

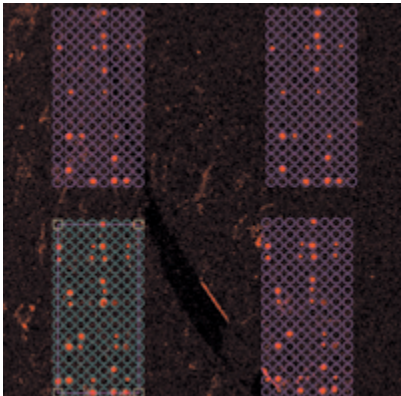
**Counting Carbs: Glycomics Gears Up**  
**Alan Dove, PhD, Contributing Editor**  
**Drug Discovery & Development - March 01, 2006**

*Molecular complications really increase after translation of DNA into proteins, when the proteins acquire branched carbohydrate structures. Multidisciplinary consortia are forming to begin deciphering this glycomic role.*

Two German telegraph operators were exchanging important bulletins one day during World War II on Enigma machines, portable devices used to encrypt and decrypt messages. The first operator sent "KYQ YSTM HOLRP VEVXQ CUHQK TPOJH LXSME JRWSM FCGHV CXYVD QFDAQ WH." The second operator replied "SXEJK BSR UQV GIB WFP TUQ." He was executed the next morning. Get it?

Like cryptographers eavesdropping on an encoded conversation, biologists studying communication between cells often get the feeling that they've been left out of a joke. Even with the genetic code cracked, numerous genomes sequenced, and sophisticated tools for protein chemistry available on demand, some of the most critical parts of cellular dialogue are still indecipherable.

The problem is all the more frustrating because it occurs in plain sight: the cells of complex eukaryotes display signaling proteins proudly on their surfaces, and the



Results from an array produced by the Consortium for Functional Glycomics. A library of carbohydrate structures is arrayed onto the surface, then probed with a test protein. This image shows the concanavalin A protein binding high-mannose carbohydrate structures on the array, run in quadruplicate to provide internal controls. (Source:

Consortium for Functional Glycomics)

interactions of these proteins often produce easily measured responses. Often, though, apparently identical protein interactions produce completely different results. In recent years, biochemists have discovered that post-translational modifications, especially glycosylation, can explain many of these cryptic conversations.

During translation, eukaryotic cell surface proteins enter the endoplasmic reticulum, where a small set of enzymes selects from a small set of carbohydrates to glycosylate the proteins. Unlike linear amino acids or nucleic acids, sugar polymers can branch, allowing these few components to generate mind-bogglingly diverse structures. Humans, for example, have only nine common monosaccharides available for this process, but the Enigma machine of glycosylation enzymes could assemble them into any of 15 million different tetrasaccharides, each of which would be considered a relatively simple glycan.

While genomic and proteomic projects press forward, a scattered collection of glycan researchers has begun to coalesce into large-scale consortia to launch the field of glycomics. A close look at one of these efforts, the Consortium for Functional Glycomics (CFG), reveals some of the challenges and promise of this nascent field. CFG is an initiative funded by the National Institute of General Medical Sciences to understand the role of carbohydrate-protein interactions at the cell surface in cell-cell communication.

**Sweet signatures**  
 Studying sugar modifications is the biochemical equivalent of an extreme sport, and much of the gear is still being developed. "There is a definite need for multiple groups to get together in these large consortia," says Stuart Haslam, PhD, director of the CFG analytical core facility based in the life sciences department at Imperial College, London.

Like most forms of chemical analysis, mass spectrometry produces a result that is cryptic to the uninitiated. Converting the spindly tracings of the mass spectrometer into a chemical structure for a protein's attached carbohydrates usually requires several days' work by an expert, precluding high-throughput use of the technique. In a key breakthrough, CFG researchers developed a sophisticated computer algorithm that can provide expert-level analysis of glycoprotein mass spectrometry in minutes rather than days, automatically drawing chemical structures for the carbohydrates.

The scientists can then work backward to determine which glycosylation enzymes were involved in modifying each protein. This reverse genetics approach is essential in glycomics. While genome sequencing has revealed 98 putative glycosylation enzymes in humans, the epigenetic algorithms that decide which enzyme acts on which substrates, and in which cell types, remain almost

#### Sugar-Coated Proteins

Though its potential remains largely untapped, even skeptics concede that glycomics could revolutionize drug development. Most pharmaceutical companies have at least a passing interest in glycans, and a few firms are now focusing intently on the field (see table 2, page 14). Not only are glycoprotein modifications a central theme in mammalian physiology and pathology, they also underlie some of the most recalcitrant problems in biotechnology.

For example, a single glycosylation enzyme in pigs causes their organs to be rejected immediately when transplanted into humans, stymieing xenotransplantation. On a larger scale,

completely opaque.

Several thousand miles west of Haslan's laboratory, another CFG core provides the other end of the reverse genetic experiment. "We breed knockout mice that are missing key genes in the synthesis of glycans, or in the glycan binding proteins that recognize them, and make those available to investigators," says James Paulson, PhD, director of the CFG and a professor of molecular biology at the Scripps Research Institute, La Jolla, Calif. Besides the transgenic mouse core, Scripps also hosts a core facility for gene expression microarrays focused on glycobiology and a "glycan library" of tools for analyzing glycan specificity and biology.

Because the CFG mission is to amass data on protein-carbohydrate interactions, it has caught the fancy of many immunologists. "We've been looking at a whole series of leukocytes—T cells, B cells, and natural killer cells—because these are the cells that possibly go through lots of [glycoprotein] interactions . . . with pathogens and also in their normal functional interactions with other cells in the host," says Haslan.

Paulson agrees that so far, the primary interest in the project has come from immunologists. As the consortium's database grows, though, its users will likely become more diverse. "Investigators are joining from different disciplines that really had very little contact with the glycobiology community," says Paulson.

All mammalian cells glycosylate at least some of their proteins, a process that requires substantial energy, so some evolutionary pressure must favor carbohydrate addition. Nonetheless, the functions of many sugar modifications are still unknown. "We might start providing information which might lead to researchers in all sorts of fields thinking about . . . how these carbohydrates could be governing interactions that their pet cell types might be involved in," says Haslan.

#### Non-Atkins Arrays

Increasing the throughput of mass spectrometry is an important step forward for glycomics, but Paulson stresses that "we're not trying to . . . determine the mammalian glycome, if you will, but we are setting up infrastructure that's relevant to that."

Consequently, the CFG's first five years of operation have been dominated by tool-building activities. First funded in 2000, the effort is now halfway through its planned 10-year life, and has just been applied for a midterm renewal. "The first couple of years, we were really just developing the tools that we now make available to all investigators . . . including those who are not members of the consortium," says Paulson.

Rapid advances in the consortium's analytical capabilities and the group's policy of making its services available at no charge have fueled an exponential rise in requests from researchers. Last year, the CFG received about 200 such requests, and Paulson expects that number to continue increasing. Besides the carbohydrate analysis core facility and a burgeoning population of transgenic mouse lines, the CFG also offers array-based screening tools for glycan binding.



A technician examines genetically modified yeast that can attach human-like glycans to their proteins, providing a streamlined system for therapeutic protein manufacturing.

Richard Alvarez, assistant professor in the department of biochemistry at the University of Oklahoma Health Sciences Center, Oklahoma City, Okla., directs the CFG's protein-carbohydrate interaction core where the arrays are being developed. To jump-start the glycomics array technology four years ago, Alvarez says he and his colleagues "looked at off-the-shelf technology that we could use immediately . . . and we also looked at libraries of oligosaccharides or carbohydrate structures that were currently available at that time."

In the team's first-generation system, carbohydrates conjugated to biotin stuck to the wells of streptavidin-coated microtiter plates. Adding a test sample to the wells lit up a fluorescent readout when a protein in the sample bound one of the carbohydrates. The microtiter plate system was quick to build and faster to use than any other glycan-screening approach available at the time, but it was large and cumbersome compared to slide-based proteomic or genomic arrays.

most therapeutic proteins and all therapeutic antibodies must be produced in cultured mammalian cells. Other systems, especially yeast, offer vastly greater protein yields and faster production times. Some therapeutic proteins, like insulin, are already produced in yeast, but many of the biggest biotechnology blockbusters require human-like glycosylation, which yeast can't provide.

"Antibodies are very sensitive to different glycosylation structures. As you change the sugars . . . those changes actually make a difference in terms of how the antibody communicates with the immune system," says Tillman Gerngross, PhD, chief scientific officer at GlycoFi Inc., Lebanon, N.H. Therapeutic antibodies with the wrong carbohydrates are substantially less potent and shorter-lived than properly glycosylated versions.

To address this, GlycoFi researchers generated a series of genetically modified yeast strains, each with a slightly different alteration in its glycosylation pathway. In a recent experiment to prove the concept, the company successfully used yeast to make a test batch of properly glycosylated Rituxan (rituximab), a monoclonal antibody sold by Genentech for the treatment of non-Hodgkin's lymphoma.

With antibody production facilities already operating at full capacity and pharmaceutical companies looking to shave mushrooming production costs, yeast-based systems have gained an eager following. Now, yeast advocates just need to convince drug approval agencies that the system yields safe products. "It is fair to say that the regulatory burden is going to be somewhat higher [than for mammalian cells], but I don't think there is an inherent issue relating to the expression system," says Gerngross.

(Source: GlycoFi Inc.)

Having tested the general concept on microtiter plates, and with the CFG's glycan synthesis team ramping up the number of new synthetic carbohydrates available, the Oklahoma team moved to a slide-based system. "This was not a novel thing; other people have published on printed [glycan] arrays on various surfaces, but we were in the nice position of having resources that allowed us to really do that on a scale that other people perhaps aren't able to do," says Alvarez. Using this second-generation technique, the investigators estimate they can easily accommodate the 1,000 new glycans the CFG expects to produce in the next five years.

Researchers interested in screening a target for glycan binding can use the new system at no charge. A scientific steering committee reviews the requests, which are generally approved if they meet basic standards and are not experiments the consortium has already done. In general, Alvarez and his colleagues take the sample and do the analysis themselves before sending results back to the researcher. But if biological containment or technical issues dictate, the group may send prepared arrays and protocols out to other labs. The only catch to the free service is that the data, like all CFG results, must be made public in the consortium's database.

Like any biochemistry experiment, though, glycomics array analysis often fails. "I would say our success rate hovers around a 30% range over time, so we have a lot of people who send us stuff where the first time we run it on the slide, we don't necessarily see a result," says Alvarez. With a universe of millions of glycans, a slide that screens fewer than a thousand might simply lack a ligand for a known glycan-binding protein. Also, purified proteins might not be folded correctly, or the affinity of the protein-glycan interaction might be too low for the array-based assay.

Echoing a familiar theme, however, Alvarez says "when we get a result, it's fantastic, it's really a 'gee whiz' kind of thing." Recent slam dunks have included new ligands for receptors involved in innate immunity and important new observations about the glycan-binding specificities of different strains of influenza virus.

### Storing sugars

For most large genomics and proteomics projects, processing the flood of data from massively parallel experiments is a challenge of its own, and bioinformatics teams struggle to design databases with sufficient capacity. In glycomics, the problem is different. "There will be a lot of data, but not a flood of data," says Paulson, but he adds that the much greater complexity of glycomics data more than offsets the smaller quantity.

"When you go to the glycome, the complication that you have is this sort of non-template-driven synthesis," says Rahul Raman, PhD, director of the bioinformatics core of the CFG and a postdoctoral associate in the division of bioengineering and environmental health at the Massachusetts Institute of Technology, Cambridge, Mass. "To get a particular glycan structure, for example, it . . . involves an expression of several genes."

Because a glycan structure does not mimic the sequences of the proteins that produced it, glycomics researchers face a code-breaking challenge. It is like knowing that the first German telegraph operator said "there's a new law that anyone caught telling puns will be hanged." Because the code uses a combinatorial algorithm rather than a defined template, the punch line remains elusive. However, placing this tidbit into context can provide clues.

For Raman and his colleagues, building the contextual relationships between glycan structures, protein-glycan interactions, and the body of relevant biochemistry literature is the main focus. "How do you store this information?" asks Raman. "You have glycans coming in different flavors, not just linear sugars, but branched sugars and so forth."

Besides the branching, there is the challenge of identifying possible errors in the synthesis process. DNA repair and RNA editing are easy to spot, because they leave clear differences in linear nucleic acid and protein sequences, but it is difficult to tell whether a particular glycan was synthesized correctly without understanding the underlying rules. Compounding the problem, the CFG data sometimes include partial structures, so the bioinformatics team must keep the database flexible to handle future updates.

"One thing that the consortium is really doing is sort of truly integrating the resources, so you can look at gene expression of the glycoenzymes . . . and you can try to correlate it with the actual profile of glycans that have been analyzed," says Raman. The portal for this integrated data set is the CFG Web site, which allows anyone to drill down from a cell type to a particular protein-glycan interaction. The interface is intuitive, and the system links to the relevant sections of the biochemistry literature at several levels. "We're trying to set it up in such a way that the data that come in will be searchable and therefore readily available to investigators that are interested in specific pieces of information," says Paulson.

While the CFG works on its own database, other glycomics consortia around the world (see table 1, above) are working on theirs, raising the risk of a Babel of incompatible data standards. "I think the first step really is to see if we can homogenize each of our repositories of structures using some standard data interchange format," says Raman, who adds that the major glycomics groups are already working on such a standard. Because the data are accumulating relatively slowly compared to the deluge of genome and gene

NAME	WEB SITE	LOCATION	INTEREST
Consortium for Structural Glycomics	http://www.structglycomics.org	La Jolla, Calif.	Identifying protein-carbohydrate interactions
GlycoProbe Consortium	www.glycoprobe.org	London, UK	Developing new carbohydrate arrays
Human Glycan Glycomics	www.hugglycomics.org	London, UK	Identifying glycan-carbohydrate interactions
IGCC	www.igcc.org	Washington, D.C.	Establishing standards for glycomics databases

NAME	WEB SITE	INTEREST
University of Toronto, Canada	www.glycomics.com	Carbohydrate-mediated vaccine for host-microbial systems
IGCC, St. Louis, MO	www.igcc.org	Genetically modified yeast to study human glycoenzymes in glycomics
IGCC, St. Louis, MO	www.igcc.org	High-throughput glycomics data and databases
Novus Technologies Inc., Madison, Wis.	www.novus.com	High-throughput glycomics data and databases

click the image to enlarge

Table 1: Major Multi-Institution Consortia Studying Glycomics

expression data, bioinformatics experts are hopeful that the standards will be set before the databases get too cumbersome.

Although some pioneering companies are already working to commercialize new findings from glycomics, the field is still in its infancy (see "Sugar-Coated Proteins" sidebar). It will likely be years before researchers have deciphered the basic rules governing the glycome. If one of those encrypted rules makes wordplay a capital crime, however, a response such as "I'll spread the noose" is probably ill-advised.

**About the Author**

*Originally trained as a microbiologist, Alan Dove has been writing about science and its interfaces with industry and government for more than a decade.*

*This article was published in G & P magazine: Vol. 6, No. 2, March, 2006, pp. 10-14.*

© 2007 Advantage Business Media

<http://www.dddmag.com/glycomics-aided-by-multidisciplinary-approach.aspx>