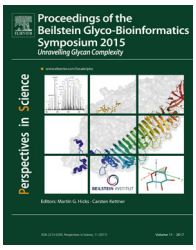
Perspectives in Science (2017) 11, 62—69

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc



Protein—glycosaminoglycan interaction networks: Focus on heparan sulfate☆

Sylvie Ricard-Blum

Institute for Molecular and Supramolecular Chemistry and Biochemistry, UMR 5246 CNRS, University Lyon 1-INSA Lyon, CPE, Raulin Building, University Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Received 4 January 2016; accepted 19 February 2016
Available online 20 October 2016

KEYWORDS
Extracellular matrix;
Heparan sulfate;
Interaction networks;
Interaction databases;
Surface plasmon resonance

Summary Sulfated glycosaminoglycans (GAGs) are complex polysaccharides, which are covalently bound to protein cores to form proteoglycans. They are mostly located at the cell surface and in the extracellular matrix (ECM) where they regulate numerous biological processes. The aim of our work is (i) to identify and characterize protein—GAG interactions occurring at the cell surface and in the ECM, (ii) to study the assembly of multimolecular complexes formed at the cell surface *via* protein—heparan sulfate interactions, (iii) to determine the roles of these complexes in the ECM maturation and assembly, which are initiated in the pericellular matrix, and in pathological situations such as angiogenesis and host—pathogen interactions, (iv) to build, contextualize and analyze the corresponding protein—heparan sulfate interaction networks to identify molecular connections between the physio-pathological processes mentioned above and to select protein—GAG complexes specifically formed in a pathological situation and which might be therapeutic targets.

Introduction

The extracellular matrix (ECM) is comprised of 274 core matrisome proteins and 747 matrisome-associated proteins in humans (Naba et al., 2012, http://matrisomeproject.mit.edu/). Most core matrisome proteins such as collagens, proteoglycans, elastin, fibronectin and laminins are multidomain proteins deposited under an insoluble form in the ECM and forming supramolecular assemblies. Matrisome-associated proteins include ECM regulators (degrading and cross-linking enzymes), secreted

E-mail address: sylvie.ricard-blum@univ-lyon1.fr

http://dx.doi.org/10.1016/j.pisc.2016.10.004
2213-0209/

growth factors and cytokines (*e.g.*, TGF-β) and ECM-affiliated proteins such as galectins. In addition to proteins, the ECM contains linear, anionic, sulfated polysaccharides, the glycosaminoglycans (GAGs), which are covalently linked to proteins to form proteoglycans. The structural and functional roles of ECM are mediated by networks of protein—protein and protein—GAG interactions, which are constantly rewired according to the physiological and pathological contexts. We focus here on protein—GAG interactions and mostly on protein—heparin/heparan sulfate interactions involved in ECM assembly, angiogenesis and infectious diseases. Our goal is to decipher the molecular mechanisms underlying the above physiopathological processes and to identify protein—GAG complexes specific of a particular disease as new potential therapeutic targets.

There are six glycosaminoglycans in mammals. Five of them (heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate and keratan sulfate) are sulfated and are covalently linked to proteins to form proteoglycans, whereas hyaluronan is not sulfated and forms aggregates of proteoglycans by interacting non covalently with individual proteoglycan molecules through link proteins. Proteoglycans are a large and heterogeneous family of forty-three members (Iozzo and Schaefer, 2015). They consist in one or numerous glycosaminoglycan chains, which are covalently linked to a core protein. Proteoglycans differ in the nature of their protein core and in the nature and number of GAG chains covalently linked to the protein core. Aggrecan comprises up to one hundred GAG chains whereas decorin bears a single GAG chain. Several proteoglycans bear two types of GAG chains and the type, number and size of GAG chains may vary for a single proteoglycan depending on tissues (*e.g.*, kidney *versus* intestinal mucosa), cells and the biological context (*e.g.*, quiescent *versus* activated cells).

Heparin is widely used as a model of heparan sulfate for *in vitro* studies. Both GAGs differ in the extent of sulfation and uronic acid epimerization. Heparin is more sulfated (1.8/2.6 sulfate/hexosamine ratio) than heparan sulfate (0.8—1.8 sulfate/hexosamine ratio) and contains a higher amount of iduronic acid (70% *versus* 30—50% for heparan sulfate) (Esko et al., 2009). Heparan sulfate has several post-synthetic modifications (*N*-, 2-*O*, 3-*O*, 6-*O* sulfation, *N*-acetylation, glucuronic acid epimerization), which lead to forty-eight possible disaccharides. Twenty-three disaccharides have been identified so far in mammals for heparan sulfate (Bülow and Hobert, 2006) which provide a huge structural heterogeneity to an heparan sulfate chain containing up to 150 disaccharides. The sulfation is not homogeneous along the heparan sulfate chain. There are regions of low — or no — sulfation, called *N*-acetylated regions (NA), highly sulfated regions (NS) and regions of intermediate sulfation called NA/NS. The NS domains and the intermediate domains are the hypervariable regions that result in different functional characteristics for heparan sulfate from different cell types and tissues. In addition, a further post-synthetic modification of heparan sulfate occurs at the cell surface, where extracellular endosulfatases, called Sulfs, catalyze the specific removal of 6-*O* sulfate groups (Vivès et al., 2014).

The structural heterogeneity of heparan sulfate chains translates into functional diversity. Indeed, heparan sulfate chains fulfill a variety of biological roles. They are involved in ECM assembly, cell-ECM interactions, cell adhesion, migration and proliferation by interacting with receptors such as integrins (Faye et al., 2009a), growth factor sequestration within the ECM, development, angiogenesis, cancer, neurodegenerative diseases (interaction with the β-amyloid peptide), host-pathogen interactions and innate immunity. The goal of our work is to translate heparan sulfate—protein interactions into functions to decipher the molecular mechanisms of action of heparan sulfate, to identify new therapeutic targets and ultimately to design small molecules inhibiting specific protein—GAG interactions based on structural and molecular features both on protein and GAG involved in the formation of the complex. Targeting molecular interactions with small molecules is one of the approaches used in pharmacology (Jin et al., 2014).

A roadmap to build, contextualize and analyze protein—GAG interaction networks

We have designed a roadmap to build, contextualize, and analyze extracellular protein—protein and protein—GAG interaction networks (Fig. 1) including the following steps: (1) identification of new interactions by surface plasmon resonance (SPR) and SPR imaging (SPRi) using ECM protein and GAG arrays we have developed (Faye et al., 2009b; Fatoux-Ardore et al., 2014; Salza et al., 2014), (2) collection of further interaction data by manual curation of the literature, (3) storage of interaction data in MatrixDB, the interaction database we have developed (http://matrixdb.ibcp.fr/, Chautard et al., 2009, 2011; Launay et al., 2015; *cf.* below), (4) querying MatrixDB and other interaction databases to build comprehensive interaction networks of a molecule (GAG or protein), (5) contextualization of the network by adding kinetics and affinity of interactions calculated by SPR, quantitative proteomic data when available, Gene Ontology terms (http://geneontology.org/, Gene Ontology Consortium, 2015), annotations from the Reactome pathway database (http://www.reactome.org/, Croft et al., 2014) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/, Kanehisa et al., 2015), and expression data from UniGene (http://www.ncbi.nlm.nih.gov/unigene, Pontius et al., 2003). The integration of the above data into the networks allows the building in MatrixDB of networks specific of a tissue, a biological process, a molecular function, a pathway or a disease. The networks are visualized using the iNavigator of MatrixDB (Launay et al., 2015) and/or Cytoscape, ''an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data'' (http://www.cytoscape.org/, Su et al., 2014). Enrichment analyses restricted so far to the proteins of the networks are performed with Cytoscape apps and the Functional Enrichment analysis tool FunRich (http://www.funrich.org/, Pathan et al., 2015).

Most tools that are currently available to analyze molecular interaction networks and to perform enrichment analyses have been designed for protein—protein interaction networks and are thus useless for protein—GAG interaction networks or for networks comprising both protein—protein and protein—GAG interactions. However, an enrichment analysis tool based on the ChEBI (Chem-