Sugars tied to the spot

Sabine L. Flitsch and Rein V. Ulijn

The interactions of sugars and proteins underlie many biological processes, and cataloguing them is a daunting task. A technique for attaching sugars to microarrays offers a promising, high-throughput solution.

Of the three main classes of biopolymers — proteins, nucleic acids and sugars — the sugars, or saccharides, are the most complex and hence the most difficult to study. Each monosaccharide building-block has multiple attachment sites, which means that sugars can be built in a variety of linear or branched fashions. For example, two common sugars, glucose and mannose, can be linked to form a disaccharide in up to 80 different ways. It takes just a few coupling steps to produce a large number of diverse biological structures, and this diversity of saccharide structure is exploited *in vivo*. A major challenge in cell biology is to decipher the sugar code, or 'glycome' — that is, to define the interactions between cell-coating sugars and proteins and work out how they recognize each other. Microarray techniques offer the possibility of high-throughput analysis and have been investigated in a series of reports[1–3], the latest by Fabio Fazio and colleagues in the *Journal of the American Chemical Society*[4].

One of the best-known examples of medically important sugar–protein interactions is the action of the human blood-group antigens A, B and O. The difference between these three antigens lies in a single carbohydrate unit in an oligosaccharide sequence on the surface of red blood cells. When different blood types are mixed, any 'foreign' A, B or O antigen is recognized by highly specific sugar-binding proteins (agglutinins) that cause blood agglutination, or clotting. Hence, transfusions fail if the blood type is not matched to that of the patient. Other types of sugar-recognizing proteins include those activated to adhere to sugars in immune responses, lectins that crosslink cell-surface sugars, antibodies that are specific to certain sugars, and enzymes such as glycosidases and glycosyltransferases.

To investigate the great diversity of oligosaccharide structure requires an automated technique, but it should also be sensitive to small quantities of material. Microarrays have been used successfully in the study of gene expression and protein interactions, and are particularly well suited to studying the glycome. By attaching complex sugars to a microarray plate, the compounds can be quickly and easily screened to work out their specific interactions with proteins. But the problem is how to link the sugar under study to the microarray surface.

A method devised by Fukui *et al.*[1] relies on the fact that the reducing end of all oligosaccharides contains a unique functional group (an aldehyde) that reacts with an aminolipid to form a so-called neoglycolipid by 'covalent conjugation'. The neoglycolipids can then be immobilized efficiently by spotting them into arrays on nitrocellulose membranes. In an ambitious experiment, Fukui *et al.* arrayed 30 different fractions of neoglycolipids generated from rabbit-brain polysaccharides (*O*-linked glycans) on nitrocellulose. Using highly specific antibodies to probe for certain carbohydrate sequences, they were able to obtain structural information on the oligosaccharides present.

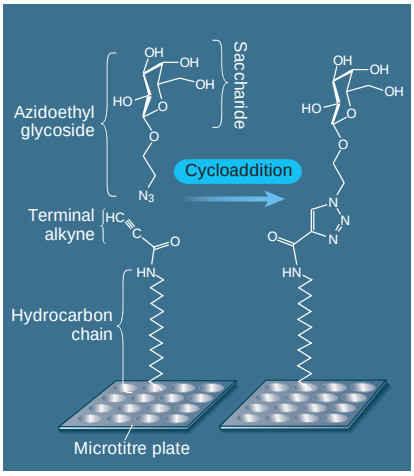

Figure 1 **Making a microarray of sugars. Fazio *et al.*[4] exploit a simple cycloaddition reaction to attach complex sugars to a microtitre plate. The alkyne group at the end of a hydrocarbon chain (attached to the plate) is a natural tether for azidoethyl glycoside, whose terminating N_3 group binds to the alkyne, forming a ring.**

Wang and colleagues[2] have shown that carbohydrates can also be spotted directly on single nitrocellulose-coated glass slides without the need for covalent conjugation. In this system, up to 20,000 spots of microbial antigens could be printed on a slide — which would be enough to include most of the known human microbial antigens. The sensitivity is such that antibodies can be detected if the microarray is tested with only one microlitre of human blood serum.

But there are drawbacks to each of these techniques: in the neoglycolipid approach, the first monosaccharide unit is lost on binding to the nitrocellulose membrane, and the Wang *et al.* method gives little control over the orientation of the immobilized polysaccharide. An exciting new development is the use of cycloaddition reactions — a class of simple, selective chemical reactions in which a ring structure is formed — by Houseman and Mrksich[3], and now by Fazio *et al.*[4]. Using a specific cycloaddition, the Diels–Alder reaction, Houseman and Mrksich showed that a series of ten different

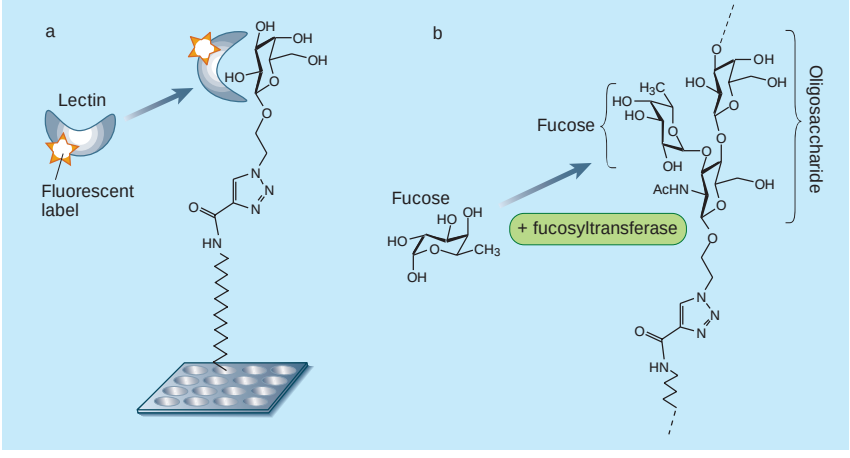

Figure 2 **Putting the sugar-microarray technique[4] to the test. Fazio *et al.* provide several examples to demonstrate the potential of their approach for high-throughput analysis of protein–sugar interactions. These include**: a, **the binding of lectin molecules to sugars on the microtitre plate traced through fluorescein labelling; and** b, **addition of a fucose group, mediated by the enzyme fucosyltransferase (and monitored using a fucose-specific lectin, not shown).**

monosaccharides could be specifically detected by five different lectins. Fazio *et al.* exploit cycloaddition between azides (which contain an N_3 group) and alkynes.

Fazio *et al.* made microarrays of oligosaccharides by immobilizing the molecules on polystyrene microtitre plates. Starting with azidoethyl glycosides, they exploited the cycloaddition reaction with the terminal alkyne group of a saturated hydrocarbon chain to bind the sugars to the microtitre plate (Fig. 1). In a series of experiments, 11 different di-, tri- and tetrasaccharides were successfully attached to the plates. Fazio *et al.* then tested their behaviour as ligands for lectins (Fig. 2a), tracking the binding of these proteins by labelling them with the fluorescent marker fluorescein ($C_{20}H_{12}O_5$).

The authors also demonstrated that the oligosaccharides attached to the microarray were still accessible as substrates for enzymatic transformations such as fucosylation, in which a fucose group is attached to the oligosaccharide through the action of fucosyltransferase (Fig. 2b). Fucosylation of a particular oligosaccharide produces the tetrasaccharide 'sialyl Lewis x', which is known to be involved in inflammation: this tetrasaccharide is produced on leukocytes (white blood cells) as a response to inflammatory stimuli and then mediates, through protein–sugar interactions, the recruitment of leukocytes from the bloodstream to inflammatory lesions. Fazio *et al.* tracked the process of fucosylation on their microarray using a fucose-specific lectin, and confirmed the results by mass spectrometry.

This work shows the potential for sugar-microarray technology, which in the future should permit systematic and high-throughput analyses of protein–sugar interactions and enzyme-inhibition screening, similar to the analyses being developed for nucleic acids and proteins. At the moment, there is still the problem of chemically generating diverse libraries of saccharides to work on, but efforts to automate oligosaccharide synthesis are under way. With microarrays, it should ultimately be possible to compare sugar expression on different proteins, cells or organs, and even to monitor how the sugar patterns on cell surfaces change as the cells develop and take on their particular function in the living system. ■

Sabine L. Flitsch and Rein V. Ulijn are in the School of Chemistry, University of Edinburgh, Joseph Black Building, West Mains Road, Edinburgh EH9 3JJ, UK.
e-mails: sabine.flitsch@ed.ac.uk
rein.ulijn@ed.ac.uk

1. Fukui, S., Feizi, T., Galustian, C., Lawson, A. M. & Chai, W. *Nature Biotechnol.* **20,** 1011–1017 (2002).
2. Wang, D., Liu, S., Trummer, B. J., Deng, C. & Weng, A. *Nature Biotechnol.* **20,** 275–281 (2002).
3. Houseman, B. T. & Mrksich, M. *Chem. Biol.* **9,** 443–454 (2002).
4. Fazio, F., Bryan, M. C., Blixt, O., Paulson, J. C. & Wong, C.-H. *J. Am. Chem. Soc.* **124,** 14397–14402 (2002).

Functional genomics

RNA sets the standard

Thomas Tuschl

One way of finding out what genes do is to inactivate them, and to study the effects, in 'model' organisms. That has now been done for many thousands of worm genes in two large-scale analyses.

Genome-sequencing projects provide a tremendous amount of information about the genetic make-up of an organism. One of the most obvious types of data to be gleaned from these ventures is a list of genes: for instance, the nematode worm *Caenorhabditis elegans* — the first animal to have its genome sequenced — has a predicted 19,757 genes that code for proteins. A list won't tell you what these proteins do, however, and one way of starting to find out is to inactivate the genes, one at a time, and see what happens. As they describe on page 231 of this issue, Kamath and colleagues[1] have done just that, using double-stranded RNAs to rapidly and transiently inactivate 16,757 of the worm's predicted protein-coding genes. Meanwhile, Ashrafi and co-workers[2] (page 268) have analysed these genes specifically to see if they have a role in regulating body fat. Together, their work sets a new standard for systematic, genome-wide genetic studies.

The analysis of gene function in animals and plants was revolutionized in 1998 by the discovery[3] of the mechanism underlying 'RNA-mediated interference' (RNAi) in *C. elegans*. In its natural form, this is a cellular defence process that protects against genome-invading transposable genetic elements and viruses. RNAi is now widely used by researchers to temporarily block gene expression; its effects on the morphology, physiology and behaviour — the phenotype — of an organism are similar to those of classical gene knockouts (deletions of segments of a gene's protein-coding regions)[4].

In nature, the trigger for this cellular defence mechanism is double-stranded RNA (dsRNA), which is normally produced from uncontrollably replicating invasive genetic elements but never from tightly regulated cellular genes. The nucleotide sequence of the dsRNA guides the defence machinery towards single-stranded messenger RNAs of the same sequence as the dsRNA, and directs their destruction. The generation of an mRNA is the first step in producing a protein from a gene, so, in the case of invasive genetic elements, destroying mRNA abolishes the synthesis of crucial proteins, such as those needed for replication. Researchers have taken advantage of this basic mechanism: introducing a dsRNA identical in sequence to a cellular gene of choice will cause the destruction of the cellular mRNA, thus disrupting protein synthesis.

Caenorhabditis elegans is remarkable in that such 'gene silencing' can result when the dsRNA has simply been eaten[5]. The bacterium *Escherichia coli* is the preferred diet of *C. elegans*, and molecular-genetic tools allow appropriate dsRNA-encoding DNA sequences (plasmids) to be introduced into *E. coli* (Fig. 1). The effects of the loss of function of a gene are best observed in the offspring (either embryos or larvae) of hermaphrodite worms that have feasted on bacteria modified in this way. In the past, this and similar approaches have been used to target about one-third of all *C. elegans* genes[6–8]. Kamath *et al.*[1] have now almost tripled the amount of available information.

The authors show that inactivation of about 10% of the genes they studied produced a defective phenotype: the offspring might die as embryos or larvae; or the larvae might grow slowly, be sterile, have morphological defects, or move in an uncoordinated way. One interesting finding was that targeting *C. elegans* genes for which there are conserved counterparts in other eukaryotes (such as yeast, fruitflies and the plant *Arabidopsis thaliana*) was more likely to produce an aberrant phenotype than targeting less conserved genes. That might have been expected, as genes that are conserved during evolution are likely to be essential. Also, a peculiarity of the *C. elegans* genome is that it contains many recently duplicated genes. Kamath *et al.* show that gene duplicates are at least partially functionally redundant, or have such specialized functions that defective phenotypes are not detectable. Simultaneous inactivation of such sequence-related genes may be necessary to uncover their functions.

The authors also analysed the protein structural domains encoded by the genes they studied. Genes that tended to have non-lethal effects when inactivated by RNAi generally encoded domains that were 'animal-specific', being found in either humans or fruitflies as well as *C. elegans*. By contrast, genes that caused death when inactivated generally encoded domains that are of more ancient origin, being found in yeast and *Arabidopsis* as well. So, the domains that are essential for survival have been preserved across evolutionarily diverse species; the domains that are not needed for survival appear to have evolved during