

A 3D structure of a glycoprotein spike that helps coronaviruses attach to cells. The spike is decorated with glycans that stop antibodies binding.

X. XIONG ET AL./J. VIROL.

THE CELL'S SWEET SIDE EXPOSED

The secrets of glycans – large sugary chains attached to proteins – are finally coming to light. **By Jyoti Madhusoodanan**

Just 22 amino acids are all that's needed to make all the world's proteins. Four nucleotide bases encode biology's blueprints in DNA. But when it comes to another, equally crucial, class of biomolecules called glycans, scientists don't even know if there is an equivalent alphabet that the cell uses to make them, says bioinformaticist Jaya Srivastava of the Indian Institute of Technology in Mumbai.

Glycans are sugar-based polymers that coat cells and decorate most proteins, forming glycoproteins. They are crucial for biological processes such as immune regulation and intercellular interactions. This makes the apparent lack of a glycan alphabet¹ surprising, and reflects an enduring issue: just how little scientists know about sugars.

More than 30 years ago, chemist Carolyn Bertozzi was astounded by the paucity of chemical information about glycoproteins. At least half of all mammalian proteins are

glycosylated – meaning they have at least one glycan attached. Without the correct sugary suffixes, proteins misfold or become unstable or non-functional. “The biological importance of glycans was well established by the 1980s,” says Bertozzi, now at Stanford University in California. “But it was very hard for biologists to answer any questions in glycoscience, because they didn't have the tools.”

Proteins and DNA could easily be manipulated in the lab, but that wasn't true of glycans. As a result, studies of sugars have lagged behind research into other macromolecules. This is in part because glycans are not synthesized using any known template, and because they can change dynamically depending on a cell's metabolic state. What's more, sugar isomers – molecules with the same chemical formula but different structures – can be used to build varied glycans, but are almost impossible to tell apart on the basis of molecular weight alone.

In 2015, the US National Institutes of Health established the Common Fund Glycoscience programme to develop overarching technologies for studying glycans in biomedicine. At the time, researchers identified a lack of tools as the greatest hurdle in glycobiology. Now, they're beginning to address it.

Bertozzi and others have pioneered methods to image glycans in living or fixed tissues. Thanks to improvements in mass spectrometry and Raman spectroscopy, researchers can more easily identify and characterize glycoproteins. Several scientists, including Srivastava, are developing open databases – such as UniCarbKB, GlyTouCan and the Glycan Mass Spectral Database – that can be used to identify sugars and common glycosylation sites on proteins. Others have focused on high-throughput techniques, including arrays that capture data from hundreds of glycans or glycoproteins at once.

“Things that used to take an entire PhD can now be done in a matter of weeks,” Bertozzi says. “To me, this feels like an inflection point for the field.”

Sugar spotting

When Bertozzi set up her first lab at the University of California, Berkeley, in 1996, she began to work on a fundamental tool: a way to visualize a sugar on a cell, in the same way that proteins can be tagged with a fluorescent marker and picked out under a microscope.

The technique she developed, now widely used, is known as bio-orthogonal chemistry. It relies on marking sugars with a small,

Work / Technology & tools

biologically unreactive chemical group that can slip undetected past the enzymes that attach glycans to proteins. Once this tagged sugar has been incorporated into a complex glycan and draped over a protein, a fluorescent dye can be snapped onto that chemical group in the cell, allowing the glycan to be visualized under a microscope.

“The key was that we needed to find two functional groups that would react with each other, but neither would react with anything else in the body,” says Bertozzi. This ‘bio-orthogonality’ is what counts: “They need to be chemically invisible in the biological world.” She and her colleagues have applied bio-orthogonal tools to identify glycoproteins that are unusually abundant in, or unique to, prostate-cancer tissues; used them to track where cells with different surface glycoproteins migrate in the zebrafish jaw during development; and more.

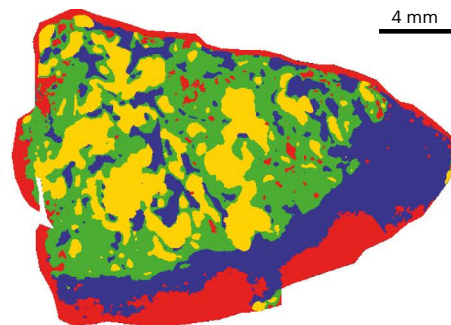
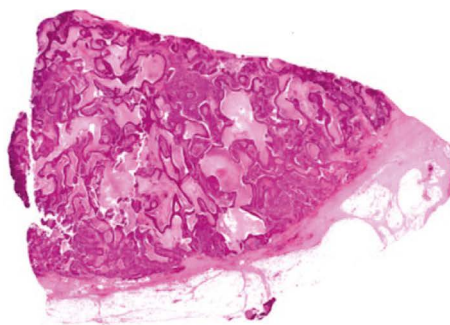
Now others are extending the concept. Instead of tagging a sugar with a chemically reactive group and then coupling it to a dye or fluorescently labelled antibody in separate steps, chemical glycobiochemist Peng Wu at Scripps Research in La Jolla, California, and his colleagues devised a way to tether the sugar to the dye directly, without the chemical group linker. That works because many of the enzymes that synthesize glycans will function even if their sugar substrate is toting a bulky fluorescent dye or labelled antibody. “The molecular weight of the sugar intermediate in reactions is 400–500 daltons,” Wu says. “No one thought it’d be possible to introduce an antibody with a molecular weight of 150 kilodaltons on to the sugar and have the reaction still work.”

In a study² last year, Wu’s team injected zebrafish embryos at the one-cell stage with two dye-labelled sugars, and tracked the tagged molecules through development using confocal microscopy. When compared with the two-step bio-orthogonal reaction, these labelled sugars yielded stronger signals from deep tissues such as the zebrafish head².

Abundant arrays

Such tools can reveal facets of glycan metabolism, but to crack the glycome, which encompasses all of a cell’s glycans, glycobiochemists require a different tool set. “High-throughput methods are essential for glycoscience to keep pace with discoveries in proteomics and other fields,” says Lara Mahal, a chemist at the University of Alberta in Edmonton, Canada.

In 2002, for instance, researchers adapted one of the original high-throughput tools of genomics, the DNA microarray, to glycoscience. The glycan array is a slide dotted with synthetic polymers that can help to identify proteins that bind to sugars, and researchers using it have identified, for example, differences between the cellular binding sites for human and avian



Various *N*-glycans can be located in human breast-cancer tissue using mass spectrometry (right) and compared with the same section of tissue coloured using typical cell stains (left).

RICHARD DRAKE

influenza viruses³. But glycan arrays present sugars at high densities and without their cellular protein and lipid partners. As a result, they might not reflect true biological interactions. So Mahal turned to nature’s original glycan binders, proteins called lectins. By putting these on an array, she created a tool that binds to all the glycans in a sample, whether they are isolated sugar fragments or are attached to proteins, lipids or other biomolecules⁴.

To reveal the diversity and abundance of glycans on proteins, researchers today are blending these approaches with a tool

“You can quickly see which proteins’ glycosylation patterns are altered in diseases.”

of metabolomics and proteomics research called MALDI mass spectrometry imaging. Mass spectrometry identifies molecules on the basis of their mass and ionic charge. Proteomics researcher Anand Mehta at the Medical University of South Carolina in Charleston and his colleagues have combined mass-spectrometry imaging with arrays of glycoprotein-binding antibodies to measure the relative amounts of glycans bound to different proteins present in samples such as human blood serum, which can contain hundreds of glycosylated proteins⁵. “You can quickly see which proteins’ glycosylation patterns are altered in cirrhosis, cancer or other diseases,” he says.

At the University of Copenhagen, glycoscientist Henrik Clausen and his team have designed a cell-based glycan array by pruning the sugars off a common cell line called HEK293, and then reintroducing the genes for 170 glycan-synthesis enzymes⁶. Subsets of cells express different enzymes – and thus, different surface glycoproteins – and serve essentially as spots on an array. But rather than imaging the results with a microarray reader, the researchers use flow cytometry, a method in which cells are scanned individually with a laser to identify bound molecules. Turning to the enzymes rather than the sugar structures alone places glycome research

in its biological context, Clausen says. “Not only do you learn what structures they bind to, you find out what genes and enzymes are involved in making that structure.”

Dissecting bonds

Clausen is also working to address another vexing aspect of glycobiochemistry. Despite significant advances in understanding sugars’ complex structures, Clausen says, “we are still quite far from being able to, in an unbiased analysis, understand which sugars are at what sites on what protein”.

Last year, the US National Institute of Standards and Technology in Gaithersburg, Maryland, provided 76 labs around the world with samples of a specific glycosylated antibody and asked them to identify the sugars present and their locations in the antibody protein. The teams reported three broad chemical groups of glycans containing sialic acid, fucose, galactose or their derivatives. But their detailed assessments varied widely⁷.

Using mass spectrometry, researchers effectively identified glycan linkages, but in many instances failed to differentiate between sugar isomers. Labs also struggled with a class of sugars called *O*-linked glycans, which are connected to an oxygen atom in an amino acid. There’s no specific amino acid or sequence that marks the location of an *O*-linked glycan, and although many analytical tools require glycans to be separated from their protein backbones, no single enzyme can cleave all such groups. *N*-linked glycans, by contrast, are attached to asparagine residues in a conserved sequence of four amino acids on proteins, and can all be sheared off the protein using an enzyme that leaves a characteristic molecular ‘scar’, Clausen says. “That’s why our understanding of the *N*-glycoproteome – not in terms of structures but where the sugars are – is decades ahead of all types of *O*-glycans.”

Last year, Clausen’s team developed a method to try to close that gap. First, the team used its cell-based glycan arrays to create a library of mass spectra from *O*-linked sugars representing more than 2,000 glycoproteins. Using this, the group was able to detect and quantify 269 *O*-linked glycans without the need for a

preset range of ions and spectra, which currently are available only for *N*-linked glycans⁸.

Other groups have adapted Raman spectroscopy, a method that uses molecules' vibrational spectra as signatures, to visualize glycans on cell surfaces. One study⁹ applied the method to living tissues, and identified glycosylation patterns that were unique to breast- and brain-cancer cells. "Most Raman studies so far have focused on simple model proteins, so to see it used on an actual biological sample is really interesting," says radiology researcher Sharon Pitteri at Stanford.

Raman spectroscopy is "a good match" for the relatively abundant sugars found in biological tissue, says Ewan Blanch, a physical chemist at RMIT University in Melbourne, Australia. But attempts to use it have been hampered by a lack of reference data. Technological advances are improving matters, Pitteri says. Historically, researchers had to cleave sugars from proteins and study glycans separately. Now, they can slice glycoproteins in different ways to study sugars in the context of protein fragments, then cleave the two apart to examine the sugar and protein individually. These tools are particularly helpful for *O*-glycans, she adds.

Mainstream merging

Researchers are also working to better integrate glycomics with wider biomedical research. Such connections can help to identify not just how glycans are altered in cancer, immune dysfunction or other diseases, but also why. "If you tell a cell biologist that his protein binds di-sialo-fucosyl-polyLacNAc, he knows nothing," Clausen says. "But if you tell him that the protein glycosylation requires these four genes to be expressed, he can go back to genetics and manipulate that glycosylation."

This step is also crucial for therapeutics, Mahal adds, because drug developers are "not likely to target the glycan, but the enzyme that makes it".

Indeed, large-scale screens frequently implicate glycan-processing enzymes in various processes and diseases, making it possible – and even necessary – for biologists to reckon with glycoscience. "When we take the bias out of biological inquiry, it often sends us back to glycoscience," Bertozzi says.

Jyoti Madhusoodanan is a science writer based in Portland, Oregon.

1. Srivastava, J., Sunthar, P. & Balaji, P. V. Preprint at bioRxiv <https://doi.org/10.1101/2020.01.15.908491> (2020).
2. Hong, S. et al. *Angew. Chem. Int. Edn* **58**, 14327–14333 (2019).
3. Xu, R. et al. *Science* **342**, 1230–1235 (2013).
4. Pilobello, K. T., Krishnamoorthy, L., Stawek D. & Mahal, L. K. *Chembiochem* **6**, 985–989 (2005).
5. Black, A. P. et al. *Anal. Chem.* **91**, 8429–8435 (2019).
6. Narimatsu, Y. et al. *Mol. Cell.* **75**, 394–407 (2019).
7. De Leoz, M. L. A. et al. *Mol. Cell. Prot.* **19**, 11–30 (2020).
8. Ye, Z., Mao, Y., Clausen, H. & Vakhrushev, S. Y. *Nature Meth.* **16**, 902–910 (2019).
9. Kopec, M., Imiela, A. & Abramczyk, H. *Sci. Rep.* **9**, 166 (2019).

Q&A



Melissa Simone Survey sleuth

Melissa Simone, a quantitative psychologist at the University of Minnesota in Minneapolis, uses surveys to study eating behaviours in people from sexual and gender minorities (LGBT+). In less than 24 hours of promoting one such survey on Twitter, she received 386 responses. But in most cases there was nobody at the keyboard. Simone's survey had been attacked by 'bots', automated online mischief-makers created by people who were probably targeting her survey for the US\$15 reward she offered.

To Simone, bots are "like fake Twitter accounts". The fact that they might be deployed to sabotage scientific studies, she says, is "mind-blowing". She spent 200–300 hours developing a battery of tests to ferret out the false responses, culling her data set to just 11.

Simone shared her findings on Twitter in September 2019. In November, she relaunched her survey, this time recruiting participants directly rather than on social media. Nature asked her about her experience.

How did you know your survey had been attacked?

My survey had a lot of open-ended questions. As I was scrolling through the answers, I noticed a response that used Latin words. I thought, "Huh, that's weird," and kept scrolling. I saw that exact response again and again. That convinced me there was something wrong.

How did you identify suspicious responses?

We used 'skip logic' to personalize the surveys. If the user clicks "yes, I am transgender", they should see questions about their experience as a trans person. But because bots are following the underlying code rather than the logic of the survey, they will click that they're cisgender (someone who identifies with the gender they were assigned at birth), and still answer questions about holding a transgender identity.

Bots were also able to skip questions that were required for all participants, and produce bundles of responses that were identical across all survey fields.

Some bots started and stopped the study at the exact same time, amounting to 488 questions answered in just 7 minutes. That's basically impossible to achieve with real respondents. It's pretty unlikely that this many people started the study at the exact same minute and finished it exactly 7 minutes later.

How can other researchers protect their own surveys?

First, never use a public survey link. Unique, personalized links can prevent people from using the same IP address to submit hundreds of responses.

Second, use 'honeypots' – questions that no person should be able to see, which resemble 'real' questions. For instance, I added a question about marital status directly after one regarding relationship status, and then hid the marital-status question from human participants. A bot wouldn't realize it should skip that question and would answer it instead.

Include open-ended questions, because that's a really good way to detect suspicious patterns. And continuously check your data throughout the lifetime of your survey. It used to be acceptable to check your data every couple of days to ensure its integrity. It's evident from my experience that that's not enough.

How has the survey relaunch gone?

We've had some bumps along the way, but things are moving along quite well. We are no longer advertising our study through Twitter or on public pages. Instead, we recruit participants exclusively from college and university campuses, through queer-specific programmes and groups. We did run into a bit of a problem when one of our partners shared our study advert on Twitter. Minutes after posting we had about 100 false responses from bots. After they removed the post, the fake responses trickled in for a day or two before things returned to normal.

Interview by Jeffrey M. Perkel

This interview has been edited for length and clarity. For further tips, check out Simone's Twitter thread: <https://go.nature.com/2wsybdk>.