate and characterize glycans will power the catalog of the human glycome that he has begun assembling with the goal of teasing out glycomic diversity in health and different diseases. People want to know how the glycome changed to address, for example, how disease might change the glycome of platelets or immunoglobulins. He is also working on 'smart' anti-glycan reagents for identifying glycans and proteins with immunohistochemistry and flow cytometry. These reagents will help scientists build the Human Glycome Atlas that will document the spatial distribution of glycans.

Cummings has separately started a project to profile what he calls the "anti-glycome," which is the immune system repertoire reacting to carbohydrates that pathogens might present to our immune system. Another aspect to pursue: comparative glycomics across plant and animal species. Glycoscience has moved from the silver age to the present golden age, with the platinum age on its way, he says. "I started in the pottery age," he says, laughing, and he's happy so much progress has occurred since then.

Early in the 20th century, carbohydrates were intensely studied, but they later took a

back seat to DNA and proteins, says molecular biologist and glycoscientist Gordan Lauc of the University of Zagreb. Just as microbiome studies struggled for attention, glycoscience's day will come.

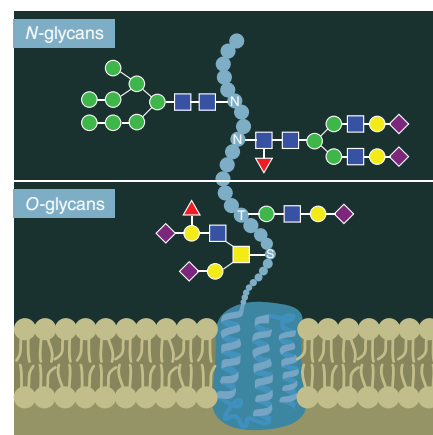
Pharma's heightened interest will help glycoscience, says Lauc. Companies want monoclonal antibodies to be consistently glycosylated to avoid batch variability. Slight changes in culturing conditions such as an oxygen-level shift can alter glycosylation.

As Marino explains, biopharmaceutical therapeutics are often based on recombinant proteins, which are glycosylated in the cells used to produce them. Inappropriate glycosylation can change the half-life of these proteins in the body or affect these proteins' immunogenicity, she says. There are opportunities for engineering, such as tuning the interactions between monoclonal antibodies and the immune system by choosing the carbohydrates with which to decorate the antibody surface.

Possibilities might be far-reaching, says Lauc, as engineers explore how to glycosylate proteins with a wide spectrum of desired characteristics. Applied glycoscience depends on basic research in glycoscience. Lauc has also founded a company, Genos, where his 40-member lab is located. The lab handles some contract research but mainly has EU grants and collaborations with US-based teams. He was part of the Euroglycoscience Forum, a five-year project to link European glycoscience labs that ended in 2014. A follow-on project focuses on the human glycome with a view to the biomedical role of glycosylation variability.

The human body could harbor as many as 10 to 100 million different glycoproteins, says Lauc. "We are not even close to having the entire human glycome," he says. He is profiling the glycome associated with the glycoprotein immunoglobulin G (IgG), a model he likes for its importance and well-explored functions. For example, glycosylation tunes IgG's inflammatory response. He is analyzing IgG glycome data from 30,000 people and plans to ramp up to 100,000 people. He uses samples in biobanks across Europe, including samples from 4,500 people in the TwinsUK study.

Lauc and his team link glycan data to phenotypic, genetic and biochemical data already gathered on these individuals to explore how the glycome might change with age or lifestyle, and to study glycosylation in conditions such as chronic inflammation, lupus, hypertension and cancer.



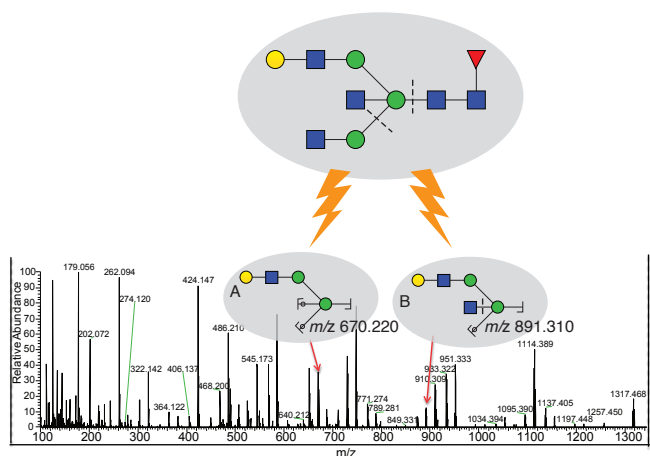
Glycans can attach to proteins at specific sites. *N*-glycans link to the amino acid asparagine (N), and *O*-glycans link to threonine (T) or serine (S). When analyzing glycoproteins, researchers use various ways to determine which proteins are glycosylated where and under which conditions.

For high-throughput glycan analysis, Lauc's team combines analytical chemistry and bioinformatics. Sample quality is important. Chemists will want to connect with clinicians to access large sample numbers and avoid statistically underpowered studies. He applies statistical tools, mass spectrometry and liquid chromatography with fluorescent labels. Chromatography is his lab's workhorse, he says, "because it has the smallest measurement error." Although mass spec instruments are powerful, Lauc finds it harder to obtain consistent quantitative results with mass spec when analyzing large numbers of samples.

Given that protein networks have varying types of glycosylation, the human glycome adds an informational layer to genomic information. There may be more layers: Lauc is currently exploring how changes to epigenetic markers affect glycosylation. The genomic data that he uses remain in the databases of the investigators responsible for each cohort, but his lab's glycan data sit in his lab's servers. Besides the challenge of combining such diverse data for analysis, it's tough to share data in glycoscience, says Lauc. For example, the community is still working on naming conventions for individual glycan structures, "so it's really not easy to read the literature," he says.

Although there are many repositories for glycoscience data, there is no central GenBank-type database. "We need some kind of public repository," says Lauc. (Please see "Computable Sugars" on [Methagora](#) for some glycoscience resources.)

T. José, New England Biolabs; E. Dewalt, A. Vartanian, Nature Research



Thermo Fisher Scientific; E. Dewalt, Nature Research

Mass spectrometry is not yet routine in glycoscience. Here, a glycan released from immunoglobulin G is ionized, and the resulting peaks help researchers identify each fragment.

### Data-juggling

Shared data have to be comprehensible to more people than the postdoc who generated them, says Raja Mazumder, a bioinformatician at George Washington University. He and William York at the University of Georgia have both received pilot funding in glycoscience informatics, and after surveying scientists around the world, they want to jointly build a community resource for computational analysis in glycoscience.

To characterize changes associated with cancer, labs want to look at glycosylation across cancer types, as they do in cancer genome research projects. But with glycosylation data in hand, such as mass spec spectra of deglycosylated peptides, researchers hunt for needed resources often without knowing what might be available among the protein databases, pathway databases and structure databases. “It’s very complicated,” says Mazumder.

Better paths to biological understanding will come from connecting the computational resources on sugars to the vast existing genomic and gene expression data. When a lab result indicates that a sugar is attached at position 92 on a glycoprotein, the team might next explore whether breast cancer samples show mutations at that site and whether such mutations lead to a loss of glycosylation.

Alternatively, a lab might find that samples from healthy people lack a sugar at a glycoprotein site that is glycosylated in cancer samples. The researchers might want to check whether the gene of interest is conserved in mice. In a protein database such as UniProt, a lab might be able to see which

amino acid is at a given position. But that will not give them the corresponding genomic data; that database is just a snapshot rather than a view of genetic variation between individuals, says Mazumder. Proteins and glycans might also vary depending on the tissue of origin. “It’s very, very hard to really traverse the different types of information that is available and make a

coherent decision,” he says. That is one aspect he hopes to be able to address with the community informatics resource he intends to build with the University of Georgia.

### Harnessed enzymes

Biotherapeutics and recombinant glycoproteins are indeed motivators for labs working on glycans and glycoproteins, says glycoscientist Paula Magnelli at New England Biolabs (NEB), where the field has long been of interest, she says, including to the company’s founder Donald Comb. Other influences come from many different areas of biology where labs encounter sugars they want to understand. The National Research Council roadmap highlights, among other aspects, a community-wide need for an enzymes toolbox to help labs study glycans. A number of academic resources have emerged, she says, such as the repository of glyco-enzyme expression constructs at the Complex Carbohydrate Research Center of the University of Georgia. Additionally, companies such as Sigma-Aldrich, Thermo Fisher, Amsbio and MP Biomedicals, as well as NEB, sell enzymes for studying glycans and glycoproteins.

Enzymes are helpful because labs cannot manipulate glycans as they can genes or proteins, says Stephen Shi, an NEB chemist. Enzyme-oriented techniques for carbohydrate analysis include the use of glycosidases that break glycosidic bonds in specific places, endoglycosidases that cleave a glycan from the peptide backbone and exoglycosidases that cut the glycan into smaller pieces. Such enzymes can be used to determine a glycan’s

identity and perhaps help engineers build a predefined glycan.

Many of the industry’s enzymes for glycoscience stem from work in academic labs in the late 1970s and early 1980s. Enzymes like exoglycosidases were sourced from animal tissues, such as bovine kidney and testes, says Shi, and some products were undercharacterized enzyme mixtures. Four years ago, NEB expanded its glycobiology products, and it uses its standard recombinant manufacturing, purification and characterization scheme to avoid impurities and batch variation. When shopping for enzymes, he says, labs should look for well-characterized enzymes and know how the reagents were expressed and purified.

Given the wealth of enzymes and the trial and error that can accompany experimental design, the company is developing kits of standardized enzyme combinations, says Beth McLeod, an NEB researcher. An experiment might involve a 96-well plate with a different enzyme combination in each well, says Magnelli. The target of interest will be digested differently in each well, which gives labs structural indications about their glycan. When developing enzymes, the NEB team keeps workflow in mind; some enzyme formulations are incompatible with mass spec, says McLeod.

Sugars are Magnelli’s passion, and the same is true for other glycoscientists eager to learn about glycans in the dynamic context of living systems. Genome analysis tells researchers which genes are present, but a deeper understanding is needed to address questions about things such as the dynamics of cancer cells or the development of Alzheimer’s disease. Glycans tend to be attached to proteins, says Magnelli, and in both organisms and cells “the DNA is the script but the proteins are the actors.” Gaining a better understanding of proteins in a living system means understanding the role of sugars because, she says, “proteins don’t come naked.”

Vivien Marx is technology editor for *Nature Methods* ([v.marx@us.nature.com](mailto:v.marx@us.nature.com)).

1. Palaniappan, K.K. & Bertozzi, C.R. *Chem. Rev.* **116**, 14277–14306 (2016).
2. National Research Council. *Transforming Glycoscience: A Roadmap for the Future* (National Academies Press, 2012).
3. Doherty, C. (ed.). *A Roadmap for Glycoscience in Europe*. (2015).
4. Yang, L., Nyalwidhe, J.O., Guo, S., Drake, R.R. & Semmes, O.J. *Mol. Cell. Proteomics* **10**, M110.007294 (2011).
5. Song, X. et al. *Nat. Methods* **13**, 528–534 (2016).