Glycans in evolution and development

Workshop on Glycoscience and Development

Catherine L.R. Merry[1+] & Sviatlana A. Astrautsova[2]

[1]Materials Science Centre, University of Manchester, Manchester, UK, and [2]Department of Microbiology, State Medical University of Grodno, Grodno, Belarus

This EMBO Workshop on Glycoscience and Development took place between 9 and 12 December, 2007, in Lille, France, and was organized by P. Delannoy, Y. Guérardel, T. Merry and J.-C. Michalski.

Introduction

It has been recognized for some time that congenital disorders of glycosylation (CDGs) can lead to numerous and diverse problems during development, particularly neural development (Freeze, 2007). This illustrates the importance of glycosylation in the process and has encouraged research that has identified many developmentally regulated proteins involved in *N*-glycosylation. Changes in the glycans attached to protein or lipid cores have also been identified in evolutionary processes—with possible relevance to the formation of the modern human brain (Angata *et al*,

2006; Schwartz & Domowicz, 2004)—and recent technological advances have made the study of glycan attachment possible for a wide range of scientists.

The EMBO Workshop on Glycoscience and Development was therefore a timely meeting, which brought together experts in the specialist field of glycobiology with developmental and evolutionary biologists—who were keen to understand better this complex topic (Varki, 2006)—and encouraged the participation of younger scientists. The workshop was a three-day event at the historic Couvent de Minimes in the old town of Lille, France.

The workshop was organized around three main themes. The first focused on recent technological advances in glycoscience, particularly those relating to developmental and evolutionary biology. The second discussed how glycosylation might affect development and focused on developmental processes in which the role of sugar modification is better understood, as well as studies of CDGs in which development is affected. The third shifted the focus to studies investigating the importance of glycosylation in evolutionary biology, including comparative studies of glycosylation in different species and the possible implications of these changes.

An encounter with bioethics

The meeting opened with a lively debate that was sponsored by the BioTethed project (www.biotethics.org) to promote the discussion of biotechnology ethics. C. Ward (Manchester, UK) highlighted the crucial need for embryonic stem-cell research (Eastham *et al*, 2007), particularly in complex areas of biology such as glycoscience. We were then fortunate to hear L. Wolpert (London, UK) talk about the ethical issues surrounding these types of study, and the need for increased public understanding of their importance and conduct (Wolpert, 2007).

State of the art and perspectives

R. Dwek (Oxford, UK) highlighted the importance of glycobiology research for drug discovery, with an introduction to the biological role of glycosylation and how glycan analogues might prove to be invaluable as new antiviral drugs. Such drugs would have the unique ability of avoiding common problems that are associated with the evolution of viral resistance. The compounds used for these applications, which are known as imino sugars (Butters *et al*, 2005), inhibit several

[1]Materials Science Centre, University of Manchester, Grosvenor Street, Manchester M1 7HS, UK
[2]Department of Microbiology, State Medical University of Grodno, Vilenskaja Street 19, 230023 Grodno, Belarus
[+]Corresponding author. Tel: +44 (0)161 306 8871; Fax: +44 (0)161 306 3586; E-mail: catherine.merry@manchester.ac.uk

Abundance of O-GlcNAc on nuclear and cytoplasmic proteins
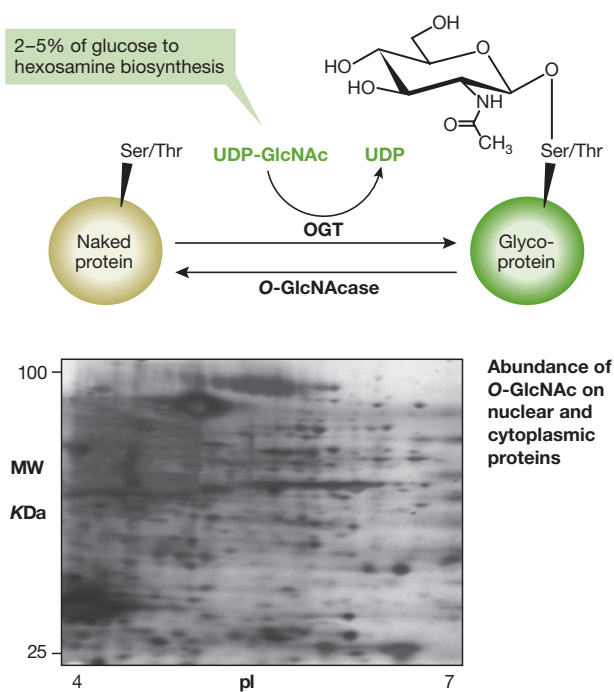
Properties of O-GlcNAc

1. It has been found in over 500 proteins and is often found to be reciprocal or competitive to phosphorylation events and cycles dynamically on a similiar timescale to these

2. Unlike most glycosylation processes the GlcNAc is not extended by addition of further monosaccharides

3. In contrast to most other types of protein glycosylation it is only localized to cytoplasmic and nuclear proteins

4. It is present in all metazoans studied including plants

5. There are specific enzymes for the addition of this residue to (O-GlcNAc transferase EC 2.4.1) and removal from (O-GlcNAcase EC 3.2.1.52) peptide chains

Fig 1 | Properties of O-linked N-acetylglucosamine (O-GlcNAc) as described by G. Hart (Baltimore, MD, USA). O-GlcNAc is a dynamic post-translational modification of nuclear and cytoplasmic proteins. Cycling of O-GlcNAc is regulated by the enzymes O-GlcNAc transferase (OGT) and O-GlcNAcase. Between 2% and 5% of glucose is metabolized to O-GlcNAc from UDP-GlcNAc, which is the donor for GlcNAcylation. GlcNAcylation is sensitive to nutrients and to cellular stress, and seems to be important in signalling, transcription and human diseases (including diabetes and neurodegenerative disorders). The inset represents a western blot of a two-dimensional gel of nuclear and cytoplasmic proteins from the human HeLa cell line, probed with an antibody against O-GlcNAc, which illustrates the large number of proteins that are modified by O-GlcNAc.

enzymes that are involved in both the biosynthesis and the catabolism of glycosylation. Using the example of deoxynojirimycin (DNJ), Dwek discussed the treatment of viral diseases such as hepatitis C. The release of intact viral particles requires the correct glycosylation of viral coat proteins. Incorrect glycosylation, which can be engineered by the application of imino-sugar derivatives, leads to protein misfolding that delays or even prevents the release of intact virus particles. Dwek described how the modification of DNJ, particularly through the extension of the alkyl chain, leads to the formation of inhibitors that effectively block viral replication at doses that, importantly, are sufficiently low to be non-toxic to the host.

Dwek also showed how the modification of these sugars allows the selective inhibition of cellular processes. He discussed a severe inherited metabolic disease involving glycolipid storage, which is known as Gaucher disease, and how the use of DNJ in a procedure known as substrate-reduction therapy allows the disease to be successfully treated by limiting the biosynthesis of a crucial intermediate.

A. Varki (La Jolla, CA, USA) introduced sialic acids, which are typically derivatives of neuraminic acid. N-glycolylneuraminic acid (Neu5Gc) is widely expressed in mammals, but not in humans. Humans lost the ability to convert N-acetylneuraminic acid (Neu5Ac) to Neu5Gc ~3 million years ago and instead express an excess of Neu5Ac (Varki, 2007). Sialic acids are often found at the termini of glycan chains where they are well placed to mediate cell–cell or cell–ligand interactions. Varki highlighted how evolutionary changes are also underway in the proteins that act as receptors for sialic acids. Among these are alterations in the sialic-acid-recognizing immunoglobulin family lectins (siglecs), which are cell-surface receptors that mediate signalling responses. Within this large and complex family, the CD33-related siglecs are undergoing particularly rapid evolution. Varki suggested that they might have a central role in the distinction of self and non-self antigens, and that sialylated pathogens, such as group B *Streptococcus*, seem to have exploited this. The bacterial outer membrane contains capsular polysaccharides that interact with the siglecs that coat the leukocyte cell surface, and this interaction denotes 'self'. A high level of CD33-related siglecs is found on the lymphocytes of great apes, but not on human lymphocytes, and it was proposed that this could explain the observed propensity of humans to suffer from disorders that are mediated by hyperreactive T-cells.

A type of glycosylation involving a single monosaccharide, N-acetylglucosamine (commonly called O-GlcNAc; Fig 1), linked directly to the peptide chain of a protein, has been shown to be as common—and probably as important—as phosphorylation in signalling processes (Zachara & Hart, 2006). G. Hart (Baltimore, MD, USA) reported that O-GlcNAc is abundant, widely distributed on all functional classes of proteins (although found mostly on transcriptional regulatory proteins) and highly dynamic, and that there is extensive cross-talk between O-GlcNAc and phosphorylation. Sometimes both modifications can occur at the same serine/threonine residues on a protein, making the choice between O-GlcNAc and phosphorylation both crucial and competitive. O-GlcNAc can act as a nutrient sensor, regulating both signalling in general and transcription in particular. However, unlike kinases that fulfil a similar function, the enzymes of O-GlcNAc cycling are regulated by the binding of accessory proteins in a manner that is more analogous to the regulation of phosphatases. For example, the ability to transfer O-GlcNAc to some transcriptional proteins is controlled by the binding of O-GlcNAc transferase (OGT) to the transcriptional repressor Sin3A or to an RNA polymerase II-targeting protein (Hart et al, 2007). O-GlcNAc is also known to have an important role in glucose toxicity through a mechanism in which hyper-GlcNAcylation of transcription factors causes them to dysregulate gene expression.

Moving from the analysis and detection of single sugars to whole organisms, G. Lippens (Lille, France) showed how magic-angle spinning (MAS) nuclear-magnetic resonance (NMR) spectroscopy can

be used to see molecules in crude mixtures and even in living cells. This could be extremely useful in studies of metabolomics, which are becoming increasingly popular.

Glycosylation defects

J. Hewitt (Nottingham, UK) introduced us to the role of a particular type of glycosylation that has been shown to be important in muscle and, more generally, in the development of the neural system. A key component of muscle, which is known as α-dystroglycan, has a unique type of glycosylation. Some genetic defects, including some forms of muscular dystrophy, have been shown to result from defective glycosyltransferases. The correct glycosylation of α-dystroglycan (Fig 2) is now known to be essential for the protein to function normally and there is substantial evidence that *O*-mannosyl glycans are key components of this glycosylation. However, the functions of many of the proteins in this pathway (such as fukutin, fukutin-related peptide and LARGE) are still unknown. Mutations in any of these genes give overlapping phenotypes, and it is likely that the extent of α-dystroglycan glycosylation correlates with the phenotype.

LARGE, in particular, seems to be crucial because, when over-expressed, it is able to bypass a deficiency in any other enzyme in the pathway. There is considerable interest in the function of this protein, as it has potential therapeutic applications. Many groups have tried to identify the activity of LARGE without success, partly because of the small amounts of fully glycosylated α-dystroglycan that can be obtained. Hewitt described how the *LARGE* gene is highly conserved, with homologues in most sequenced animal genomes—*Drosophila* is a notable exception. The predicted amino-acid sequence of the sponge protein shows approximately 40% identity to human LARGE. Although the phenotype of a mouse-null mutant of *LARGE* is mainly myopathy, during embryogenesis the mouse gene is predominantly expressed in developing neuronal tissues with little expression in developing muscle. Postnatally, *LARGE* is expressed in skeletal muscle, heart muscle and the central nervous system (CNS). Within the CNS, LARGE mRNA is seen in neurons of the cortex, the cornu ammonis (CA), and the dentate gyrus regions of the hippocampus and the olfactory bulb. Within the cerebellum, *LARGE* expression is confined to the Purkinje cells. This tight regulation of expression indicates that LARGE might have site-specific functions within the mature CNS.

T. Butters (Oxford, UK) described in more detail how sugar analogues have complex and discrete activities in inhibiting metabolic processes, which should allow varied applications as therapeutics. *N*-butyl-DNJ (NB-DNJ) has two cellular enzyme targets: endoplasmic reticulum α-glucosidases I and II in the glycoprotein-processing pathway, and ceramide glucosyltransferase in the glycosphingolipid-biosynthesis pathway. The latter activity has been selected for the development of an approved therapy for Gaucher disease. The activity of NB-DNJ in the endoplasmic reticulum has been revealed through the use of a new free-oligosaccharide assay that detects the products of protein misfolding. *In vivo* experiments have shown a transrenal pathway for clearance of these oligosaccharide biomarkers.

Glycosylation and evolution

This session began with a detailed account by R. Oriol (Paris, France) of an important class of glycosyltransferases, the members of which are known as fucosyltransferases (FUTs; Coullin *et al*, 2003). Using methods akin to tracing one's family tree to establish the relationships between the complex families of enzymes that are involved in glycan
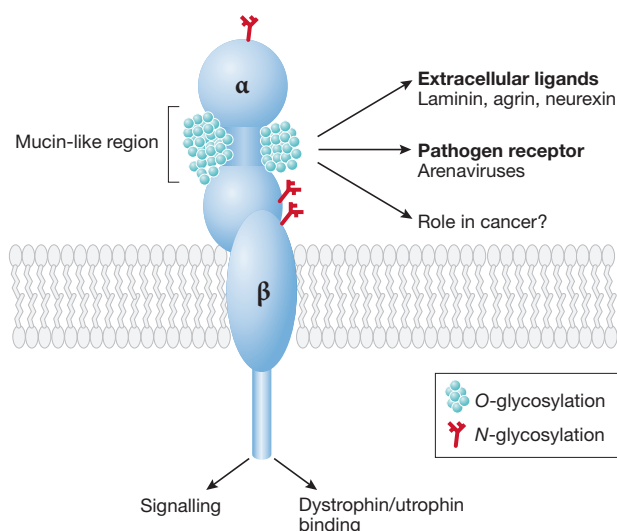
Fig 2 | Structure of α-dystroglycan as described by J. Hewitt (Nottingham, UK). Diagrammatic structure of the α-dystroglycan molecule showing its different regions and functional activities. The glycans under discussion are linked to the peptide in the mucin-like SEA (sea urchin sperm protein, enterokinase and agrin) extracellular domain of the molecule. *O*-GlcNAc, *N*-acetylglucosamine; OGT, *O*-GlcNAc transferase.

biosynthesis, Oriol described the two main types of glycosyltransferase in eukaryotes: enzymes that use a nucleotide-monosaccharide donor and enzymes that use dolichol-*P*-monosaccharide as the donor substrate. Each family derives from a common ancestor that underwent successive duplications followed by divergent evolution. An example of this occurrence in enzymes that use a nucleotide-monosaccharide donor is shown in Fig 3. Two main families of FUTs have been identified: one that contains α2-FUTs, α6-FUTs and protein-*O*-FUTs (Fig 3); and one that contains α3-FUTs and α4-FUTs. Each of these families has specific conserved peptide motifs that are signatures for the corresponding transferase activity; however, they share a common general type II transmembrane structure with a short amino-terminus cytoplasmic tail followed by a single transmembrane domain (TMD). Their active sites reside in the Golgi lumen with the exceptions of FUT10, FUT11 and protein-*O*-FUT (POFUT), in which the active sites are in the endoplasmic reticulum. In contrast to these structurally and functionally related families, the dolichol-related glycosyltransferases are hydrophobic multispan transmembrane proteins of the endoplasmic reticulum membrane, with ~12 TMDs distributed along the protein. Here again, there are two main families: enzymes that use dolichol-*P*-mannose (Dol-*P*-Man) and enzymes that use dolichol-*P*-glucose (Dol-*P*-Glc) as donor substrates. The analysis of ClustalW alignments shows that, within the enzymes using Dol-*P*-Glc, there are two related families, the α2-glucosyltransferases and the α3-glucosyltransferases, which derive from a common ancestor. Among the Dol-*P*-Man-related enzymes, there are three main families: the α2-mannosyltransferases and the PIG enzymes, SMP3 and PIGB; the α6-mannosyltransferases; and the α3/4-mannosyltransferases. In addition, the protein-*O*-mannosyltransferases belong to this multispan superfamily, using Dol-*P*-Man as a donor substrate, and they all probably derive from a
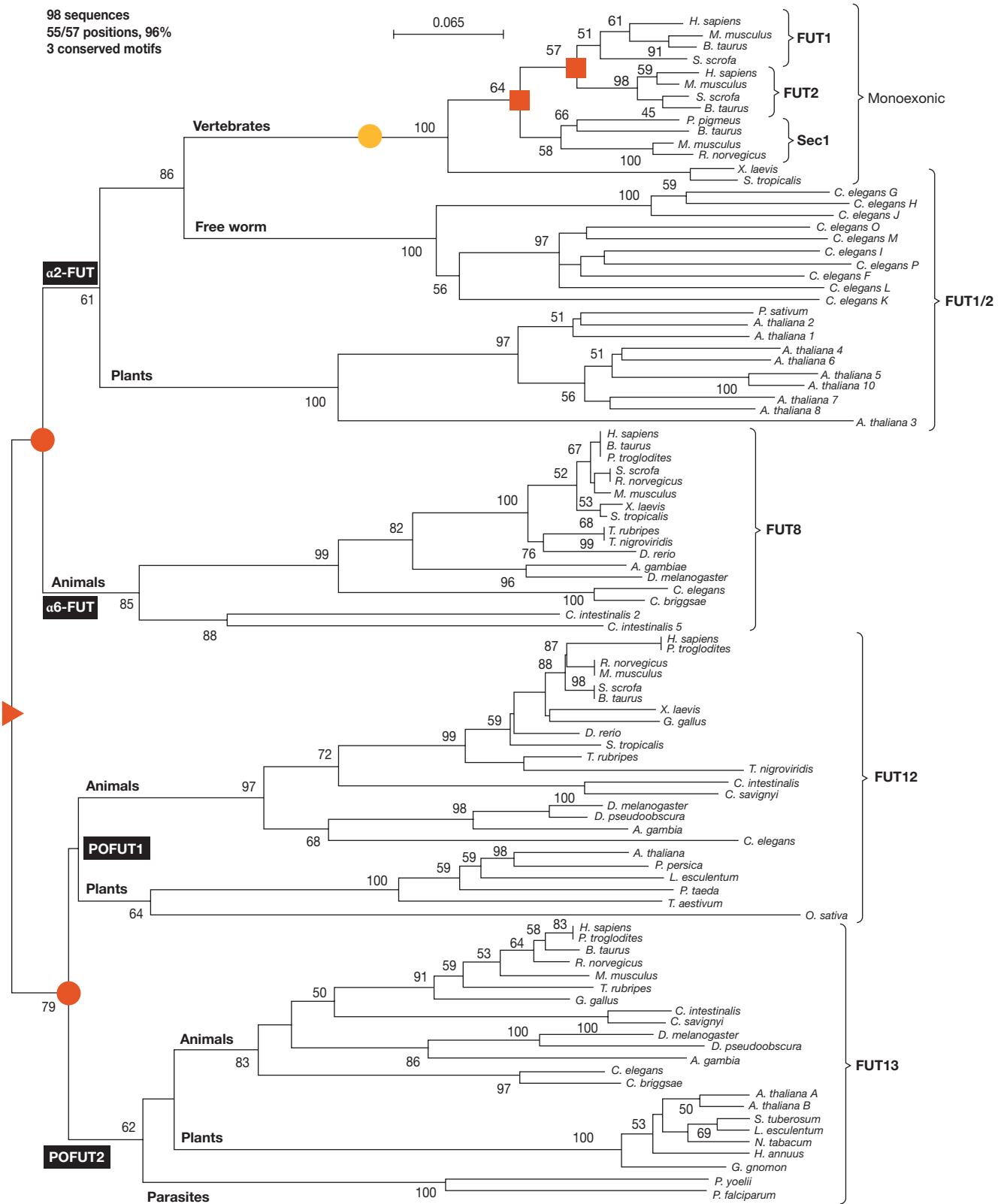
◄ **Fig 3 |** Evolution of fucosyltransferases as described by R. Oriol (Paris, France). Phylogenetic tree of α2-FUTs, α6-FUTs and protein-*O*-FUTs (POFUTs) illustrating the common origin (red triangle, at the root of the tree), the two duplications at the origin of the α2/6-FUT and the POFUT1/2 branches (red circles), and the two duplications at the origin of the three families of vertebrate α2-FUTs (red squares). The yellow circle shows the position of a retro-transposition event, which occurred in the vertebrate branch of α2-FUT and induced the loss of all the introns in the coding sequences of these genes, which are therefore monoexonic. FUT, fucosyltransferase; Sec1, Secretor locus-encoded α(1,2)fucosyltransferase 1.

single common ancestor of the Dol-*P*-Man-related α-mannosyltransferase superfamily, by successive duplications followed by divergent evolution (Fig 3).

This evolutionary theme was continued by C. West (Oklahoma City, OK, USA) for *O*-glycosylation, in which the sugars are attached to the peptide in a different way to the *N*-linked modifications described above. West showed that the soil amoeba *Dictyostelium* is an excellent model system in which to study how environmental signals regulate development (West *et al*, 2007). Starvation induces the amoebae to aggregate into a slug, which migrates to the soil surface. Here it 'culminates', forming a fruiting body of spores and stalk cells. Culmination is O_2-dependent and it has been shown that the enzyme prolyl-4-hydroxylase-1 (P4H1) acts as an O_2 sensor during this stage of *Dictyostelium* development (West *et al*, 2007). West described how culmination normally requires O_2 levels of more than 10% and that disruption of the *P4H1* gene increases this requirement so that culmination is blocked at ambient O_2 levels. By contrast, overexpression of P4H1 reduces the O_2 requirement of culmination to less than 5%. Because P4H1 is an orthologue of the prolyl hydroxylases that sense O_2 levels in animals, West suggested that it functions as part of an ancient mechanism for O_2 sensing that pre-dates the evolution of animals.

The session on development continued with a presentation by G. Vasta (Baltimore, MD, USA), who described experiments in the zebrafish, which has a series of glycan-binding proteins known as galectins that are also found in humans. By comparing their structure–function relationships, it is possible to trace how the functions of galectins, which include important modulations of the immune system, have developed. I. Wilson (Vienna, Austria) gave a talk with the provocative title 'How is a worm more complex than a fly?' and showed that there are many glycosylation systems in *Caenorhabditis elegans* that can be compared with those in *Drosophila melanogaster*. Wilson proposed that together, these two organisms provide a more complete system for the study of glycosylation.

Glycosylation and development/differentiation

Y. Van Kooyk (Amsterdam, The Netherlands) showed how glycosylation forms a 'code' by which cells can be recognized. She illustrated this with reference to the way in which the body recognizes self and foreign antigens, for example on pathogenic cells, and how our immune system relies on these glycosylation codes. The precise role of *O*-glycosylation mentioned previously was examined in more detail by K. Ten Hagen (Bethesda, MD, USA), who described studies in *Drosophila* showing that the many varied and subtle changes in glycosylation all have some role in development. Epithelial tubes are essential for the proper function of a diverse array of eukaryotic organs, and are present within the kidney, lung and vascular system. The formation of these tubes is complex, relying on many genes to control tube diameter and length, and the cell–cell junctions within the tube. Although many of these genes have now been identified, others remain unknown. Ten Hagen introduced a new class of gene involved in mucin-type protein *O*-glycosylation, the members of which are required for maintaining epithelial cell shape, polarity and paracellular barrier function in the *Drosophila* embryonic tracheal system. Mutations in one member of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase family (*Pgant35A*) are recessive-lethal and result in tracheal tubes that are irregular in diameter and morphology. It was also found that diminished levels of the apical determinant Crbs (*Crumbs*) and the luminal marker *2A12* were present concomitant with increased staining in cytoplasmic vesicles within tracheal cells. The presence of GalNAc-containing glycoproteins was also greatly reduced along the apical region of the tracheal system. Ten Hagen suggested that *Pgant35A* is required to establish the proper apical composition of tracheal cells by influencing the apical delivery of proteins and/or glycoproteins. Disruption of the normal apical content results in altered cell morphology and a loss of paracellular barrier function. Hence, there seems to be a previously unrecognized requirement for mucin-type *O*-glycosylation in epithelial tube integrity that has obvious implications for epithelial morphogenesis in higher eukaryotes (Tian & Ten Hagen, 2007).

J. Dennis (Toronto, Canada) discussed the other main form of glycosylation, *N*-glycosylation, and described how key enzymes that are involved in this process are turned on and off during developmental processes. When this control system goes wrong—for example, in cancer—the changes in glycosylation can have serious effects. The final talk in this session was given by R. Gerardy-Schahn (Hanover, Germany) who discussed how a form of glycosylation called polysialic-acid glycosylation decorates the neural cell-adhesion molecule (NCAM) and thereby regulates cell–cell interactions during neural development and disease. The simplistic view that this large and negatively charged molecule is simply anti-adhesive has been found wanting in the light of many fascinating genetic studies using animals that are defective in elements of NCAM biosynthesis. In particular, Gerardy-Schahn highlighted the importance of understanding models that are designed to uncover the functions of 'naked' NCAM (that is, without polysialic acid) as opposed to the 'clothed' form or forms lacking NCAM altogether (Hildebrandt *et al*, 2007). This talk neatly led us back to the theme of new technological developments, in this case the use of complex developmental models to study a key glycosylation event in neurogenesis.

Finally, T. Merry (Manchester, UK) gave a summary of the workshop, and concluded that glycosylation is involved in a wide range of systems at many different stages. Glycosylated structures are frequently under tight spatio-temporal regulation, and this is further complicated by observations that specific combinations of these structures might be required for a particular biological effect. As highlighted throughout the meeting, model systems have proved to be exceptionally useful, and the mechanisms uncovered using these techniques have often been conserved during evolution over long periods of time and in different species. However, 'an exception proves the rule', and the importance of structures that are specific to a single species—for example, the precise type of sialic acid

in humans—cannot be underestimated, demonstrating that caution should be used in extrapolating conclusions between different species. Merry concluded that there is a need to engage more with traditional evolutionary biologists, to better integrate the methods and findings of developmental glycobiology into the mainstream. The exciting and valuable studies discussed in this EMBO workshop, which highlighted the involvement of glycosylation in various areas of development, will help to convince all that glycans have many important roles.

REFERENCES

Angata K, Lee W, Mitoma J, Marth JD, Fukuda M (2006) Cellular and molecular analysis of neural development of glycosyltransferase knock out mice. *Methods Enzymol* **417:** 25–37

Butters TD, Dwek RA, Platt FM (2005) Imino sugar inhibitors for treating the lysosomal glycosphingolipidoses. *Glycobiology* **15:** 43R–52R

Coullin P, Crooijmans RP, Fillon V, Mollicone R, Groenen MA, Adrien-Dehais C, Bernheim A, Zoorob R, Oriol R, Candelier JJ (2003) Cytogenetics, conserved synteny and evolution of chicken fucosyltransferase genes compared to human. *Cytogenet Genome Res* **103:** 111–121

Eastham AM, Spencer H, Soncin F, Ritson S, Merry CL, Stern PL, Ward CM (2007) Epithelial–mesenchymal transition events during human embryonic stem cell differentiation. *Cancer Res* **67:** 11254–11262

Freeze HH (2007) Congenital disorders of glycosylation: CDG-I, CDG-II and beyond. *Curr Mol Med* **7:** 389–396

Hart GW, Housley MP, Slawson C (2007) Cycling of *O*-linked β-*N*-acetylglucosamine on nucleoplasmic proteins. *Nature* **446:** 1017–1022

Hildebrandt H, Mühlenhoff M, Weinhold B, Gerardy-Schahn R (2007) Dissecting polysialic acid and NCAM functions in brain development. *J Neurochem* **103:** 56–64

Schwartz NB, Domowicz M (2004) Proteoglycans in brain development. *Glycoconj J* **21:** 329–341

Tian E, Ten Hagen KG (2007) A UDP-GlcNAc:polypeptide *N*-acetylgalactosaminyltransferase is required for epithelial tube formation. *J Biol Chem* **282:** 606–641

Varki A (2006) Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126:** 841–845

Varki A (2007) Glycan-based interactions involving vertebrate sialic-acid-recognising proteins. *Nature* **446:** 1023–1029

West CM, van der Wel H, Wang ZA (2007) Prolyl 4-hydroxylase-1 mediates O$_2$ signalling during development of *Dictyostelium*. *Development* **134:** 3349–3358

Wolpert L (2007) Is cell science dangerous? *J Med Ethics* **33:** 345–348

Zachara NE, Hart GW (2006) Cell signalling, the essential role of O-GlcNAc! *Biochem Biophys Acta* **1761:** 599–617

Catherine L.R. Merry   *Sviatlana A. Astrautsova*