CHEMBIOCHEM
Viewpoints

Revisiting the Language of Glycoscience: Readers, Writers and Erasers in Carbohydrate Biochemistry

Simone Dedola,[a] Michael D. Rugen,[b, c] Robert J. Young,[d] and Robert A. Field*[a, b, e]

The roles of carbohydrates in nature are many and varied. However, the lack of template encoding in glycoscience distances carbohydrate structure, and hence function, from gene sequence. This challenging situation is compounded by descriptors of carbohydrate structure and function that have tended to emphasise their complexity. Herein, we suggest that revising the language of glycoscience could make interdisciplinary discourse more accessible to all interested parties.

The perception that glycoscience—the chemistry and biology of carbohydrates—is both complex and ubiquitous in nature[1,2] has led to the notion that "*carbohydrates in molecular biology are like dark matter in the universe… poorly studied yet crucial to a full understanding of how things actually work*".[3] In contrast to nucleic acids and proteins (DNA codes for RNA codes for protein), the lack of template-encoding disconnects the "glyco code"[4] from direct gene sequence control. This results in carbohydrate biosynthesis and the biological function of glycans being dependent upon a series of protein–carbohydrate interaction events. Overall, the concerted actions of lectins, glycosyltransferases and/or glycoside hydrolases achieve the integrity of mature bioactive glycan structures. The intricacies of this landscape are made worse by the tendency of the glycoscience community to emphasise the complexities of the field, perhaps making it less accessible to the casual reader—the informed non-expert—than it needs to be. The glycoscience community are not alone in this shortcoming, as highlighted by a Comment in *Nature* that suggests that "*Antibiotic resistance has a language problem. A failure to use words clearly undermines the global response to antimicrobials' waning usefulness*".[5] Technological[6] and informatics[7] advances in glycoscience, alongside combinations of the two,[8] are providing new ways to cut through the complexity, whilst comprehensive books of glycobiology topics provide entries in to the field.[9] The introduction of stylized symbol nomenclature for glycans (SNFG; Figure 1) also represents an important step towards simplifying communication within and between interested disciplines[10] along with guidelines for experimental design and data curation,[11] and a repository for glycan structures.[12]

As discussed recently by Gabius,[13] there are notable parallels between aspects of glycobiology and the epigenetic regulation of chromatin structure and function. The latter processes, which occur with exquisite precision, are typically referred to in stripped-down terms as a series of *read*, *write* and *erase* events, making the field immediately accessible to outsiders. Indeed, this approach emulates computer programming's create, read, update and delete (CRUD)[14]—the four basic functions employed for persistent data storage.[15] Herein, we consider the potential to recapitulate glycoscience language in the terms of epigenetic vocabulary.

In simple terms, epigenetics concerns small chemical changes (marks) in the chemical structure of chromatin—typically the histone proteins that organize and package DNA in chromosomes.[16] Dynamic changes in these epigenetic protein marks impact on the physical accessibility of gene sequences for expression, rather than on the alteration of the genetic code per se. The profound biological consequences of these processes have attracted enormous attention over the past decade, given their central role in life and their disruption in disease.[17] The molecular hallmarks of epigenetic regulation comprise a dynamic series of enzymatic modification steps that introduce or remove marks to the histone protein structure. Epigenetic writers, which introduce epigenetic marks on amino acid residues of the histones, include histone acetyltransferases (HATs, which N-acetylate lysine), histone methyltransferases (HMTs, which N-methylate lysine), protein arginine methyltransferases (PRMTs) and protein kinases (which O-phosphorylate serine/threonine), amongst others. Epigenetic readers, which bind to epigenetic marks and amplify their impact on DNA packaging and hence gene accessibility for expression,

[a] Dr. S. Dedola, Prof. R. A. Field
Iceni Diagnostics, The Innovation Centre
Norwich Research Park, Norwich, Norfolk NR4 7GJ (UK)
E-mail: rob.field@icenidiagnostics.com

[b] Dr. M. D. Rugen, Prof. R. A. Field
Department of Biological Chemistry, John Innes Centre
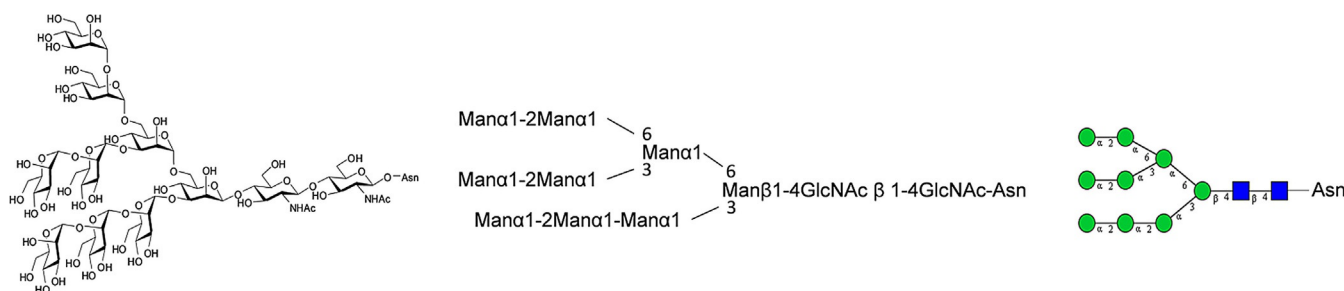Norwich Research Park, Norwich, Norfolk NR4 7UH (UK)

[c] Dr. M. D. Rugen
Present address: Cobra Biologics, Science Park, University of Keele
Newcastle-under-Lyme, Staffordshire ST5 5SP (UK)

[d] Dr. R. J. Young
Medicinal Chemistry, Medicines Research Centre, GlaxoSmithKline
Stevenage, Hertfordshire SG1 2NY (UK)

[e] Prof. R. A. Field
Present address:
Department of Chemistry and Manchester Institute of Biotechnology
University of Manchester
Manchester, M1 7DN (UK)

The ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201900377.

Figure 1. Representing glycan structures: simplification and standardization with stylized SNFG. Taken from ref. [10].

include proteins containing bromodomains, chromodomains and Tudor domains. Epigenetic erasers, such as histone deacetylases (HDACs), lysine demethylases (KDMs) and phosphatases, catalyse the removal of epigenetic marks (Figure 2).

The impact of the lysine N-acetylation epigenetic mark is perhaps simplest to appreciate. Writing this mark results in the loss of a positive charge on the lysine side chain of a histone, thus removing the potential for interaction with the negatively charged DNA backbone and causing loosening the DNA–histone complex. The resulting opening up of the chromosome structure enables the localized activation (turning on) of gene expression. In the opposite sense, erasing a lysine acetylation mark drives a tighter assembly of the histone–DNA complex and silencing (turning off) gene expression.

The general principle of readers, writers and erasers prompts consideration of potential parallels between epigenetics and the control of glycan biosynthesis, structure and function. That is, does the notion of lectin readers, glycosyltransferase writers and glycosyl hydrolase erasers ring true in glycobiology? A

convenient segue from epigenetics into glycoscience is provided by the reversible O-GlcNAc modification of Ser/Thr residues in proteins.[18] This central metabolic "rheostat"[19] comprises a nutrient status-responsive, post-translational modification that impacts on protein–protein and protein–nucleic acid interactions. In turn regulating of cellular events including transcription and signal transduction, with implications in diabetes, Alzheimer's disease and cancer.

So how does the O-GlcNAc cycle work? O-GlcNAc transferase (OGT) writes and O-GlcNAcase erases, providing a simple and reversible modification cycle that is orthogonal to protein phosphorylation and which has far-reaching physiological impact (Figure 3).[20]

In addition to glycosyltransferase writers and glycosyl hydrolase erasers, there are also potential readers in glycoscience—a function performed by lectins[21] and the carbohydrate-binding modules (CBMs)[22] in multidomain CAZymes. The full read, write, erase combination in glycoscience is most easily exemplified by the proofreading and editing cycle associated with
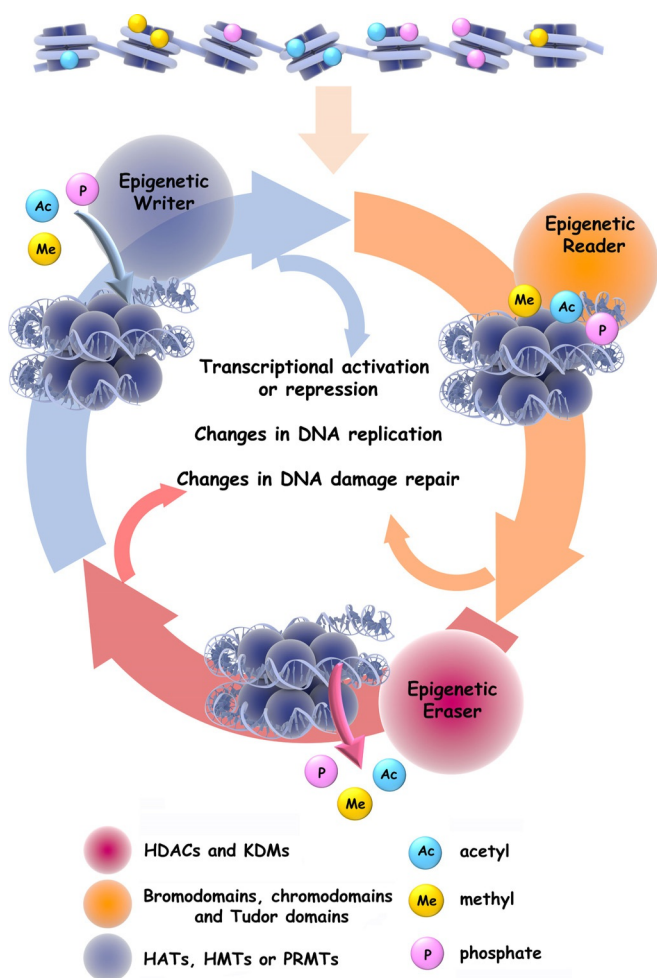
Figure 2. Epigenetic writers, readers and erasers. DNA packaged around histones gives a condensed genomic information package (top) that can be selectively unwound by epigenetic modification (e.g., acetylation, methylation of phosphorylation) to expose genes for transcription (turn on). Abbreviations used are given in the text. Adapted from ref. [17a].

N-linked glycoprotein biosynthesis. These processes are essential to ensuring the correct integrity and dynamics of cell-surface glycoproteins, which contribute to the glycocalyx that dominates cell–cell interactions in the maintenance of healthy tissue and which underpin sperm–egg interactions during fertilisation, but which also serve as cellular receptors for a wide range of microbial pathogens.[9]

Asparagine-linked protein N-glycosylation starts in the endoplasmic reticulum, whereas the peptide chain is unfolded, and proceeds through protein folding to the Golgi apparatus, where the glycan components are processed to a mature state. This requires a highly organised distribution of processing machinery to achieve the fidelity and quality control needed to ensure biological function.[23] Approximately 80% of the proteins entering the secretory pathway are glycosylated in the ER and most of the proteins assembled in the ER feature N-linked oligosaccharides. Most of the glycoproteins featuring mature *N*-glycans are, as described by Aebi, "precisely heterogeneous" in their carbohydrate composition—a result of kinetically controlled processing.[24] Nonetheless, to reach their final
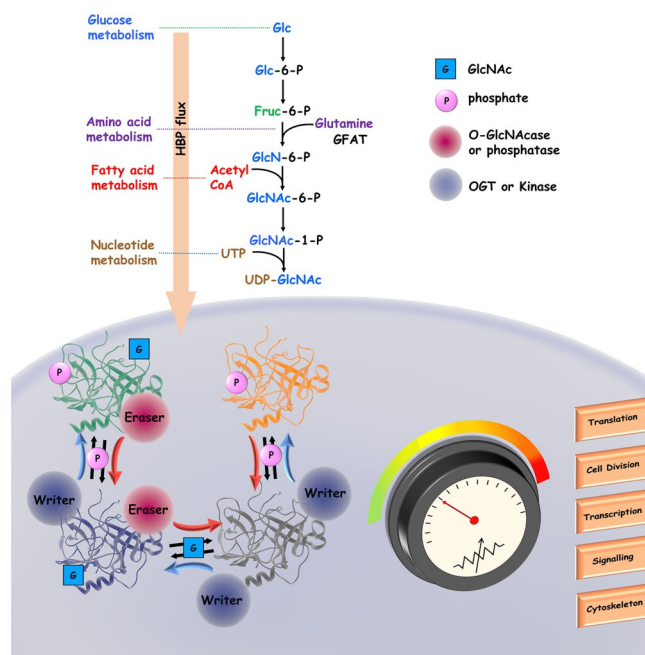

Figure 3. The *O*-GlcNAc cycle and its impact on the modulation of cellular processes. Adapted from ref. [19].

mature and bioactive form, in the early stage of biosynthesis all N-linked glycoprotein are homogeneously glycosylated. This is a result of a precise lectin chaperone (reader) based proofreading mechanism in the ER, which discriminates between correctly folded and misfolded glycoproteins (Figure 4).[25] Here the oligosaccharide plays a key role in presenting each glycoprotein for scrutiny by the sophisticated biological checkpoint process, which is referred to as glycoprotein quality control.[26]
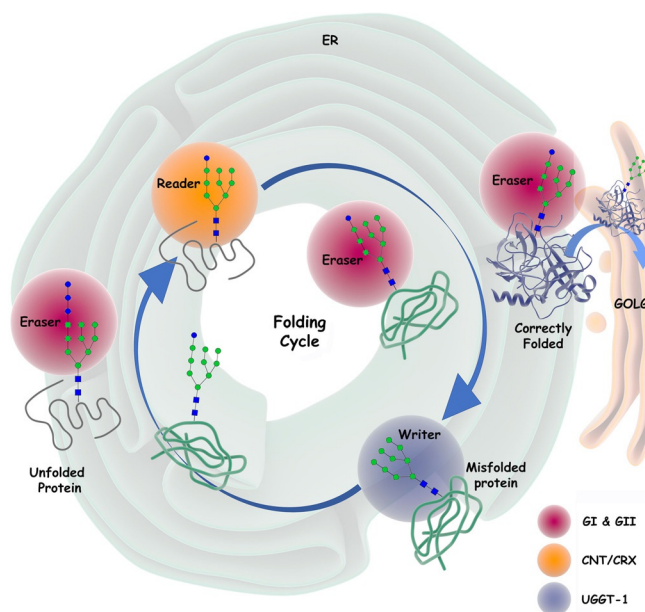

Figure 4. Carbohydrate writers, readers and erasers oversee the quality control of glycoprotein folding in the ER by modification of then-linked glycan high mannose oligosaccharide core structure.

This process ensures that only correctly folded glycoproteins are transported to the Golgi for further glycan processing in to mature glycoproteins. Unfolded and misfolded glycoproteins are retained in the ER for further folding attempts and are eventually degraded if the correctly folded status is not achieved.

The glycoprotein quality control system presents clear parallels to the read, write and erase processes of epigenetic regulation. In the first step of glycosylation, $Glc_3Man_9GlcNAc_2$ is transferred en bloc from an oligosaccharyl dolichol diphosphate to the nitrogen of an asparagine side chain in the nascent polypeptide chain by the writer oligosaccharyltransferase (OST).[27] Immediately after the $Glc_3Man_9GlcNAc_2$ is transferred, the eraser glucosidase-I[28] cleaves off the terminal glucose (Glc) residue, which is necessary to prevent the glycoprotein product rebinding to the OST. Subsequently, the eraser glucosidase-II[29] catalyses cleavage of a second glucose residue and the resulting monoglucosylated polypeptide is promptly sequestered by the calnexin (CNX)[30] and calreticulin (CRT)[31] lectin chaperone[32] readers. These chaperones prevent aggregation of the unfolded glycopolypeptide chains, and assist in their correct folding by presentation to the oxidoreductase ERp57, which is responsible for effecting correct disulfide bond formation.[33] Once the folded glycoprotein is released from the lectin chaperones readers, the eraser glucosidase-II removes the final glucose residue and the glycoprotein undergoes inspection by the UDP-glucose:glycoprotein glycosyltransferase (UGGT)[29] (Figure 4).

If the correct glycoprotein folding is not accomplished, UGGT serves as a writer and re-glucosylates the misfolded glycoprotein in preparation for recycling to the chaperones/ERp57 machinery[34]—the so-called calnexin/calreticulin cycle (Figure 4).[35] Following repeated failed folding attempts, the glycoprotein is degraded by the endoplasmic reticulum associated degradation system (ERAD).[36] If correct folding is achieved, the glycoprotein is transported into the Golgi apparatus for further processing of the glycan to provide the mature glycoprotein.

Conclusion

It is widely recognised that carbohydrates play important roles in biological molecular recognition, and have a profound impact on human health and medicine. Nonetheless, there is merit in simplifying the language of glycoscience to make it more accessible to the uninitiated. In turn, this might facilitate a focus on the principles and implications of glycosylation in biology, rather than risking drowning in the detail of structural complexity. The notion of accessible vocabulary in glycoscience is not new: it was already evident in Hood, Huang and Dreyer's 1977[37] description of differentiation antigens as cell-surface "area codes"; and the potential of cell-surface carbohydrates, lectins, enzymes and carbohydrate-binding antibodies in Feizi's 1981[38] "cellular addresses", "postmen, policeman and traffic signs" "involved in the obedient interpretation of area codes". Similar thoughts were explored in Brandley and Schnaar's 1986[39] "potential carbohydrate "language" involved in inter-

cellular interactions", while Hakomori's 2002[40] "glycosynapse"—microdomains of glycolipids—seeks to draw parallels to the "immune synapse" assembly that contributes to cell adhesion and signalling. As highlighted in the cross-disciplinary article by Bertozzi and Kiessling in 2001,[41] "chemical tools have proven indispensable for studies in glycobiology". Perhaps it is time to revisit the terminology of glycoscience, to make interdisciplinary communication more straightforward and to support marketing and engagement beyond the immediate field. Reference to lectin readers, glycosyltransferase writers, and glycosyl hydrolase erasers could therefore be worth wider (re)consideration.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

Keywords: carbohydrates · epigenetics · glycoscience vocabulary · glycosyl hydrolase · glycosylation · glycosyltransferase

[1] T. W. Rademacher, R. B. Parekh, R. A. Dwek, *Annu. Rev. Biochem.* **1988**, *57*, 785–838; R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683–720.

[2] A. Varki, *Glycobiology* **1993**, *3*, 97–130; A. Varki, *Glycobiology* **2017**, *27*, 3–49.

[3] A. Varki quoted in S. Borman, *Chem. Eng. News* **2012**, *90*, 28–29; See also: National Research Council (US) Committee on Assessing the Importance and Impact of Glycomics and Glycosciences; *Transforming Glycoscience: A Roadmap for the Future*, National Academies Press, Washington (DC), **2012**.

[4] H.-J. Gabius, H.-C. Siebert, S. André, J. Jiménez-Barbero, H. Rüdiger, *ChemBioChem* **2004**, *5*, 740–764; *The Sugar Code: Fundamentals of Glycosciences* (Ed.: H.-J. Gabius), Wiley-VCH, Weinheim, **2009**.

[5] M. Mendelson, M. Balasegaram, T. Jinks, C. Pulcini, M. Sharland, *Nature* **2017**, *545*, 23–25.

[6] J. E. Turnbull, R. A. Field, *Nat. Chem. Biol.* **2007**, *3*, 74–77.

[7] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic Acids Res.* **2009**, *37*, D233–D238; V. Lombard, H. G.

Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490–D495.

[8] S. V. Bennun, D. Baycin Hizal, K. Heffner, O. Can, H. Zhang, M. J. Betenbaugh, *J. Mol. Biol.* **2016**, *428*, 3337–3352.

[9] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, *Essentials of Glycobiology*, 3rd ed., Cold Spring Harbor Laboratory Press, New York, **2015–2017**.

[10] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, et al., *Glycobiology*, **2015**, *25*, 1323–1324. See also: S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, T. Lütteke, N. O'Boyle, N. H. Packer, P. Stanley, P. Toukach, A. Varki, R. J. Woods, The SNFG Discussion, *Glycobiology*, **2019**, *29*, 620–624.

[11] W. S. York, S. Agravat, K. F. Aoki-Kinoshita, R. McBride, M. P. Campbell, C. E. Costello, A. Dell, T. Feizi, S. M. Haslam, N. Karlsson, et al., *Glycobiology* **2014**, *24*, 402–406.

[12] M. Tiemeyer, K. Aoki, J. Paulson, R. D. Cummings, W. S. York, N. G. Karlsson, F. Lisacek, N. H. Packer, M. P. Campbell, N. P. Aoki, A. Fujita, M. Matsubara, D. Shinmachi, S. Tsuchiya, I. Yamada, M. Pierce, R. Ranzinger, H. Narimatsu, K. F. Aoki-Kinoshita, *Glycobiology* **2017**, *27*, 915–919.

[13] H.-J. Gabius, *BioSystems* **2018**, *164*, 102–111.

[14] J. Martin, *Managing the Database Environment*, Prentice Hall, New Jersey, **1983**, p. 381.

[15] M. Heller, "*REST and CRUD: The Impedance Mismatch*", InfoWorld, **2007**, https://www.infoworld.com/article/2640739/rest-and-crud–the-impedance-mismatch.html.

[16] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41–45.

[17] Reviewed in: a) K. J. Falkenberg, R. W. Johnstone, *Nat. Rev. Drug Discovery* **2014**, *13*, 673–691; b) C. D. Allis, T. Jenuwein, *Nat. Rev. Genetics* **2016**, *17*, 487–500.

[18] J. A. Hanover, M. W. Krause, D. C. Love, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 312–321; S. A. Yuzwa, D. J. Vocadlo, *Chem. Soc. Rev.* **2014**, *43*, 6839–6858.

[19] G. D. Hart, *J. Biol. Chem.* **2019**, *294*, 2211–2231.

[20] Little is currently known about readers for O-GlcNAc, although proteins involved in signalling have recently been implicated. X. Yang, K. Qian, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 452–465; C. A. Toleman, M. A. Schumacher, S. H. Yu, W. Zeng, N. J. Cox, T. J. Smith, E. J. Soderblom, A. M.

Wands, J. J. Kohler, M. Boyce, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 5956–5961.

[21] M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593–1608; J. Arnaud, A. Audrey, A. Imberty, *Chem. Soc. Rev.* **2013**, *42*, 4798–4813.

[22] A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, *Biochem. J.* **2004**, *382*, 769–781; C. Carvalho, N. N. Phan, Y. F. Che, P. J. Reilly, *Biopolymers* **2015**, *103*, 203–2145.

[23] X. Zhang, Y. Wang, *J. Mol. Biol.* **2016**, *428*, 3183–3193.

[24] M. E. Losfeld, E. Scibona, C.-W. Lin, T. K. Villiger, R. Gauss, M. Morbidelli, M. Aebi, *FASEB J.* **2017**, *31*, 4623–4635.

[25] A. Tannous, G. B. Pisoni, D. N. Hebert, M. Molinari, *Sem. Cell Dev. Biol.* **2015**, *41*, 79–89.

[26] T. Anelli, R. Sitia, *EMBO J.* **2008**, *27*, 315–327.

[27] E. Mohorko, R. Glockshuber, M. Aebi, *J. Inherited Metab. Dis.* **2011**, *34*, 869–878.

[28] M. K. Barker, D. R. Rose, *J. Biol. Chem.* **2013**, *288*, 13563–13574.

[29] C. D. Alessio, J. J. Caramelo, A. Parodi, *Semin. Cell Dev. Biol.* **2010**, *21*, 491–499.

[30] J. D. Schrag, J. J. M. Bergeron, Y. Li, S. Borisova, M. Hahn, Y. D. Thomas, M. Cygler, *Mol. Cell* **2001**, *8*, 633–644.

[31] M. Michalak, J. Groenendyk, E. Szabo, L. I. Gold, M. Opas, *Biochem. J.* **2009**, *417*, 651–666.

[32] L. Halperin, J. Jung, M. Michalak, *IUBMB Life* **2014**, *66*, 318–326.

[33] A. Hettinghouse, R. Liu, C. J. Liu, *Pharmacol. Ther.* **2018**, *181*, 34–48.

[34] L. Ellgaard, E.-M. Frickel, *Cell Biochem. Biophys.* **2003**, *39*, 223–247.

[35] L. Lamriben, J. B. Graham, B. M. Adams, D. N. Hebert, *Traffic* **2016**, *17*, 308–326.

[36] R. Benyair, N. Ogen-Shtern, G. Z. Lederkremer, *Sem. Cell Dev. Biol.* **2015**, *41*, 99–109.

[37] L. Hood, H. V. Huang, W. J. Dreyer, *J. Supramol. Struct.* **1977**, *7*, 531–559.

[38] T. Feizi, *Trends Biochem. Sci.* **1981**, *6*, 333–335.

[39] B. K. Brandley, R. L. Schnaar, *J. Leukocyte Biol.* **1986**, *40*, 97–111.

[40] S.-i. Hakomori, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 225–232.

[41] C. R. Bertozzi, L. L. Kiessling, *Science* **2001**, *291*, 2357–2364.