Glycosylation: A Phenomenon Shared by All Domains of Life

Anne Dell and Federico Sastre

Abstract This chapter provides insights into why proteins are glycosylated and how their glycosylation can be characterized by mass spectrometry. The covalent attachment of carbohydrates to proteins during their biosynthesis is a phenomenon shared by all domains of life. Indeed the majority of proteins in living systems are glycosylated. Their carbohydrates play critical roles in a myriad of biological processes especially those involving recognition. They do this via engagement with carbohydrate binding proteins called lectins. For example mammalian sperm-egg engagement in the first step of fertilization involves carbohydrate-lectin recognition, and the human egg is coated with a carbohydrate sequence called sialyl Lewisx which also plays important recognition roles in the immune system.

1 An Overview

The genome sequencing projects of the past two decades have yielded many surprises, the most startling of which is unquestionably the revelation that the total number of genes in humans is not very different from many model organisms such as worms, fruit flies and simple plants. This discovery has cast a spotlight on the correlate that biological complexity is not linearly related to the number of genes among species. Why might this be the case? A variety of explanations can be offered, arising from different fields of biological research. For example, molecular biologists might suggest transcriptional regulation or epigenetic modifications as key factors. Others

A. Dell (✉) · F. Sastre
Division of Molecular Biosciences, Faculty of Natural Sciences, Centre for Integrative Systems Biology, Imperial College London, London SW7 2AZ, UK
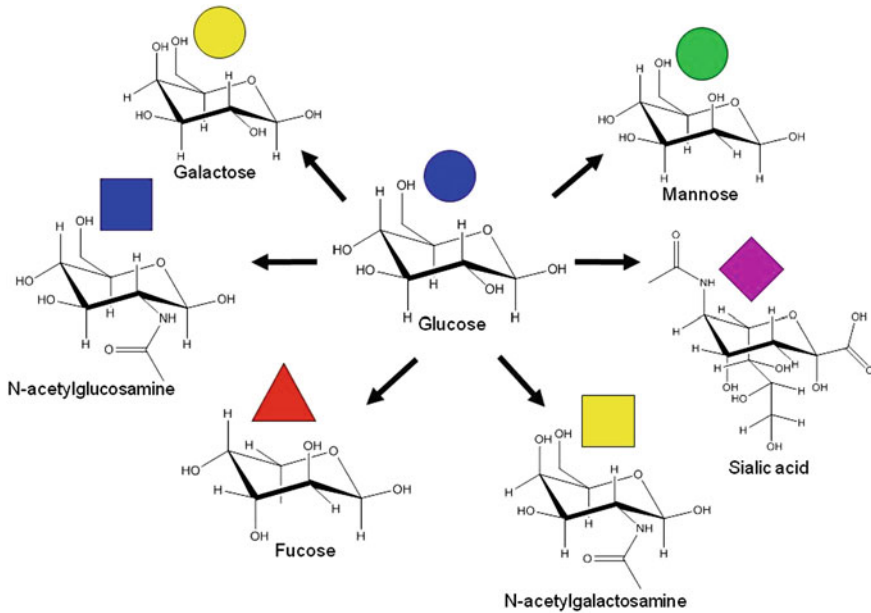e-mail: a.dell@imperial.ac.uk

Fig. 1 Structure and symbolic representations of common carbohydrates. Glucose is central to carbohydrate biosynthesis because it is made de novo from carbon dioxide and water during photosynthesis

would cite alternative splicing. We, too, believe that these phenomena contribute to biological complexity. Nevertheless we would argue that the greatest amplification of genomic information occurs after genes have been translated into proteins when the latter become modified by a myriad of functionalities. Moreover, one type of post-translational modification, namely glycosylation, results in the greatest diversity of the products of gene expression in all forms of life [1].

Today it has been well-established that protein N- and O-glycosylation (the covalent attachment of carbohydrate sequences to the side-chains of asparagine and serine or threonine, respectively) is a phenomenon shared by all domains of life. In addition, protein glycosylation has been demonstrated to be an essential requirement, rather than just an intriguing decoration. For example, correct glycosylation ensures that the plethora of proteins which eukaryotic cells use to transmit, receive and respond to chemical, electrical and mechanical signals, are expressed in functionally active forms in the right places. The information such glycoproteins mediate is essential for cells to pass through the different stages of development that occur in an organism [1, 2].

Carbohydrates have enough structural diversity to play a pivotal role as informational molecules on cell surfaces Fig. 1.

Importantly, they are in "the right place" to act as such. All eukaryotic cells are coated with a carbohydrate layer, referred to as the glycocalyx. It consists of glycoproteins and glycolipids embedded in the cell membrane, together with proteoglycans,

another class of carbohydrate biopolymer, which may be loosely associated with the eukaryotic cell surface. Prokaryotes also express glycoproteins on their surfaces. Among the prokaryotic glycoproteins, the best understood are S-layers, pilins and flagellins, plus a selection of cell surface and secreted proteins which are known to be involved in adhesion and/or biofilm formation [3, 4]. Significantly, complex carbohydrates are often highly branched and each residue can be linked to another in any of several positions on each sugar ring. This allows the formation of a large number of oligosaccharide structures from a relatively small repertoire of building blocks. Indeed even greater diversity is often conferred by the addition of functional groups such as sulfates, phosphates, acetyl and methyl groups.

How do carbohydrates on cell surfaces fulfil their "information" roles? This is most often achieved by engagement with partner molecules on other cells thereby triggering adhesive and/or signalling events. These carbohydrate binding partners are called lectins [5]. Thanks largely to the Consortium for Functional Glycomics (CFG) (which was funded by the US National Institutes of Health to provide tools and resources to the international research community to understand the role of carbohydrate-protein interactions), scientists from all disciplines can readily access information pertaining to how surface carbohydrates and complementary lectins on opposing cell surfaces mediate cell-to-cell recognition. Thus the CFG website [2] provides a rich source of information and data which facilitates the engagement of researchers, unfamiliar with carbohydrates, with experts working in the field of glycobiology. Figure 2 illustrates several of the best understood biological interactions where carbohydrate-lectin recognition plays a central role. The meaning of the symbols used in the figure is explained in Fig. 2.

The tools of modern mass spectrometry have been crucial for unravelling the carbohydrate mediated processes exemplified in Fig. 2 [6–10]. Mass spectrometry is an enormously powerful tool for high sensitivity sequencing of complex carbohydrates. Its versatility permits the analysis of all families of glycopolymers. Moreover, complex mixtures of glycoproteins are not a problem for mass spectrometric analysis. Indeed, glycomic methodologies are capable of defining the carbohydrate sequences constituting the glycocalyx of tissues or cells without the need for time consuming purifications [11]. Glycomics research of the past decade, much of it supported by the CFG, has yielded substantial quantities of public data which are facilitating worldwide research addressing the roles of carbohydrates and lectins in complex systems [2, 12].

We hope that those reading this article are now stimulated to learn more about glycosylation and biological complexity. If this is the case, the CFG website is an excellent place to start your journey [2]. To whet your appetite we end our article with an introduction to an evolving story in glycobiology which has as its central character a famous carbohydrate moiety called sialyl Lewis[x] (Fig. 2).

First identified in rat brain glycoproteins in the 1970s, this carbohydrate was revealed, by the emerging glycomic strategies of the mid 1980s, to be present on human white blood cells and enriched in cancer cells. A few years later the Selectins were discovered. These constitute a lectin family that recognise sialyl Lewis[x] as their primary ligand. The Selectins play pivotal roles in lymphocyte trafficking and recruit-
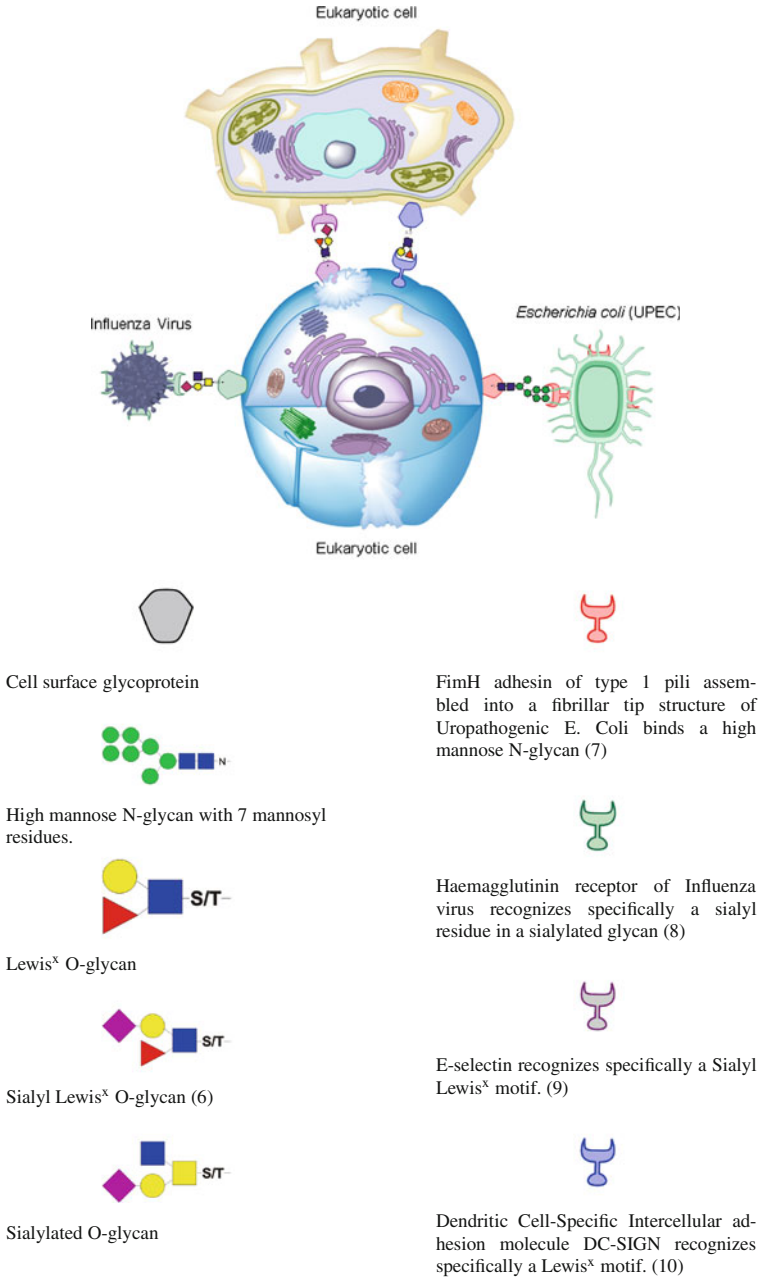
Symbols used in Fig. 2

Fig. 2 Glycan–lectin recognition is key to cell-to-cell communication

ment of neutrophils to sites of inflammation. Their discovery energised the field of glycobiology and spawned numerous biotech companies intent on developing new anti-inflammatories and anti-cancer agents. The high hopes for glyco-therapeutics that prevailed in the early 1990s continue to this day, but are now tempered by realism i.e. it takes a very long time to understand the processes mediated by carbohydrate recognition and even longer to develop effective therapies based on intercepting these processes.

Very recently sialyl Lewis[x] has re-appeared in the headlines of both scientific and lay articles. This is because of exciting discoveries concerning human reproduction [13]. Ultra-high sensitivity mass spectrometric analyses have now provided the first molecular insights into the recognition processes occurring at the very start of human life, when a single sperm first engages with the surface of a human egg. This research has shown that multiple sialyl Lewis[x] sequences are attached to the proteins constituting the jelly-like coat of the human egg, which is called the zona pellucida. Remarkably the density of sialyl Lewis[x] moieties on the human egg is orders of magnitude higher than on white blood cells, consistent with it playing a pivotal role in sperm recognition [13]. Interestingly human sperm do not express any of the known Selectins. Hence the race is now on to find the putative "Selectin-like" molecule on sperm that binds to the sialyl Lewis[x] sequence on the human egg.

References

1. A. Varki, R. Cummings, J. Esko, H. Freeze, P. Stanley, C. Bertozzi, G. Hart, M. Etzler, *Essentials of Glycobiology*, 2nd edn. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2009)
2. Consortium for Functional Glycomics (CFG) web page, http://www.functionalglycomics. org/fg/
3. C. Szymanski, B. Wren, Protein glycosylation in bacterial mucosal pathogens. Nature Rev. Microbiol. **3**(3), 225–237 (2005)
4. A. Dell, A. Galadari, F. Sastre, P. Hitchen, Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. Int. J. Microbiol. **2010**, 148178 (2010). doi:10. 1155/2010/148178
5. N. Sharon, H. Lis, History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology **14**(11), 53–62 (2004)
6. M. Fukuda, E. Spooncer, J. Oates, A. Dell, J. Klock, Structure of sialylated fucosyl lactosamino-glycan isolated from human granulocytes. J. Biol. Chem. **259**(17), 10925–10935 (1984)
7. C. Jones, J. Pinkner, R. Rotht, J. Heusert, A. Nicholes, S. Abraham, S. Hultgren, Fimh adhesin of type 1 pili is assembled into a fibrillar tip structure in the enterobacteriacea. Biochemistry. Proc. Natl. Acad. Sci. U.S.A. **92**, 2081–2085 (1995)
8. W. Weis, J. Brown, S. Cusack, J. Paulson, J. Skehel, D. Wiley, Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature **333**, 426–431 (1988)
9. M. Sperandio, C. Gleissner, K. Ley, Glycosylation in immune cell trafficking. Immunol. Rev. **230**(1), 97–113 (2009)
10. K. van Gisbergen, M. Sanchez-Hernandez, T. Geijtenbeek, Y. van Kooyk, Neutrophils mediate immune modulation of dendritic cells through glycosylation-dependent interactions between mac-1 and dc-sign. J. Exp. Med. **201**(8), 1281–1292 (2005)

11. S.J. North, J. Jang-Lee, R. Harrison, K. Canis, M. Nazri, A. Trollope, A. Antonopoulos, P.C. Pang, P. Grassi, S. Al-Chalabi, T. Etienne, A. Dell, S.M. Haslam, Chapter two-mass spectrometric analysis of mutant mice. Methods Enzymol. **478**, 27–77 (2010)
12. A. Antonopoulos, S.J. North, S.M. Haslam, A. Dell, Glycosylation of mouse and human immune cells: insights emerging from n-glycomics analyses. Biochem. Soc. Trans. **39**(5), 1334–1340 (2011)
13. P.C. Pang, P. Chiu, C.L. Lee, L.Y. Chang, M. Panico, H.R. Morris, S.M. Haslam, K.H. Khoo, G.F. Clark, W.S. Yeung, A. Dell, Human sperm binding is mediated by the sialyl-lewis^x oligosaccharide on the zona pellucida. Science **333**(6050), 1761–1764 (2011)