Biochimica et Biophysica Acta 1850 (2015) 186–235

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Review

A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code

Dolores Solís [a,b,*], Nicolai V. Bovin [c,**], Anthony P. Davis [d,***], Jesús Jiménez-Barbero [e,****], Antonio Romero [e,*****], René Roy [f,******], Karel Smetana Jr. [g,*******], Hans-Joachim Gabius [h,********]

[a] Instituto de Química Física "Rocasolano", CSIC, Serrano 119, 28006 Madrid, Spain
[b] Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), 07110 Bunyola, Mallorca, Illes Baleares, Spain
[c] Shemyakin & Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul Miklukho-Maklaya 16/10, 117871 GSP-7, V-437, Moscow, Russian Federation
[d] School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK
[e] Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu, 9, 28040 Madrid, Spain
[f] Department of Chemistry, Université du Québec à Montréal, P.O. Box 8888, Succ. Centre-Ville, Montréal, Québec H3C 3P8, Canada
[g] Charles University, 1st Faculty of Medicine, Institute of Anatomy, U nemocnice 3, 128 00 Prague 2, Czech Republic
[h] Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 München, Germany

ARTICLE INFO

Article history:
Received 29 October 2013
Received in revised form 13 March 2014
Accepted 18 March 2014
Available online 28 March 2014

Keywords:
Conformer
Crystallography
Dendrimer
Glycan
Lectin
Neoglycoconjugate

ABSTRACT

Background: The most demanding challenge in research on molecular aspects within the flow of biological information is posed by the complex carbohydrates (glycan part of cellular glycoconjugates). How the 'message' encoded in carbohydrate 'letters' is 'read' and 'translated' can only be unraveled by interdisciplinary efforts.
Scope of review: This review provides a didactic step-by-step survey of the concept of the sugar code and the way strategic combination of experimental approaches characterizes structure–function relationships, with resources for teaching.
Major conclusions: The unsurpassed coding capacity of glycans is an ideal platform for generating a broad range of molecular 'messages'. Structural and functional analyses of complex carbohydrates have been made possible by advances in chemical synthesis, rendering production of oligosaccharides, glycoclusters and neoglycoconjugates possible. This availability facilitates to test the glycans as ligands for natural sugar receptors (lectins). Their interaction is a means to turn sugar-encoded information into cellular effects. Glycan/lectin structures and their spatial modes of presentation underlie the exquisite specificity of the endogenous lectins in counterreceptor selection, that is, to home in on certain cellular glycoproteins or glycolipids.
General significance: Understanding how sugar-encoded 'messages' are 'read' and 'translated' by lectins provides insights into fundamental mechanisms of life, with potential for medical applications.

1. Introduction

The alphabet of nucleotides is the basis for storing genetic information in the germ line, starting its downstream flow to effector molecules by transcription and ensuring replication. Its size is confined to four characters. Intuitively, adding letters to this alphabet would make more sequence combinations of nucleotides possible, hereby increasing the informational contents of oligomers. The downside of this at first sight attractive perspective likely is impairing copying fidelity. Summing up thorough theoretical investigations based on structural, energetic and information-theoretic studies, "a certain alphabet size (probably four) seems to be optimal as a compromise between stability and evolvability, between fidelity and catalytic efficiency, and between information density and error resistance" [1]. The 5′,3′-phosphodiester linkage between nucleotides gives the nucleic acid chain its direction, for copying and for translation. The genetic code, relating two sequence types (i.e. nucleic acids and proteins) in a linear (one-dimensional)

* Correspondence to: D. Solís, Instituto de Química Física "Rocasolano", CSIC, Serrano 119, 28006 Madrid, Spain. Tel.: +34 915619400; fax: +34 915642431.
** +7 4953307138; fax: +7 4953305592.
*** +44 1179546334; fax: +44 1179251295.
**** +34 918373112x4370; fax: +34 915360432
***** +34 918373112x4244; fax: +34 915360432.
****** +1 5149873000x2546; fax: +1 5149874054.
******* +420 224965873; fax: +420 224965770.
******** H.-J. Gabius, Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 München, Germany. Tel.: +49 8921802290; fax: +49 892180992290.
E-mail addresses: d.solis@iqfr.csic.es (D. Solís), bovin@carb.ibch.ru (N.V. Bovin), Anthony.Davis@bristol.ac.uk (A.P. Davis), jjbarbero@cib.csic.es (J. Jiménez-Barbero), romero@cib.csic.es (A. Romero), roy.rene@uqam.ca (R. Roy), ksmet@lf1.cuni.cz (K. Smetana Jr), gabius@tiph.vetmed.uni-muenchen.de (H.-J. Gabius).

http://dx.doi.org/10.1016/j.bbagen.2014.03.016
0304-4165/

manner, underlies the conversion of triplet coding to proteins, to sequences which are arranged in bioactive shapes [2]. Considering the high degree of flexibility of peptides, proteins can, in part, be considered as a framework to let the embedded functionally relevant peptides adopt a distinct (bioactive) conformation and to enable conformational switches. Thus, it is essential that proteins reach their optimal conformation, suited for activity. To minimize detrimental deviations, every assistance possible to folding (or removal of misfolded waste) is welcome, and this is where the functionality of glycans already sets in. As follows, we introduce the concept of sugar coding by a case study of protein glycosylation, a ubiquitous process which has formerly been mostly interpreted to alter physicochemical properties of the protein.

N-glycosylation (the formation of an *N*-glycosidic linkage between the amide nitrogen of asparagine and reducing-end *N*-acetylglucosamine (β-Glc*N*Ac) of a glycan) is a common process in the three urkingdoms of life (eukaryotes and prokaryotes divided into archae- and eubacteria). It is mostly started co-translationally by the transfer of a lipid-linked 14mer oligosaccharide (i.e. GlcNAc_2 Man_9 Glc_3, with Man for mannose and Glc for glucose) to the protein's acceptor sequence (N-X-S/T-X; X≠P), in eukaryotes at the entry side of the endoplasmic reticulum (ER) [3–13]. Of note, at this stage all glycoproteins share the same sugar part. As a consequence of the glycan addition when leaving the cytoplasm, eukaryotic proteins on the lumenal side of the ER, in the Golgi regions, transport vesicles or lysosomes and on the plasma membrane as well as in the extracellular space carry this modification, and at high frequency in spatially accessible β-turns as examining mature glycoproteins delineated [6,14–18]. The thorough profiling also uncovered that, despite the phylogenetically ubiquitous occurrence, a fine-tuning of this type of protein modification has developed. In interspecies comparison, presence of potential *N*-glycosylation sites can differ to a notable degree so that, focusing on an endothelial intercellular adhesion molecule (ICAM-1), "examining the role of *N*-glycosylation in murine ICAM-1 and extrapolating the results to the human protein would seem imprudent" [19]. Turning back to the just mentioned possibility for co-translational additions providing assistance in folding, it would mean missing a splendid opportunity if the *N*-glycans were not recruited to this task. Fittingly, they are, and – due to the common origin – the sequence regions both in the core, that is in direct vicinity to the protein surface [20–22], and in outer branches [6,23,24] are at this stage of *N*-glycan functionality strictly constant at every site of glycosylation, a factor favoring a general role. Later, branch ends are subject to the same route of processing by stepwise truncations of the preformed glycan block [5,6,25–27]. As a consequence of the rather bulky nature of the added glycan and its structural uniformity, every *N*-glycan is capable of aiding the folding process by directly interacting with the protein via its core and by facilitating recognition of molecular chaperones via branch ends [23,24,28,29].

At this stage, Glc, which is not a part of mature glycoproteins, plays a prominent role in quality control, and at every stage of the stepwise removal of the three Glc moieties from the *N*-glycan. They are presented spatially accessible on a branch end. After the first of the three Glc moieties (in α1,3-linkage) had been cleaved from the *N*-glycan right after protein *N*-glycosylation by oligosaccharyltransferase ($t_{1/2}$ < 2 min) (in eukaryotes a multi-subunit enzyme, in prokaryotes a single-subunit protein [30]), this done so rapidly to preclude reversibility of the transfer by hydrolytic cleavage of glycan from the protein by this enzyme, the diglucosylated glycan becomes a ligand for malectin, an ER-resident protein [31] (please also see Table 1, comment on [368] in entry 20). Combined with ribophorin I and its capacity to interact with misfolded proteins, this complex may retain dysfunctional products in the ER, to start the process of quality control via glycan recognition of *N*-glycans [32]. The next step of trimming of the glycan chain (now at the Glcα1,2Glc linkage) makes the monoglucosylated contact point to two molecular chaperones (calnexin and calreticulin) available, and taking away the third Glc moiety (cleavage of Glcα1,2Man linkage) abrogates their binding. A sensor surveying conformations of glycoproteins in the ER (a bifunctional protein, which will restore presence of this Glc residue crucial for chaperone contact by its enzymatic UDP-Glc:glycoprotein transferase activity) will ensure signal presence for the association with these chaperones based on Glc recognition by reglucosylation if necessary [6,33–36]. Glucosylation of *N*-glycans thus is a transient phenomenon, an investment to make *N*-glycosylation an irreversible process and to help in ER-based quality control via chaperones.

As to the core, it is engaged in interaction with the protein (please see above). Its loss of contact to the protein is even later a sign for unfolding, 'read' by a receptor targeting the innermost GlcNAc_2 structure, and therefore a diagnostic signal for the need to direct these dysfunctional glycoproteins to proteasomal degradation by ubiquitinylation [37,38]. The indicated importance of the *N*-glycan implies the potential of defects in glycan assembly/processing to cause diseases. Indeed, careful analysis of *N*-glycan profiles in serum samples of patients using isoelectric focusing (started on the glycoprotein transferrin by the Swedish neurologist H. Stibler as diagnostic test for alcohol abuse in 1976) led to the discovery of anomalies (false-positives in the highly specific and sensitive assay to confirm abstinence), as class referred to as congenital diseases of glycosylation (CDG) [39–41]. For example, the absence of the α-glucosidase-I initiating the trimming cascade causes lethality in humans. The thorough delineation of causes for these diseases thus teaches intriguing lessons on functional aspects of *N*-glycans, reveals differences in comparison to animal models and allows therapeutic options to be devised [40,42–44] (for information on animal models, please see Table 1, entry 37).

Following this involvement in folding and quality control, *N*-glycans then have a second phase in their 'social' life. They undergo a substantial remodeling in the Golgi regions to prepare (equip) them for new functions (for details, please see Section 8). The stepwise size reduction to the pentasaccharide stage (Man_3 GlcNAc_2) and subsequent neosynthesis open the way to the known complexity of the cellular *N*-glycome, in terms of branching, the type and length of the branches, the terminal structures and the patterns of their substitution [3,5,6,45,46]. A fingerprint-like profile thereby takes the place of the initial sequence identity, an unmistakable hint to more than a random variation [47]. Based on total cellular glycomics, a high degree of glycome dependence on the cell type was found, prompting these authors to describe glycomes to have "utility as unique cellular descriptors" [48]. The same applies to the glycan chains of glycolipids, a key chemical platform of glycan presentation especially in the nervous system [49,50], and of proteoglycans [51,52]. Using glycan-binding laboratory tools, glycophenotyping of cells and tissues can conveniently be accomplished (for respective information, please see Table 1, entries 32 and 33).

Beyond such descriptive aspects, already structurally rather small changes can make their presence felt for protein properties such as (enzyme or receptor) activity [53]. Status of oligomerization or stability is other affected parameter, as exemplarily shown for a mammalian growth factor [54–57]. Considering that *N*-glycosylation (with the Asn-Glc*N*Ac linkage in β-configuration first described in 1961 [58]) is just one from at least 41 bonds between carbohydrates (13 different sugars known to be a part of the conjugate) and amino acids (eight different acceptor sites detected; hydroxylysine and hydroxyproline, acceptors of Glcα1,2Galβ1,*O* in collagen or arabinogalactan chains in plant glycoproteins [59–61], are produced by distinct hydroxylases, three in man, at least two in plants [62,63]) within proteins [64,65] (for further information on mucin-type *O*-glycosylation and mucins, the other frequently occurring type of protein glycosylation, please see [66]), carbohydrates (attached to proteins and lipids) have so broad a distribution profile that postulating a truly fundamental role is reasonable. Having started off with a look at the strictly ordered processes during *N*-glycosylation and then broadening the scope to sugars as part of cellular glycoconjugates, this assumption prompts to examine more closely the properties of carbohydrates, in comparison to nucleotides and amino acids.