Hand in glove

BENJAMIN G DAVIS

With their language of sophisticated handshakes and secret signs, sugars play a vital role in cellular communication. Cracking their code offers huge therapeutic potential

Image Bank

Molecular freemasonry: sugars convey vast amounts of information

At its most basic level, genetics implies that every living organism can be described using a code of just four letters. Each letter represents a base within a DNA molecule — adenine (A), thymine (T), guanine (G) and cytosine (C). The order these bases appear in the DNA determines the molecule's function. Each of these letters is coded for by a molecule of ribonucleic acid and they, in turn, act as a blueprint for the production of proteins, providing the structure and function for all organisms.

The idea that the rich variety of life can be reduced to just four letters once seemed overwhelming, but it is now accepted. Indeed, scientists can manipulate this code, and study it in an effort to understand the evolutionary cryptography that Nature has been playing with it for billions of years. The various genome projects seek to map all of these codes as an invaluable guide to the corresponding protein sequences and a source of molecular knowledge of ourselves. But what if this isn't the final code? What if there is an extra, as yet virtually untapped layer of complexity? There are good reasons for thinking that this may be the case.

Simple parallels show that codes are just a form of communication — a mechanism for transmitting information. Electronically, we choose the simplest: zero or one, off or on. This simple system derives its power from sheer bulk — many zeros and ones. DNA takes this a step further to a set of four. But a vocabulary of four words does not lend itself well to subtle conversation. 'IS IT NOT GOOD?' 'GOOD IS NOT IT.' 'IT IS NOT GOOD.'

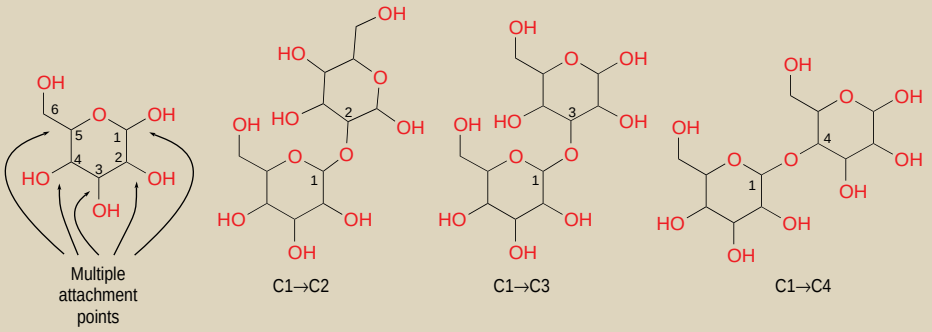
Of the three classes of abundant biomolecules — nucleic acids, peptides and carbohydrates — science has understandably focused on the simplest of these: the nucleic acids. Perhaps in doing so it is overlooking a higher level of subtlety and precision in the way our bodies function. We are comfortable with linear communication, for example, the lines of letters on this page. The simplicity allows us to comprehend the meaning. But why should communication be limited to one dimension? What if we added a second or a third? Nature is not flat.

In contrast with nucleic acids and peptides, which typically link to each other through only two attachment points, carbohydrates have many potential attachment points. Instead of just a

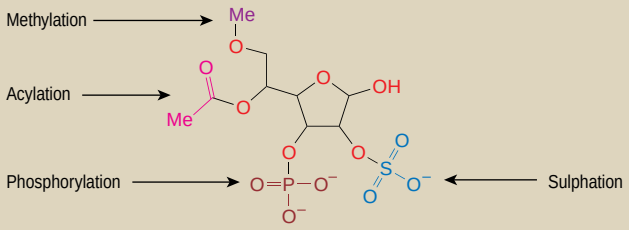
Carbohydrate connectivity

1

Sugars offer a wide variety of linkages, as can be seen in the joining of two pyranoses



Further modifications such as different ring size, methylation, acylation, phosphorylation and sulphation can expand the code



left or a right, a back or a front, they offer multiple connection points — they branch. This means that carbohydrates are unrivalled in the density of information that they can convey. Precise differences in the nature of the linkages between two carbohydrate residues — for example, C1–C2, C1–C3, C1–C4, C1–C6 for two pyranoses — contrast with the linear nature of proteins and nucleic acids (see Figure 1). Comparing the permutations of hexamer formation illustrates this point. Whereas DNA (with a basis set of four) and amino acids (with a basis set of 20) can construct a biological language for information transfer of 4096 and 6.4×10^7 'words', respectively, carbohydrates have access to more than 1.05×10^{12} variations.[1] Add to this the additional variety afforded by anomeric stereochemistry, ring size and subunit modification — for example, sulphation, phosphorylation, methylation or acylation — and it is clear that the language of carbohydrates has exquisite eloquence. This language has been christened 'glycocode' — a term that well represents the potential level of complex information that carbohydrates can convey.

Reassessing sugar

We tend to think of sugars only as sources of energy or structure, but in reality they are much more. The presence of carbohydrate units in naturally occurring structures has a dramatic effect on their physical, chemical and biological properties.[2] Carbohydrates project from cell surfaces in abundance with what was thought for a long while to be seemingly little function. Scientists now realise that all along these sugary hands have been passing on subtle messages — freemasonry on a molecular scale.

Over the past 25 years, example after example has demonstrated that when Nature wants to say more than 'IT IS GOOD', sugars are the language of choice. They give the interactions between biomolecules a precision that prevents mistakes by allowing only those molecules with the correct sugar structure — or glycocode — to cause an effect. Their remarkable structural diversity means that carbohydrates can mediate highly specific and therefore complex processes.

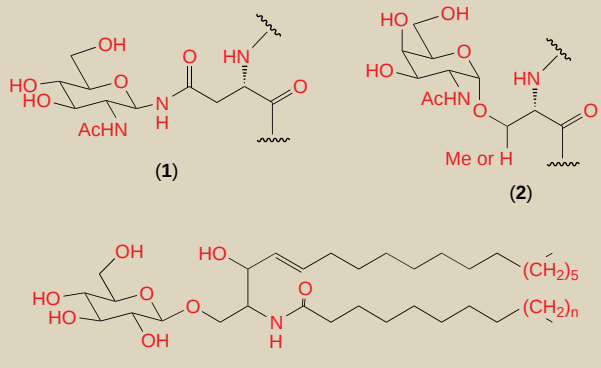
The sugars involved are usually presented as conjugates. Mainly they appear as glycoproteins of two types: *N*-linked (**1**), where the sugar is bound to the side chain of asparagine residues, and *O*-linked (**2**), where the sugar is bound to the side-chain of serine or threonine residues (see Figure 2). These proteins are found either at the surfaces of cell membranes or in soluble form. The other major class of glycoconjugate is glycolipids, such as (**3**). These contain a lipophilic 'head' that buries itself in the lipid bilayer of cell membranes, attached to a carbohydrate portion that protrudes from the membrane.

Decoding the message

Glycocodes are usually decoded by sugar-binding proteins called lectins.[3] Despite their very shallow binding sites, lectins are remarkably specific in their binding of multivalent complex carbohydrate structures.[4] Binding takes place in part of the lectin known as the carbohydrate recognition domain. Typically, this involves hydrogen bonding from backbone and side-chain amide group donors to oxygen lone pair acceptors, and from carbohydrate hydroxyl group donors to backbone and side-chain carbonyls. Protein-bound calcium ions can also play a role both in coordinating carbohydrate hydroxyl groups and in 'shoring up' the lectin's backbone, helping to position correctly potential hydrogen-bond donors and acceptors. Moreover, these polar interactions are amplified by non-polar van der Waals interactions between aromatic side-chains and hydrophobic 'patches' on the carbohydrates. The resulting cradle of non-covalent interactions places a large degree of spatial constraint on which ligands can bind and is part of the source of lectin specificity — if the glove fits, it's worn.

Glycocode basic building blocks

2



When more than one sugar of the right type and in the right orientation are clustered together, there is a rapid increase in both affinity and specificity by the corresponding lectin. This 'cluster' or 'multivalent effect' is only partly explained by an increase in local carbohydrate concentration, and a high degree of cooperativity is implied. As a result, other 'wrong' sugars present will not inhibit any process that the lectin concerned mediates. In addition, the specificity of this type of binding is

very finely-tuned. It relies not only on the complementarity between the individual lectin binding sites and a particular sugar ligand, but also on the relative arrangement of the binding sites to each other in space and therefore the corresponding display of each sugar ligand relative to the next.

In Nature, multivalent arrays are built on the surface of cells either by branched oligosaccharides of glycoproteins, which act like sugary hands with each 'finger tip' grasping a lectin 'bowling ball', or by the sliding together of glycolipids within the lipid bilayers to form carbohydrate-dense, 'sticky' patches.[5] Diverse carbohydrate bonding gives way to diverse oligosaccharide patterns (first-order patterns of sugars in clusters). This gives way to diverse cluster patterns (for example, second-order arrangements of different clusters on the same glycoprotein), which in turn gives way to diverse glycoconjugate patterns (such as third-order arrangements of glycoproteins on a cell surface). In essence, glycocode is a biological fractal with each layer of structural diversity generating another layer of still greater diversity.

There is also a corresponding number of multireceptor display methods, ranging from the presence of many single-binding proteins on one cell surface, such as the asialoglycoprotein receptor in the liver which forms hexamers, through to single proteins that have more than one binding site — for example, mannose-binding snowdrop lectin which has three per monomer.

The effects of glycocode can be viewed on three levels, though of course these only represent points on a spectrum of activities and the interplay between these areas is a crucial part of the subtle balancing act that sugars play.

Hands that defend

This first group of effects can be thought of as 'semi-structural' as they depend more on the physical properties of the sugars than on their precise 3-D structure. For example, the presence of sugars on amino acid side-chains protects proteins from thermal degradation or hydrolysis by peptidases, enzymes that cleave proteins. These effects can be attributed either to the polarity of these residues or to their role as steric shields in protecting potential areas of attack in the protein. In this context, these sugary hands are protective, fending off enzymes by virtue of their size or changing polarities of local regions using polar hydroxyl 'finger tips'.

Another fascinating example of this type of more physical effect is that of antifreeze proteins such as (**4**) (see Figure 3). Found in the blood of deep-sea fish, these proteins disrupt ice crystallisation, allowing the fish to survive at temperatures as low as –4°C.

Hands that shake

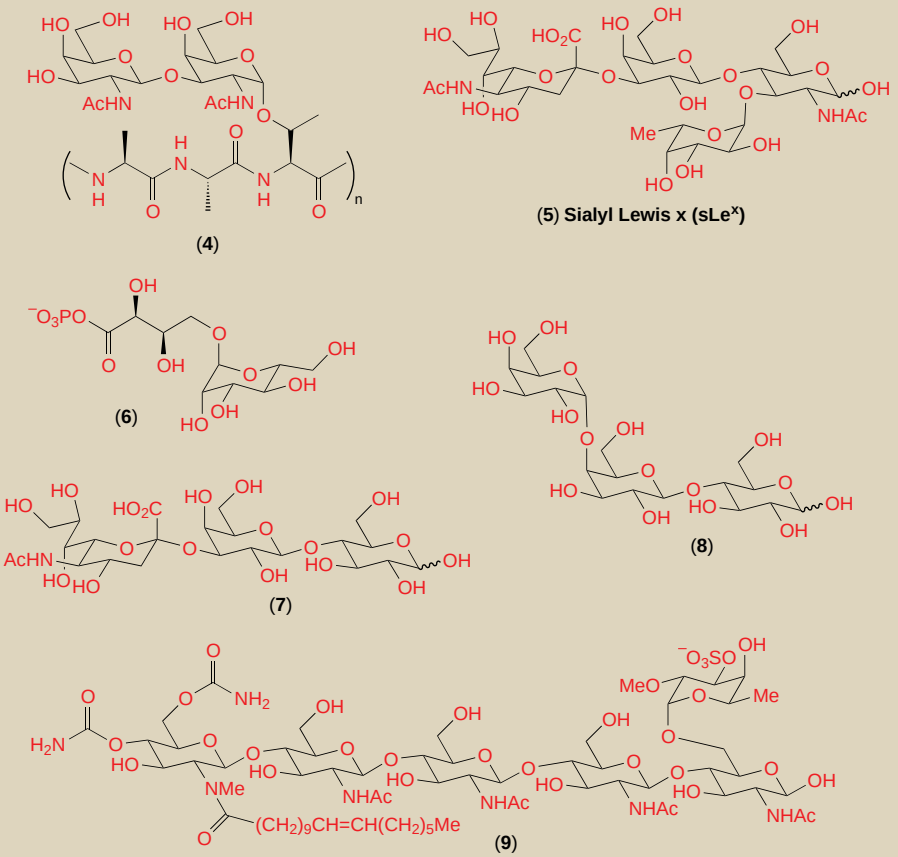
Correct spatial arrangement is crucial to the next level of sugar role: binding and docking. The past 25 years have seen a revolution in glycobiology as scientists have realised the key role sugars play in highly specific 'handshaking' reactions. One important example is that of the inflammatory response.[6,7]

Damage to tissue as a result of injury or infection sets in motion a complex process designed to recruit white blood cells — the body's foot soldiers — to the scene. To deal with the problem, the blood cells need to know where it is and then have to be able to get out of the blood stream into the damaged tissue. However, this process — by virtue of making the walls of the blood vessels in the location 'leaky' — allows for a build up of fluid in the tissue, resulting in inflammation.

Damage to tissues surrounding a blood vessel causes an influx of signalling molecules to the injury. These encourage selectins — a particular class of sugar-binding lectins — to be expressed on the inner surface of blood vessels (E-selectin) and on platelets in the blood stream (P-selectin). Handshaking interactions between the selectins and the sugars on the cells' surfaces encourages white blood cells to roll along the blood vessels to the damage site. Here the cells stick to the blood vessel walls via protein–protein interaction, completing this two-way binding process by expressing L-selectins. Ultimately, the white blood cells are able to escape through the vessel wall into the surrounding tissue. While this is a well-controlled process in healthy individuals, if too many blood cells escape the blood vessel, they can cause damage to healthy tissue, resulting in a range of problems including septic shock, arthritis, asthma and heart disease.

The precise identity of the ligands on the carbohydrates that bind to the selectins is not known, but the tetrasaccharide sialyl Lewis x (sLe^x, (**5**)) is bound by all types of selectin and serves as a useful benchmark (see Figure 3).[7] Several copies of sLe^x are present in glycoproteins, such as GlyCAM-1 and PSGL-1, that are present in high concentrations on the surface of certain white blood cells. This opens the door to a host of therapeutic applications using sugars to inhibit the initial selectin–blood cell interactions and so regulate the early stage of inflammation.

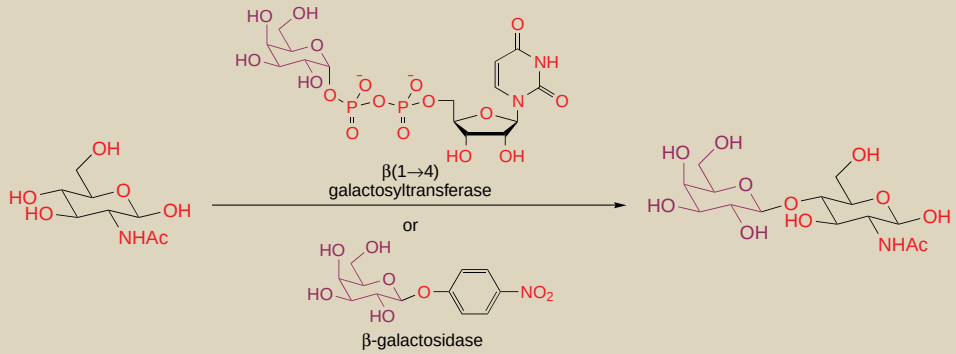
Examples of sugar 'hands' and 'fingers' **3**



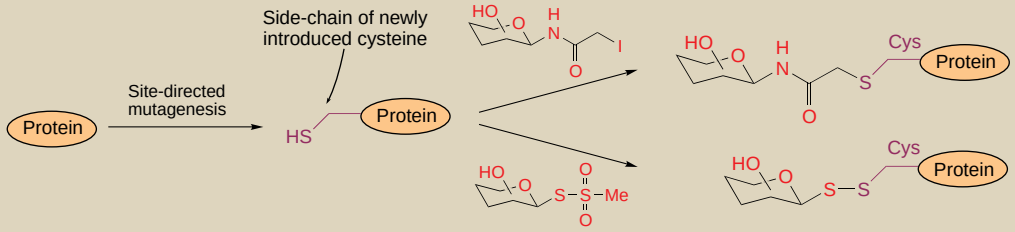
Creating glycocode

4

(a) Exploiting Nature's machinery — enzymes — allows complex sugars to be formed quickly and easily



(b) A combination of protein mutagenesis and chemical modification provides high levels of control in code construction



For example, the sLe^x mimic (**6**) is an 800-times more potent selectin binder than sLe^x itself.[6] Furthermore, the high specificity of this type of docking means that these compounds will inhibit only this interaction, providing a treatment that is potentially devoid of side effects. Another example of this kind of docking interaction involves *Helicobacter pylori*, the bacteria that cause gastric ulcers. These microorganisms attach themselves to gut cells during the infection process by binding to extracellular glycoproteins containing the sugar sialic acid. As part of an anti-ulcer strategy, this adhesion can be inhibited by conjugates glycosylated with 3′-sialyllactosyl residues (**7**), such as the phase II drug candidate NE-0080.[8]

Several toxins also act as a result of sugar handshaking. Notorious examples are ricin and abrin from castor beans and the rosary pea, respectively. These related proteins can act as both galactose-binding lectins and base-specific RNA-*N*-glycosidase enzymes. As enzymes, they hydrolyse a specific purine base from the sugar backbone of the large RNA subunit of the ribosome (the site of protein synthesis in a cell). This single modification inactivates the ribosome, stopping the synthesis of cellular proteins and ultimately causing cell death. This has earned this family of proteins the apt acronym RIPs — ribosome inactivating proteins.

The toxin enters the cell by binding to sugars on its surface. This is followed by a process of internalisation and release into the cell's cytoplasm where ribosomes are then inactivated. Just one internalised RIP molecule is needed to kill a cell. The lectin subunits of RIPs account for their extreme cytotoxicity (three orders of magnitude more potent than cyanide) — proteins which only have RNA-*N*-glycosidase activity are, in general, much less toxic than those carrying a second lectin subunit.

A similar mechanism of action is shown by the toxin released by *Escherichia coli* O157:H7 — which causes 'hamburger disease' — and by *Shigella dysenteriae*. This toxin gains access to kidney cells by binding only to the trisaccharide (**8**) on the surface of kidney red-blood cells. Treatments being developed exploit this specificity. For example, attaching many copies of (**8**) to a polymer provides an ingestable treatment — such as *Synsorb Pk*, which is now in phase III clinical trials — that 'grabs hold' of the toxin in a selective manner, thereby clearing it from the body.[9]

Cracking the glycocode should enable treatments to be devised that are virtually devoid of side effects

Molecular masons

The final and, as yet therapeutically untapped, type of role takes sugar docking one step further by eliciting a response as a result of their interactions. Unlike handshaking, these interactions do more than facilitate; they cause the transmission of a signal, often across cellular membranes. Two examples are Nod factors and the Mannan-binding lectin (MBL).

Nod(ulation) factors, such as hexasaccharide (**9**), are released by *Rhizobium* bacteria, which are responsible for nitrogen fixation.[10] These Nod factors bind to lectins on the root hairs of legumes and initiate a cascade of signalling and deformation within root hair cells. As a result, nodules develop on the roots where these symbiotic bacteria reside. Nod factor structures are highly specific to both the corresponding plant and bacterial species, so that only a particular species of bacteria will nodulate a given legume species. Sugars with different stereochemistry or substituents cause little or no such effect, again illustrating the importance of the glycocode in sending messages.

The complement system, which is an essential part of the innate immune response, can be started by the binding of the serum protein mannan-binding lectin (MBL) to the sugar mannose, which is found on many bacterial cell surfaces.[11] Upon binding, MBL activates the complement system via a long cascade of protein-mediated signals starting with an associated peptidase enzyme called MASP-1. This protein cuts, and in doing so activates, another enzyme called MASP-2, which in turn cuts and activates a further peptidase enzyme (C4) and so on. Further cycles of cutting and activation lead first to the 'tagging' of the bacteria that MBL first bound to and then the selective killing of this tagged intruder by a collection of hydrolysing enzymes. Thus, the precise docking of a bacterial mannose glycocode sparks off a rapid and remarkable signal that leads to a powerful antibacterial activity.

Investigating glycocode

Access to well-defined biomolecules to probe the nature of sugar-based interactions is essential. The elucidation of the mechanisms of many of the examples discussed above and their consequences is a major goal in glycoscience and has driven,

and continues to drive, the synthesis of novel glycoconjugates. These are the tools of the glycobiology trade.

Unlike the biosynthesis of proteins and nucleic acids, there appears to be no associated mechanism for proof-reading and correcting differently glycosylated biomolecules — the result is mixtures. Therefore, glycoproteins occur naturally in a number of forms (glycoforms)[12] that have the same peptide backbone, but differ in both the nature and site of glycosylation. The different properties shown by each component within these mixtures presents problems in determining their exact function through structure–activity relationships. It has even been suggested that these naturally-occurring mixtures of glycoforms provide a spectrum of activities that can be biased in one direction or another as a means of rapid fine-tuning.[13] Consequently, the very few studies that have compared single glycoforms successfully[13] have required abundant sources and extensive separation. The urgent need for sources of pure glycocodes is being satisfied by chemistry.

The potential of this research is enormous — more and more sugars will emerge as codes that may be used to control disease with an unrivalled precision

Sugar chemistry has often been perceived as tricky or difficult, but a host of exciting and powerful techniques give the lie to this myth (see Figure 4). One approach adapts the natural machinery for processing carbohydrates — glycosidase and glycosyltransferase enzymes.[14] These environmentally benign catalysts allow the efficient formation of complex sugars. For example, the conventional synthesis of sLe^x (**5**) takes 31 steps, but this has been superseded by a one-pot system that uses three glycosyltransferases, one for each of the sugar–sugar links in the tetrasaccharide.[15]

New, highly selective methods for attaching these sugars to their protein or lipid carriers are also being developed.[16] For example, by combining protein mutagenesis with chemical modification, both the site of glycosylation and the sugar code to be placed there can be controlled. Such glycosylation techniques will allow scientists to probe glycocode with absolute precision — in effect testing out each letter of the code.

Therapeutics and the future

By cracking the glycocode, scientists can use it to send, or more usually to block, messages in ways that will be interpreted by the body with a high degree of fidelity. The fact that these interactions are so remarkably specific, indicates that the drugs that are being developed as false codes will be just as specific as the genuine codes. This should enable treatments to be devised that are virtually devoid of side effects. Moreover, this precision will give a much greater chance of success in trials, leading to drastically shorter lag times to market and lower development costs.

The potential of this research is enormous — more and more sugars will emerge as codes that may be used to control disease and biological function safely and with an unrivalled precision. Carbohydrate–protein interactions in sperm–egg recognition could be exploited by using glycoconjugates in contraception or fertility treatments, and the cellular recognition processes identified in cancer are examples of newly emerging glycocodes that may have profound effects once deciphered.

It is a sobering thought that as a result of the way carbohydrates are introduced to cell surfaces, the cataloguing of the human genome will tell us little of the sugar code. For example, the biosynthesis of glycoproteins falls into two halves: translation — the reading of the letters of the nucleic acid code to produce the protein backbone — and modification — the attachment to or alteration of groups on the protein after backbone construction. While the former can be confidently predicted through genomics, the latter is only known through hindsight. Only by examining the protein after modification and relating its properties to its structure can we begin to understand some of the relationships and principles. The challenges of this discipline — glycomics — are enormous but the rewards are a far more direct knowledge of biological function.

References

1 Laine, R.A., *Glycobiol.*, 1994, **4**, 759-67
2 Varki, A., *Glycobiol.*, 1993, **3**, 97-130
3 Boyd, W.C., *Science*, 1954, **119**, 419
4 Sharon, N. & Lis, H., *Essays Biochem.*, 1995, **30**, 59-75
5 Roy, R., *Curr. Opin. Struct. Biol.*, 1996, **6**, 692-702
6 Simanek, E.E., McGarvey, G.J., Jablonowski, J.A. & Wong, C.H., *Chem. Rev.*, 1998, **98**, 833-62
7 Lasky, L.A., *Annu. Rev. Biochem.*, 1995, **64**, 113-39
8 Zopf, D. & Roth, S., *Lancet*, 1996, **347**, 1017-21
9 Takeda, T., Yoshino, K., Adachi, E., Sato, Y. & Yamagata, K., *Microbiol. Immunol.*, 1999, **43**, 331-7
10 Price, N.P.J., *Carbohydr. Res.*, 1999, **317**, 1-9
11 Taylor, P., Botto, M. & Walport, M., *Curr. Biol.*, 1998, **8**, R259-61
12 Rademacher, T.W., Parekh, R.B. & Dwek, R.A., *Annu. Rev. Biochem.*, 1988, **57**, 785-838
13 Rudd, P.M., *et al*, *Biochemistry*, 1994, **33**, 17-22
14 Crout, D.H.G. & Vic, G., *Curr. Opin. Chem. Biol.*, 1998, **2**, 98-111
15 Ichikawa, Y., *et al*, *J. Am. Chem. Soc.*, 1992, **114**, 9283-98
16 Davis, B.G., *J. Chem. Soc., Perkin Trans 1*, 1999, 3215-37

Dr Davis is a lecturer in organic and biological chemistry at the Department of Chemistry, University of Durham, South Road, Durham DH1 3LE, UK. Email: Ben.Davis@durham.ac.uk