Glycans – the third alphabets of life

Sambasivan Venkat Eswaran

Amino acids and nucleotide bases (adenine, thymine, guanine and cytosine) constitute the first and the second alphabets of life. Glycans, i.e. carbohydrates bound to proteins are considered as the third alphabets of life. In addition to glucose, other monosaccharides like mannose, sialic acid, fucose, etc. are constituents of glycans. Like the genetic code, the importance of the sugar code is now being understood. Glycans are involved in the classification of blood groups. Studies on mechanism of binding of viruses to glycans leading to influenza and H1N1 flu have led to the development of new drugs to treat such infections. Glycans in nanotechnology (glyconanotechnology) is an emerging area of study as carriers for vaccinations, drug delivery in cancer detection, and as energy sources.

Scientists and the general public alike have become accustomed to the amino acids and the nucleotides as being the two sets of alphabets of life. The N-terminal and the C-terminal ends are the elongation points in amino acids, while the 3′- and 5′-ends are points for linear extension of nucleotides. Undoubtedly, amino acids have played the most important role in the origin of life. Some of these are referred to as 'essential' amino acids and are the building blocks of proteins and enzymes, which have led to science of proteomics. Ever since the structure of DNA was unraveled by Watson and Crick, nucleotides have continued to occupy centre stage and led to the emergence of genetic engineering, biotechnology and genomics. The complete successful sequencing of the human genome has led to the hope that the era of personalized medicine is almost here.

The glycans

It is in this context, that another class of compounds, the glycans have emerged as the 'third alphabets of life'. There was much confusion in literature[1] whether carbohydrates are impurities in proteins or whether proteins were the contaminants in this class of compounds. The IUPAC-IUB definition of glycans removed this confusion and established that these are compounds of carbohydrates bound to proteins, keeping nucleotides out of their purview. Glycans, could be co- or post-translational modifications of proteins. Unlike the amino acids and the nucleotides, the carbohydrates offer multiple positions for elongation and branching and therefore provide many more possibilities for encoding and storing information. Reading and decoding such data will obviously pose many more challenges than hitherto has been the case. Under the term 'sugar code', Gabius[2] talks about new alphabets, words, the antennae, postal codes read by a process similar to that a courier delivery person follows. Tools for reading sugar coded information are available and are the receptors called lectins (from legere; to read in Latin). Two major types of glycans are the *N*-glycans and the *O*-glycans. The former is attached to proteins through Asn Xxx Ser/Thr, where Xxx can be any amino acid other than proline. *O*-glycans, on the other hand, are linked via hydroxyl amino acids. These possess a reducing end and a non-reducing end.

Which are then the new alphabets, words? These naturally come from the monosaccharaides; Glucose has been the most well known member form this class. However, from the point of view of this paper, D-galactose, D-mannose, D-xylose, D-glucoronic acid, N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), L-fucose and sialicacid are also equally important.

One letter abbreviations for the known amino acids are now in common use and everyone is familiar with the use of alphabets A, T, G, C in the context of DNA and the genetic code. Symbols and characters of new letters/alphabets for monosaccharides of the 'sugar code' have already been generated, e.g. a filled circle with different colors represents D-glucose (Glc; blue), D-mannose (Man, green) and D-galactose (Gal, yellow) respectively. Other symbols used include, filled/crossed square, divided diamond, filled/divided triangle, flat rectangle, filled star, flat hexagon and pentagon[3]. These look very impressive and remind one of a veritable Periodic Table.

As early as 1966, the importance of sialic acid as a receptor for influenza

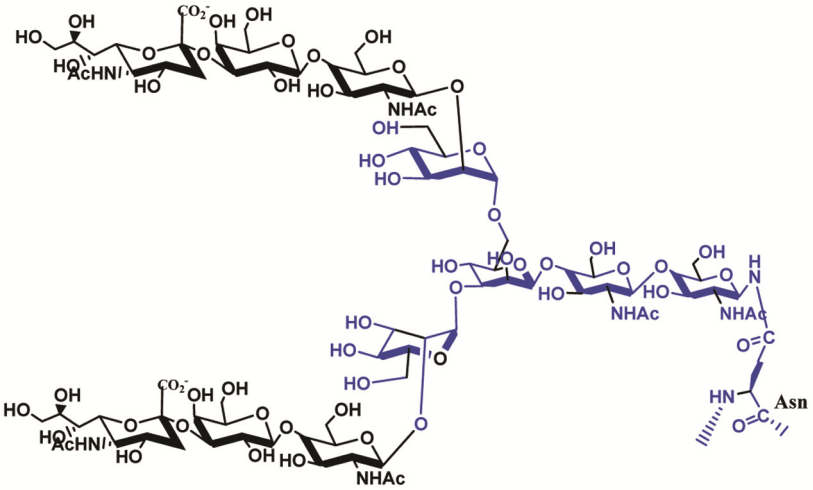

Figure 1. Structure of a *N*-glycan with a core pentasaccharide (in blue), with two antennae (in black). Figure adapted from ref. 4.

virus with a core pentasaccharide (in blue) and the antennae (in black) is well known (Figure 1).

Soya bean agglutinin (SBA), $Man_9(GlcNAc)_2$ (Figure 2) is one of the earliest examples of a glycoprotein and is found in human urine up to 100 mg/day and is referred to as uromodulin when isolated from human pregnancy urine[5], the amino acid sequence remaining the same.

SBA is routinely used in hospitals during bone marrow transplants, is used to treat patients of leukemia and is used for immune deficient 'bubble children'. Blood serum is a rich source up to 150 known glycoproteins[6]. Hen egg white ovalbumin (42 kDa) was isolated and crystallized at the turn of the previous century[7]. Pine apple stem gives a homogeneous glycoprotein[8]. The copper containing blue blood of scorpions[9] contains haemocyanin. Glycoproteins have also been isolated for molluscs[10]. The best known examples of glycoproteins are erythropoietin, interferons, blood clotting factors whose sales exceed US$ 3 billion per annum. Phytoagglutinin has been synonymous with plant lectins but that changed after the identification of the first mammalian glycan (from rabbit liver), and has the structure GlcNAc (β1-4) MurNAc, where GlcNAc (β1-4) stands for *N*-acetyl glucosamine and MurNAc stands for *N*-acetyl muramic acid. Similarly, lysozyme digests of bacterial cell walls yielded a tetrasaccharide. Recombinant glycoproteins have come into contention as these can be produced in larger amounts as and when required. It has been stated 'that sugar specific adhesion might be pre requisite for bacterial colonization and infection was not considered at all though it was known that influenza virus requires its attachment to sialic acid on cells'. Seven such papers by Ofek *et al.*[11] have been cited 4500 times. Understanding the importance of the adhesion of leukocytes and their hampered flow being linked to asthma, arthritis and inflammation are leading to anti-adhesion therapy being developed. Infection of mouse bladders with a mannose-specific *E. coli* was markedly diminished by methyl alpha mannoside[12]. Much work has been done on ^{1}H-NMR and 2D-NMR by the Utrecht group led by Vliegenthart (loc. cit.) over four decades and these studies help in tracking the changes, especially before and after a periodic acid (HIO_4) treatment.

GlcNAc–MurNAc is a disaccharide fragment commonly present in bacteria. The very minor difference ($NHCOCH_3$ and OH) between, for example, blood group A and B can make a difference between life and death, if the wrong blood group is transfused to a patient[4]. The crystal structure of a H1N1 virus is shown in Figure 3.

H1N1 virus causing swine flu attaches to the exposed sialic acids, the virus then enters the host cell and takes over the cellular genetic machinery. As it starts bursting out its own copies, it ensures that the original binding to sialic acid is cleaved by neuraminidase to ensure that it does not become a hindrance for the

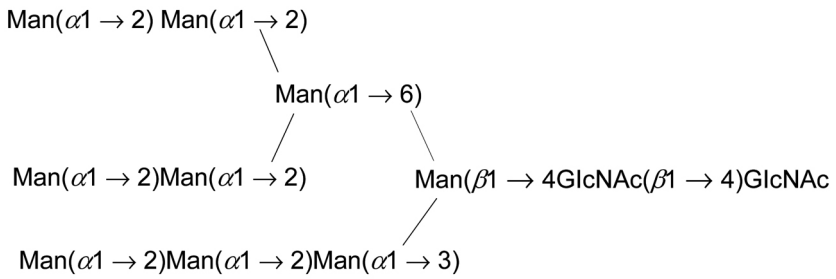

Figure 2. Oligomannoside, $Man_9GlcNAc_2$ from soya bean agglutinin.

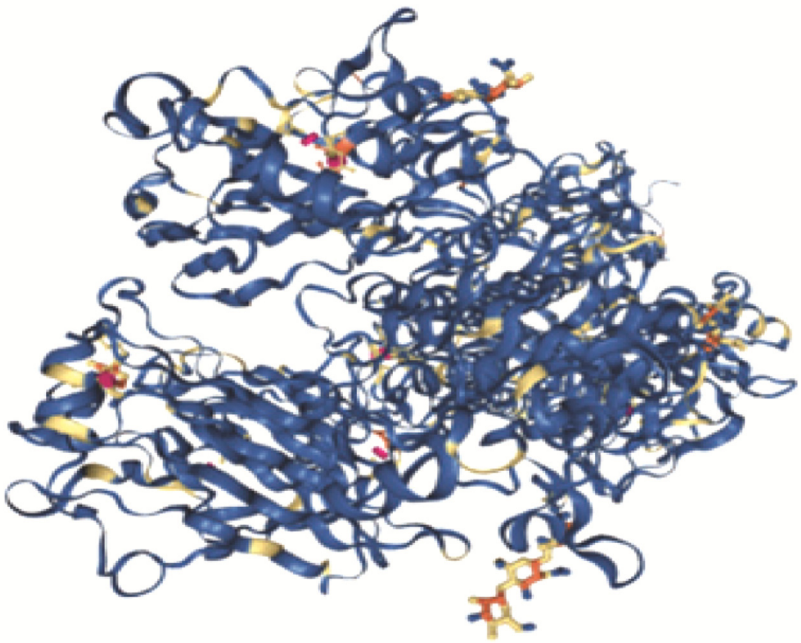

Figure 3. Crystal structure of a H1N1 influenza virus, 2009; doi:10.2210/pdbSLG/pdb.

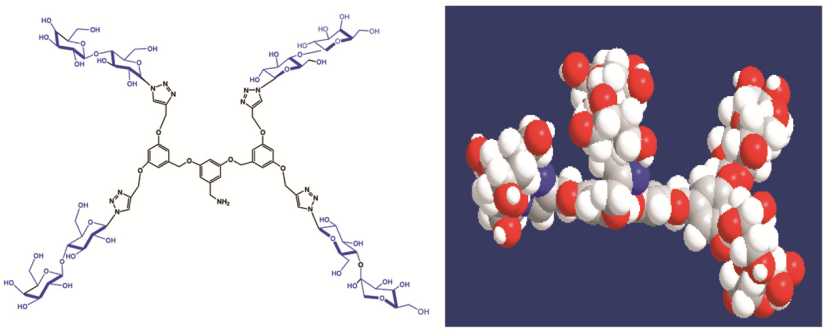

Figure 4. A synthetic glycan adapted from ref. 1 (c) and its 3D structure.

virus to spread. The name H1N1 stems from the involvement of both hemagglutinin (H) and neuramidinase (N). A detailed study of the cleavage step showed the involvement of a sp^2 hybridized carbocation intermediate and the design of a drug resembling this intermediate led to the successful development of the drug Tamiflu to treat H1N1 flu. Sulfated glycans present on cell surface are important for normal growth of axons[13,14]. These interact with selectins (L, P and E), which, too, are glycoproteins; low sulfated levels retard axon growth. Synthetic glycans are now available which could serve as vaccines (Figure 4).

Glycans and nanotechnology

With diminishing petroleum reserves, plant glycans and lectins are increasingly being sought after as, 'energy sources, building materials, clothes, paper products, animal feed, and food and beverage additives'. In this context, immense possibilities arise, when glycans and nanotechnology (glyconanotechnology) come together, e.g. in their use as 'vehicles for vaccination and drug delivery, for detection of cancer and other pathogens'. The agglutinin (RCA 120) of the castor oil plant, *Ricinus communis* and the cholera toxin induce a reversible colour change in gold nanoparticles, which is useful in colorometric studies of binding of biological molecules. α-D-mannosyl [60]fullerenes ('sugar balls') prepared via the alkyne-azide 'click' reaction have been shown to inhibit erythrocyte aggregation and are nearly-spherical glycans, with 0.7-nm diameter, constituting the smallest 'nanoparticles' for glyconanotechnology[15]. It has been stated that 'translation of fundamental studies to biomedical glyconanotechnology appears to be imminent'.

Summary

Glycans, carbohydrates bound to proteins, have emerged as the third alphabets of life, whereas amino acids and the nucleotide bases (adenine, thymine, guanine and cytosine) are the first two alphabets of life. Although the genetic code is well established, the sugar code is now being actively explored. Glycans are important in blood group classification and soya bean agglutinin is commonly used in hospitals during bone marrow transplants. Studies on binding of influenza virus to sialic acids in the onset of influenza and H1N1 flu have led to the successful development of drugs like Tamiflu. Biomedical glyconanotechnology is emerging as an attractive area of study as vehicles for vaccines, drug delivery and detection of cancer and also as energy sources, in food and beverage additives.

1. Sharon, N., *Acta Anat*, 1998, **161**, 7–17.
2. (a) Hans-Joachim Gabius (ed.), *The Sugar Code: Fundamentals of Glycosciences*, Wiley-Blackwell (ISBN: 978-3-527-32089-9), 2009; (b) Roth, J. and Gabius, J.-H., *Histochem. Cell Biol*., 2017, **147**, 111–117, doi:10.1007/s00418-016-1521; (c) Andre, S., Kaltner, H., Manning, J. C., Murphy, P. V. and Gabius, J.-H., *Molecules*, 2015, **20**, 1788–1823.
3. Varki, A. (ed.), *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, New York, 2017, 3rd edn; https://www.ncbi.nlm.nih.gov/books/NBK310274/
4. Bertozzi, C. R., https://www.youtube.com/watch?v=WCbg-kOY_8E
5. (a) Vliegenthart, J. F. G., *Proc. Jpn. Acad. Ser*., 2017, **B93**, 64–93; (b) Tamm, I. and Horsfall, F. L., *Proc. Soc. Exp. Biol. Med*., 1950, **74**, 108–114.
6. Yang, Z., Hancock, W. S., Chew, T. R. and Bonilla, L., *Proteomics*, 2005, **5**(13), 3353–3356.
7. Neuberger, A., *Trends Biochem. Sci.*, 1988, **13**, 398–399.
8. Scocca, J. and Lee, Y. L., *J. Biol. Chem.*, 1969, **244**, 4852–4863.
9. Ali, S. A., Zaidi, Z. H. and Abbasi, A., *Comp. Biochem. Physiol. A Physiol.*, 1995, **112**(1), 225–232; Debeire, P. and Montreuil, J., *Carbohydr. Res.*, 1986, **151**, 305–310.
10. (a) Van Kuik, J. A., Sybesma, R., Kamerling, J. P., Vliegenthart, J. F. G. and Wood, E. J., *Eur. J. Biochem*., 1987, **169**, 399–411; (b) Van Kuik, A. J., Sijbesma, R., Kamerling, J. P., Vliegenthart, J. F. G. and Wood, E. J., *Eur. J. Biochem*., 1986, **160**, 621–625.
11. Ofek, I., Mirelman, D. and Sharon, N., *Nature*, 1977, **265**, 623–625.
12. Aronson, M., Medalia, O., Schori, L., Mirelman, D. and Sharon, N., *J. Infect. Dis.*, 1979, **139**(3), 329–332.
13. Cavalcante, L. A. *et al.*, *Braz. J. Med. Biol. Res.*, 2003, **36**(8), 993–1002; http://dx.doi.org/10.1590/S0100-879X2003000800005.
14. Velasco, S., Díez-Revuelta, N., Hernández-Iglesias, T., Kaltner, H, André, S. Gabius, H. J. and Abad-Rodríguez, J., *J. Neurochem*., 2013, **125**(1), 49–62; doi:10.1111/jnc.12148.
15. Penadés, S., Davis, B. G. and Seeberger, P. H., *Essentials of Glycobiology* [Internet], Chapter 58 in ref 5, Cold Spring Harbor, NY, 3rd edn.

ACKNOWLEDGEMENTS. I thank Dr Alok Adholeya, Director, Teri Deakin Nano Biotechnolgy Centre, Gurugram, for research facilities. I thank Mr Sujeet Kumar Thakur for help with the drawings.

Sambasivan Venkat Eswaran is in the Teri Deakin Nano Biotechnology Centre, Teri Gram, Gwal Pahari, Gurgaon–Faridabad Expressway, Gurugram 121 002, India.
e-mail: sv.eswaran@gmail.com