Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Deciphering functional glycosaminoglycan motifs in development

Robert A Townley[1] and Hannes E Bülow[2,3]

Glycosaminoglycans (GAGs) such as heparan sulfate, chondroitin/dermatan sulfate, and keratan sulfate are linear glycans, which when attached to protein backbones form proteoglycans. GAGs are essential components of the extracellular space in metazoans. Extensive modifications of the glycans such as sulfation, deacetylation and epimerization create structural GAG motifs. These motifs regulate protein–protein interactions and are thereby repsonsible for many of the essential functions of GAGs. This review focusses on recent genetic approaches to characterize GAG motifs and their function in defined signaling pathways during development. We discuss a coding approach for GAGs that would enable computational analyses of GAG sequences such as alignments and the computation of position weight matrices to describe GAG motifs.

Addresses
[1] Department of Biological Sciences, Columbia University, New York, NY 10027, United States
[2] Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, United States
[3] Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, United States

Corresponding author: Bülow, Hannes E (hannes.buelow@einstein.yu.edu)

Current Opinion in Structural Biology 2018, **50**:144–154

This review comes from a themed issue on **Carbohydrates**

Edited by **Hugues Lortat-Jacob** and **Ralf P Richter**

https://doi.org/10.1016/j.sbi.2018.03.011

Introduction

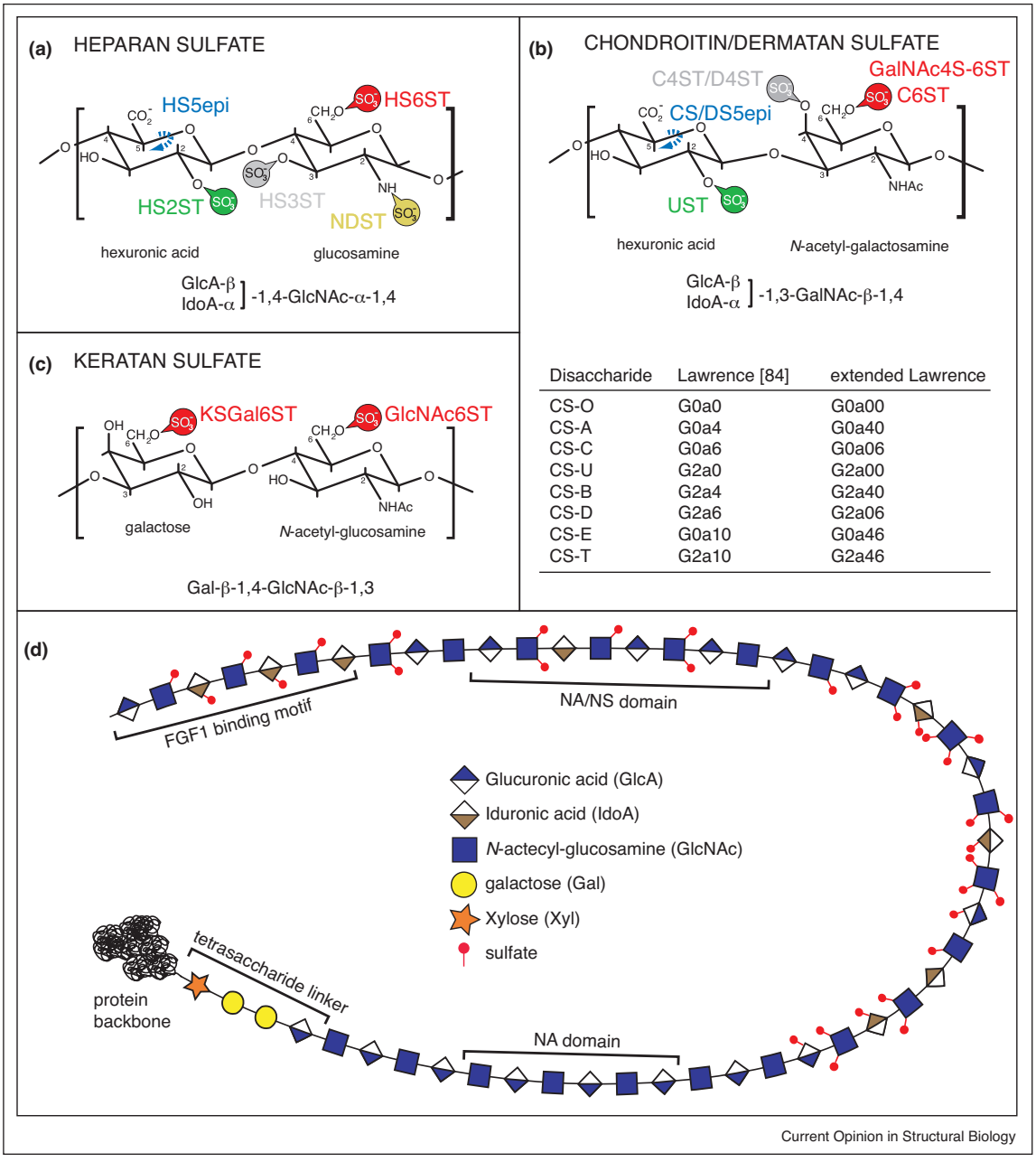
Glycosaminoglycans (GAGs) are linear glycans of repeating disaccharides and are ubiquitous components of cellular membranes and the extracellular space in metazoans. They play important biological roles in many aspects of development and physiology of multicellular organisms [1–8]. GAGs are classified based on the type of disaccharide repeats into heparan sulfates (HS, hexuronic acid/*N*-acetyl-glucosamine), chondroitin/dermatan sulfates (CS/DS, hexuronic acid/*N*-acetyl-galactosamine), and keratan sulfate (KS, galactose/*N*-acetyl-glucosamine) (Figure 1a–c). Hyaluronic acid (HA) comprises glucuronic-acid/*N*-acetyl-glucosamine disaccharides (although with different linkages), yet is not attached to a protein backbone. All GAGs (except HA) show diverse and molecularly complex modification patterns. These modification patterns include epimerizations, sulfations and deacetylations of select positions along the glycan chains, which create motifs that can function as protein binding sites (Figure 1d). Modification patterns are introduced in the Golgi, and not based on a template. Consequently, GAG structures display substantial heterogeneity within tissues and, an enormous molecular diversity across different tissues and over developmental time.

Determining direct structure–function relationships of GAG motifs *in vivo* remains challenging, because no tools exist that allow either the manipulation or removal of defined GAG motifs *in vivo* or, the routine characterization of GAG motifs *in vitro*. Progress has been made in the biochemical characterization of some GAG motifs, although these approaches remain difficult and time consuming. Similarly, the function of some more or less defined GAG sequences *in vitro* has been described using cell culture experiments. Here, we will review recent progress to establish characteristics of GAG sequences and their functions in defined signaling pathways during development. We will focus on HS, where most is known about the structural characteristics of motifs that mediate function *in vivo*. In a second part, we will briefly review established concepts of functional CS/DS motifs, although some of this data comes from *in vitro* experiments. We will not review KS, for which very little is known about functional domains during development, or HA, which is unmodified and thus does not contain motifs. In a final section we will discuss a novel coding approach that can be used to (1) depict and characterize functional GAG motifs, (2) describe the range of sequences bound by a protein and (3) capture the variation that is inherent in all GAG preparations from natural sources.

HS motifs in development

HS motifs can range significantly in length from a few disaccharides to dodecamers. Their binding to their cognate proteins can be broadly categorized as occurring via at least three different mechanisms [9,10]. The first mechanism is dominated by electrostatic interactions and is mediated primarily by negatively charged groups (carboxyl and sulfate groups) on the glycans and basic amino-acid residues in the protein ligands. Second, HS sequences can bind to

Figure 1



Disaccharide	Lawrence [84]	extended Lawrence
CS-O	G0a0	G0a00
CS-A	G0a4	G0a40
CS-C	G0a6	G0a06
CS-U	G2a0	G2a00
CS-B	G2a4	G2a40
CS-D	G2a6	G2a06
CS-E	G0a10	G0a46
CS-T	G2a10	G2a46

General structures of GAGs.
(a) Schematic of a heparan sulfate disaccharide with all possible modifications and the respective enzymatic activities required to introduce them are indicated: NDST (*N*-decacetylase-sulfotransferase), GLCE (*C5*-glucuronyl-epimerase), HS2ST (HS-2-*O*-sulfotransferase), HS3ST (HS-3-*O*-sulfotransferases) and HS6ST (HS-6-*O*-sulfotransferase).
(b) Schematic of a chondroitin/dermatan sulfate disaccharide with all possible modifications indicated. Dermatan sulfate is characterized by the presence of iduronic acid. The respective enzymatic activities required to introduce modifications are indicated: CS/DS5epi (*C5*-glucuronyl-epimerase), UST (uronyl-2-*O*-sulfotransferase), C4ST/D4ST (chondroitin/dermatan-4-*O*-sulfotransferase), C6ST (chondroitin-6-*O*-sulfotransferase), GalNAc4S-6ST (GalNAc4S-6-*O*-sulfotransferase). A table lists the structural characteristics of specific CS disaccharides using the coding system of Lawrence *et al.* [84] as well as our extended Lawrence system (cf. Figure 3).
(c) Schematic of a keratan sulfate disaccharide with all possible modifications and the respective enzymatic activities required to introduce them are indicated: KSGal6ST (KS galactose-6-*O*-sulfotransferase), GlcNAc6ST (KS *N*-actetyl-glucosamine-6-*O*-sulfotransferase).
(d) An HS chain attached to a protein is shown. Note the characteristic tetrasaccharide linker by which the chain is attached to a serine in a protein and, the structural segregation into domains with a high degree (NS), low degree (NS/NA) or no (NA) sulfation of the nitrogen on glucosamine. A putative FGF1 binding motif [97] is shown.