“In analogy to the current situation in Cosmology, glycans can be considered as the “dark matter” of the biological universe: a major and critical component that has yet to be fully incorporated into the “standard model” of biology.”

Ajit Varki and Stuart Kornfeld

Outlining the topic and its importance

Glycomics and glycoproteomics define powerful technologies that help scientists to capture this biological “dark matter” and study all glycans and glycoproteins expressed by cells, tissues or organisms at a given time, space and condition.2–4 While both glycomics and glycoproteomics methods have existed for several years,5–8 key advances in the separation sciences, mass spectrometry and informatics have recently allowed for a more complete implementation and integration of glycomics and glycoproteomics in the life sciences. As exemplified by papers published in this themed issue in Molecular Omics “Glycomics & Glycoproteomics: From Analytics to Function”, studies that concertedly apply two or more glyco-centric ‘omics techniques to uncover new mechanisms in various research areas in glyobiology such as pathogen–host interactions (Cain et al., DOI: 10.1039/D0MO00032A; Mule et al., DOI: 10.1039/D0MO00043D; Delannoy et al., DOI: 10.1039/C9MO00173E; Mthembu et al., DOI: 10.1039/ C9MO00175A) cancer and biomarkers (Chandler et al., DOI: 10.1039/D0MO00098D; Acharya et al., DOI: 10.1039/C9MO00061E) and embryogenesis (Qu et al., DOI: 10.1039/D0MO0005A), are becoming common in the literature. This themed issue also illustrates that mass spectrometry-based glycoproteomics studies often use multiple orthogonal fragmentation schemes to produce complementary glycan structural information to aid glycopeptide characterisation at scale, with sensitivity, speed and accuracy (Pap et al., DOI: 10.1039/C9MO00160C; Chalkley et al., DOI: 10.1039/C9MO00178F). Other new technology improvements (Cordova et al., DOI: 10.1039/C9MO00181F; Zhang et al., DOI: 10.1039/C9MO00180H; Sethi et al., DOI: 10.1039/D0MO0019A), informatics advancements (Chalkley et al., DOI: 10.1039/ C9MO00178F; Uh et al., DOI: 10.1039/ C9MO00174G; Phung et al., DOI: 10.1039/ C9MO00125E) and pathway analyses (del Solar et al., DOI: 10.1039/D0MO00023J; Donald et al., DOI: 10.1039/C9MO00168A) also contribute to an ever-expanding analytical toolbox that now allows scientists to explore not only the structure but also the biosynthesis and function of glycans and glycoproteins.

The field has for years recognised that no single technology can provide an in-depth, often overlooked, understanding of the compositional, structural and functional heterogeneity inherently associated with protein glycosylation. Nonetheless, integration of glycomics and glycoproteomics with the many interfacing ‘omics disciplines has only commenced recently. Multifaceted analytical strategies and the integration of glycomics and glycoproteomics with proteomics, transcriptomics, metabolomics and genomics are critical to dive deeper into the immensely complex and dynamic glycoproteome, the underpinning glycosylation machinery and their effects on cellular activity. Glycobiology, arguably more than other biological research areas, benefits from the synergy that can be accessed using multi-omics approaches as, for example, demonstrated in “glycomics-assisted glycoproteomics” studies.9–11 In such studies, the glycomics data not only provide essential information on the fine detail of the attached glycan structures and define the boundaries of the glycome, but also enable more effective characterisation of the accompanying glycoproteomics data.

Cite this: Mol. Omics, 2021, 17, 8

DOI: 10.1039/d0mo90019b

rsc.li/molomics

Morten Thaysen-Andersen, Daniel Kolarich and Nicolle H. Packer

*Contributed equally.
and give insights into the role of glycans in protein function.

The body of research papers compiled in this themed issue and elsewhere clearly shows that glycomics and glycoproteomics have matured considerably over the past years. As the wider scientific community increasingly accesses these technologies, and the workflows and data collection, interpretation, reporting and sharing are being streamlined, we are beginning to unlock the real potential of glycomics and glycoproteomics to generate new insights into how the heterogeneous glycosylation of proteins is involved in so many biological processes.

Future directions of the field

At present, the mass spectrometric technologies, when applied to mixtures of intact glycopeptides, mostly assign the monosaccharide compositions of glycans to peptides, and only in favourable cases to discrete sites. However, more intricate structural features of glycans are often required to understand the effect of glycosylation on protein function in such areas as cellular interactions, protein cleavage, immune recognition and in biomarker discovery. Technologies that can uncover the heterogeneity of detailed structures decorating each protein site remain at this time a “holy grail” of glycobiology. Excitingly, innovative developments in “structure-focused glycoproteomics” employing, for example, diagnostic tandem mass spectrometry fragment ions,12–14 sequential exoglycosidase digestions on glycopeptides,15,16 differential liquid chromatography and ion mobility based retention,17–21 and linkage-specific derivatisation of glycopeptides,22 to annotate certain glycan structural features together with site information, are beginning to appear in the literature.

The other missing link for biological researchers is the means to be able to more easily interpret the volume of data obtained by the specialised mass spectrometric approaches that are being constantly expanded to cover top down, middle down and bottom up analyses of glycoproteins.23–25 The development of glycobioinformatics tools and curated databases, together with the implementation of standardization of reporting guidelines such as MIRAGE (Minimum Information Required For a Glycomics Experiment)26–29 (https://www.beilstein-institut.de/en/projects/mirage/), to cope with these data are essential for the continued penetration of glycobiology into the many streams of life sciences. The agreed implementation of a common “glyco-language” via the “Symbol Nomenclature for Glycans (SNFG)”,30,31 together with the assignment of unique accession numbers to the glycan structures (https://glytoucan.org/) has started to unify and facilitate the depiction and data sharing of complex biomolecular structures across the chemical and biological sciences.

Despite this progress, the most common response of scientists to the question as to why they have not explored the glycobiology of their research question at hand has been that it is “too hard” – this reluctance can only be addressed by the availability of user friendly methods, software and accurate knowledge bases of glyco-biological information. This information needs to be not only that acquired by the technologies of glycomics and glycoproteomics, but the seamless integration of such datasets with information delivered by well-established genomics, transcriptomics, proteomics, metabolomics and/or lipidomics technologies for an integrated and holistic understanding of cellular biology in health and disease, such as has now begun with the international glycoinformatics initiatives of GlyConnect (https://glyconnect.expasy.org/), Glygen (https://glygen.org/) and GlyCosmos (https://glycosmos.org/) under the collaborative GlySpace Alliance.32

Once the dynamics and functional glycan features are determined on proteins within a complex biological system it is important to translate these findings into useful practice. Coordinated efforts in glycobio-technology and clinical glycoscience that aim to connect fundamental and applied glycobiology and to translate findings in this field into products including new therapeutic targets to treat, cure and ultimately prevent disease and new biomarkers to diagnose and prognose diseases, are ongoing.33–35

Not to be forgotten in the future, of course, is that glycomics does not only apply to the structure and function of proteins, but also refers to the characterisation of glycans that are found attached to lipid anchors or found as free glycans, proteoglycans, bacterial lipopolysaccharides and peptidoglycans. At this time, the technologies for the analysis of these molecules are in their infancy, and are practised by few laboratories world-wide, but must be taken into consideration as we come to realise that multi-omics analyses must be addressed across the entire breadth of molecular sciences if we are to understand the complex systems that are biology.

References


