International Journal of Biological Macromolecules 84 (2016) 372–379

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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

Review

Phylogeny, structure, function, biosynthesis and evolution of sulfated galactose-containing glycans

Vitor H. Pomin[a,b,*]

[a] Program of Glycobiology, Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro, Rio de Janeiro, RJ 21941-913, Brazil
[b] University Hospital Clementino Fraga Filho, Federal University of Rio de Janeiro, Rio de Janeiro, RJ 21941-913, Brazil

ARTICLE INFO

Article history:
Received 15 November 2015
Received in revised form
30 November 2015
Accepted 11 December 2015
Available online 19 December 2015

Keywords:
Galactose
Galactosyltransferase
Glycobiology
Glycosylation
Molecular evolution
Sulfated galactans
Sulfation patterns
Sulfotransferase

ABSTRACT

Glycans are ubiquitous components of all organisms. The specificity of glycan structures works in molecular recognition in multiple biological processes especially cell–cell and cell–matrix signaling events. These events are mostly driven by functional proteins whose activities are ultimately regulated by interactions with carbohydrate moieties of cell surface glycoconjugates. Galactose is a common composing monosaccharide in glycoconjugates. Sulfation at certain positions of the galactose residues does not only increase affinity for some binding proteins but also makes the structures of the controlling glycans more specific to molecular interactions. Here the phylogenetic distribution of glycans containing the sulfated galactose unit is examined across numerous multicellular organisms. Analysis includes autotrophs and heterotrophs from both terrestrial and marine environments. Information exists more regarding the marine species. Although future investigations in molecular biology must be still performed in order to assure certain hypotheses, empirical evidences based on structural biology of the sulfated galactose-containing glycans among different species particularly their backbone and sulfation patterns clearly indicate great specificity in terms of glycosyltransferase and sulfotransferase activity. This set of information suggests that evolution has shaped the biosynthetic machinery of these glycans somewhat related to their potential functions in the organisms.

Contents

1. Introduction

Carbohydrates, proteins, nucleic acids and lipids are ubiquitous components of all living beings and due to this reason can be considered the essential biological macromolecules of life [1]. Each macromolecule has its own physical, biological and chemical properties, but carbohydrate is by far the one that shows the greatest variations in terms of structure. This is a consequence from the existence of multiple monosaccharides, glycosidic linkage types, anomericity, positions and levels in post-polymerization chemical modifications such as acetylation, phosphorylation and sulfation [1]. These structural features can lead to infinite number of polysaccharides assuming all possible chemical permutations. However, not all polysaccharide structures occur naturally. This is because carbohydrates, unlike proteins and nucleic acids, are not

* Correspondence to: R. Prof. Rodolpho Paulo Rocco, 255, HUCFF 4A01, Ilha do Fundão, Rio de Janeiro, RJ 21941-913, Brazil. Fax: +55 21 3938 2090.
E-mail addresses: pominvh@bioqmed.ufrj.br, vhpomin@gmail.com

http://dx.doi.org/10.1016/j.ijbiomac.2015.12.035
0141-8130/

synthesized through a template-driven mechanism, but rather a direct product from the levels of expression, activity and substrate-specificity of related anabolic enzymes together with the proper availability of substrates and sugar-donors for biosynthesis [2]. Because not all enzymes, substrates and sugar-donors exist at the same time at the same system, not all polysaccharide structures are found in nature.

Among numerous monosaccharide types, glucose (Glc) is undoubtedly the most famous representative. This is due to a series of factors. (i) Glc is the principal precursor for several organic molecules and for the most other monosaccharides. (ii) Glc is the central molecule in energy-related metabolism (bioenergetics). (iii) Glc is involved as structural component of many biologically active glycoconjugates in the organisms. And (iv), Glc is the center molecule in several pathophysiological events. The best known pathology example is diabetes mellitus. Conversely, other monosaccharide types such as galactose (Gal) are also related to such events. Gal is the C4 epimer of Glc. In Gal the OH at C4 position points axially while in Glc it points equatorially (Fig. 1). This slight structural modification is enough to change considerably the physicochemical and biological properties of these two hexoses.

Carbohydrates are involved not only in energy-related metabolism (bioenergetics), but also in vital biological processes especially cell–cell and cell–extracellular matrix signaling events [3]. These signaling events are essentially triggered by functional proteins [4]. In order to achieve their proper roles in the different systems, the signaling proteins must be, in most cases, controlled by molecular interactions with carbohydrate moieties of glycoconjugates such as *N*- or *O*-linked glycoproteins, glycolipids and proteoglycans (PGs) displayed at the cell surface or embedded in the extracellular matrix (ECM) [3,4]. The carbohydrate-protein interactions underlying the signaling processes are very structure-specific [4]. Proteins can bind and interact with just certain epitopes (interacting sequences) of the carbohydrate moieties. These binding motifs are frequently constituted of just few monosaccharides in length, usually no longer than a decasaccharide. The specificity of carbohydrate recognition in interactions with protein depends on the chemical information encoded in the binding motif of the glycan chain. *Sugar code* has been conventionally named to designate this structural specificity in bioactive glycan sequences [5–7].

Gal together with Glc as well as many other monosaccharide types is a common composing unit of the biologically active sequences of glycocojugates in multicellular organisms. For instance, in *N*-linked glycoproteins, Gal units are generally located outside the common core structure and may become rich in complex structures [8]. On the other hand, Gal usually constitutes the principal unit in the backbone of *O*-linked sugar chains apart from the peripheral region [8]. One structural modification that improves affinity for protein interactions, especially cationic proteins, and also enhances structural specificity in molecular recognition is sulfation. All monosaccharide building blocks are liable of this chemical modification, depending of course on the type of sulfotransferase expressed and active in the system together with the proper substrate available for reaction. Sugar-donors in enough amounts must be also considered in this biosynthetic system. Among all possible monosaccharides eligible for sulfation, Gal seems to be the unit distributed in a broader range of species and more frequently modified by such reaction [8]. Moreover, sulfated Gal is likely the commonest sulfated monosaccharide type found in the various glycoconjugates. Although *O*-sulfation can occur at positions C2, C3, C4 or C6 of Gal, sulfation happens more often at C3 and C6 positions [8]. Sulfation position is biologically important because only at certain ring positions it will be able to improve the protein binding [9]. In fact, sulfation at certain positions can conversely compromise the formation of the carbohydrate–protein complex [9]. Of course the resultant intermolecular complex depends intimately and ultimately on the protein type.

For example, P-selectin and laminin are important functional protein examples whose activities must be controlled by molecular interactions with cell sulfated Gal-containing motifs on cell surface carbohydrates. P-selectin, a key protein involved in cell migration during inflammation and tumor processes, recognizes and binds specifically to cell membrane sulfated cerebrosides (sulfatide also known as 3-*O*-sulfogalactosylceramide) that are concentrated in lipid rafts. Metabolic inhibition of sulfation on cerebrosides can significantly abrogate the P-selectin binding [10]. Laminin, a key ECM protein of the basement membrane and responsible for cell differentiation, migration and adhesion, must anchor to sulfatide in order to initiate basement membrane reassembly [11]. Laminin fails to trigger this process when sulfation is absence in the ligand structure. Laminin–sulfatide interactions are necessary to create functional membrane microdomains capable to induce myelination [12]. Disruption of laminin–sulfatide interactions impedes oligodendrocyte differentiation and myelin-like membrane formation [8]. The two examples of proteins above-described are known to recognize and bind the sulfated Gal-containing region of glycolipids in mammalian-derived cells. But a great number of other proteins from multiple other systems are also necessary to bind to the sulfated Gal-containing glycan motifs of cell-coating glycoconjugates in order to present their proper functions to the systems.

Since sulfated Gal units are extremely relevant in terms of biological activity and found in glycoconjugates of various species, this paper aims therefore at analyzing the phylogenetic distribution of the glycan structures containing this particular functional unit. Analysis is made across different multicellular organisms. Representative autotroph and heterotroph species from both terrestrial and marine habitats either closely related or evolutionary distant were all considered in the investigation. Focus will be given not only on the sulfation patterns, but also on the constitution of the backbones (additional composing monosaccharide types adjacent to the sulfated Gal units), anomericity, enantiomericity as well as position and types of glycosidic bonds of the composing residues in the functional motifs. This structural observation although empirical will help to raise some correlations between structure and phylogeny. Such analysis helps to understand the biological functions of the studied glycans in the different species as well as their biosynthetic aspects, particularly regarding the activities of galactosyltransferases and specific sulfotransferases. Data concerning the marine organisms exist more than those about terrestrial species. This is a natural consequence from the fact that marine organisms such as algae and invertebrates are capable to synthesize polysaccharides composed exclusively or mostly of sulfated Gal units. These macromolecules are widely known as sulfated galactans (SGs). Carrageenans and agarans are the famous examples of SGs.

2. General aspects regarding structure, function and phylogeny

2.1. Terrestrial heterotrophs

The best known and likely the most abundant sulfated Gal-containing glycan in terrestrial animals is keratan sulfate (KS). KS is a glycosaminoglycan (GAG) composed of disaccharide building blocks of alternating 4-linked *N*-acetyl β-D-glucosamine (GlcNAc) and 3-linked β-D-Gal units [13–15]. Sulfation can occur at both units but exclusively at the C6 positions and more often at the GlcNAc unit. KS chains are generally found structurally attached to a protein core forming the KS PGs [16]. KS chains can be either *N*-linked to asparagine (Asn) residues (named as KS-I) or