Perspectives in Science (2016) xxx, xxx–xxx

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

**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/

Please cite this article in press as: Ricard-Blum, S., Protein–glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004

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-

Please cite this article in press as: Ricard-Blum, S., Protein—glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004

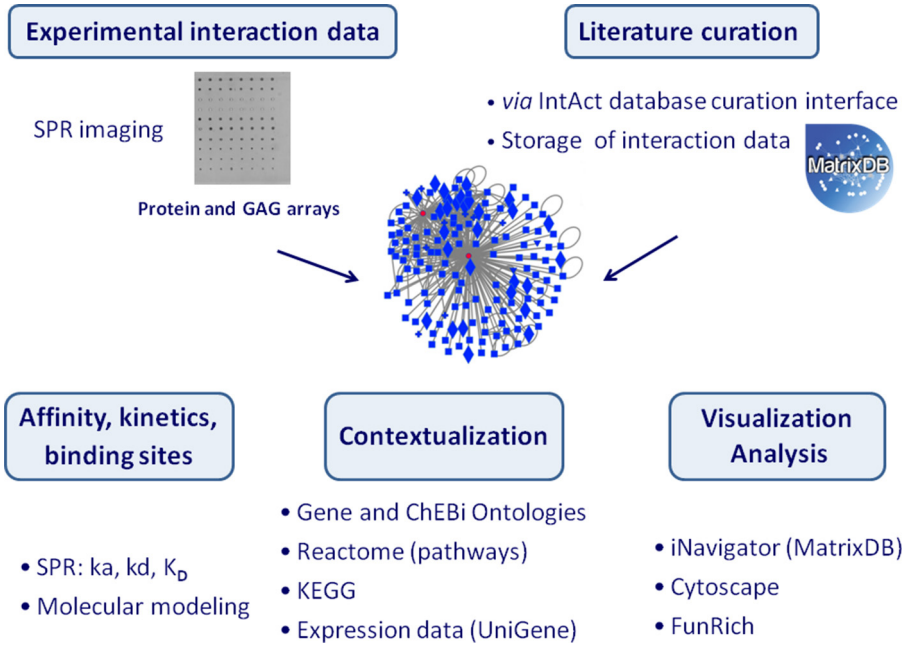


**Figure 1** Roadmap to build extracellular and pericellular protein—GAG interaction networks. The heparin—heparan sulfate interaction network comprises 150 and 52 partners (in blue) for heparin and heparan sulfate (in red) respectively. Diamonds: multimeric proteins (*e.g.*, collagens, laminins), Crosses: protein fragments (*e.g.*, endostatin, a fragment of collagen XVIII), Squares: proteins. The protein—GAG network has been built with the iNavigator of the MatrixDB database. (KEGG: Kyoto Encyclopedia of Genes and Genomes, SPR: surface Plasmon Resonance).

ical Entities of Biological Interest) ontology and called BiNChE has been recently developed for small molecules including glycosaminoglycan constitutive disaccharides (http://www.ebi.ac.uk/chebi/tools/binche/, Moreno et al., 2015). It has been integrated into the Gene Ontology resource (Hill et al., 2013) and we will use BiNChE for future analyses of protein—GAG interactomes.

## Identification of protein—GAG and pathogen—GAG interactions by SPR imaging (SPRi)

SPR binding assays are performed in real time and do not require labeling of the interactants. One interactant is immobilized on a gold surface and its potential partner is injected in solution at various flow rates over the immobilized interactant. The binding is detected by surface plasmon resonance, which is an optical phenomenon. SPRi binding assays are carried out in a Biacore Flexchip instrument to identify new protein—GAG interactions, whereas classical SPR experiments are performed in a Biacore T100 instrument to calculate kinetic parameters and affinity of newly identified protein—GAG interactions. Both Biacore systems we use are from GE Healthcare and belong to the Protein Science Facility of UMS 3444/US 8 (Lyon, France).

New protein—glycosaminoglycan interactions are identified by SPRi with protein and glycosaminoglycan arrays comprised of about one hundred molecules (proteins and GAGs) spotted in triplicate on Gold affinity chips (GE Healthcare). GAGs are spotted (1 mg/ml) either as bound to a protein core under a proteoglycan form when available or as free GAG chains from several species (heparan sulfate from bovine kidney and from intestinal mucosa) and/or of high- and low-molecular weight (heparin and hyaluronan). Proteoglycans and GAGs are physically adsorbed onto the bare gold surface without any chemical reaction. Purified proteins (Faye et al., 2009b; Salza et al., 2014) or intact pathogens (*Leishmania* parasites, Fatoux-Ardore et al., 2014) are injected at a flow rate of 300—500 μl/min (proteins) or 150 μl/min (pathogens) and recirculated over the spotted GAGs for 25 min (proteins) or 2 h (pathogens). Using this approach we have shown that 83.5 and 62.5% of the 24 strains of *Leishmania* parasites we have tested bind to heparin and heparan sulfate, respectively, whereas only 8.3% the 24 strains of *Leishmania* interact with chondroitin or dermatan sulfate (Fatoux-Ardore et al., 2014).The role of sulfate groups in the binding of proteins or pathogens to heparin is investigated by spotting selectively desulfated heparins (Iduron, UK) on the arrays. We have shown that 6-*O*-sulfate groups of heparin play a crucial role in its binding to *Leishmania* parasites (Fatoux-Ardore et al., 2014).

Several *Leishmania* strains interact with heparin but not with heparan sulfate. This might be due to differences in the structures of these GAGs. Heparan sulfate is less sulfated than heparin, contains clustered sulfate groups organized in highly sulfated domains, is more flexible and has a longer and more bent structure than heparin (Khan et al., 2013). Heparan sulfate may thus interact differently with pathogens or proteins compared with heparin. It is thus not correct to automatically infer that pathogens or proteins interacting with heparin also bind to heparan sulfate.

Please cite this article in press as: Ricard-Blum, S., Protein—glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004

## Manual curation of the literature: storage of protein—GAG interaction data in MatrixDB, a database focused on ECM interactions

To collect protein—GAG interaction data from the literature we have built a molecular interaction database called MatrixDB (http://matrixdb.ibcp.fr/, Chautard et al., 2009, 2011; Launay et al., 2015) focused on interactions involving at least one extracellular matrix protein and/or one glycosaminoglycan. MatrixDB was the first interaction database and is currently the only one to systematically curate protein—GAG interaction data. MatrixDB belongs to the International Molecular Exchange consortium (IMEx, http://www.imexconsortium.org/) and follows the IMEx curation rules (Orchard et al., 2012). MatrixDB performs curation of the papers published in Matrix Biology and of other papers related to ECM assembly, ECM maturation, angiogenesis and infectious diseases through the curation platform developed by the IntAct database and currently used by eleven interaction databases (http://www.ebi.ac.uk/intact/, Orchard et al., 2014). Protein—protein and protein—GAG interaction data curated by MatrixDB are freely available on our website, where they can be downloaded under a tabular format and/or as interaction networks built with the iNavigator we developed (Launay et al., 2015). Furthermore MatrixDB has implemented a Proteomics Standard Initiative Common Query InterfaCe (PSICQUIC) for computational access to molecular interaction data resources (Aranda et al., 2011). The query of a single database having implemented PSICQUIC returns the interaction data stored in other interaction databases having also implemented the PSICQUIC service.

One of the first issues to address for the curation of protein—GAG interactions was to select an identifier for glycosaminoglycans. In agreement with the IMEx curation rules we have selected the identifiers of the Chemical Entities of Biological Interest for glycosaminoglycans (ChEBI, http://www.ebi.ac.uk/chebi/init.do, Hastings et al., 2013) based on their constitutive disaccharides. This identifier does not take into account the structural diversity of the disaccharides of GAGs described above but neither PubChem identifier (Kim et al., 2016), nor the IUPAC International Chemical Identifier for chemical substances (InChI™, Heller et al., 2013, 2015) do it.

## *In vitro* and *in silico* localization of binding sites for protein—GAG interactions

The role of sulfate groups in the binding of GAGs to proteins is deciphered by SPRi as described above or by SPR using inhibition experiments. In this case, proteins are preincubated with selectively desulfated heparins before injection over immobilized heparin. We have shown that *N*-sulfate and 2-*O* sulfate groups of heparin strongly contribute to its binding to a fragment of collagen V (HepV, Ricard-Blum et al., 2006) and of collagen XVIII (endostatin, Ricard-Blum et al., 2004), respectively. Inhibition experiments are also carried out with heparin and heparan sulfate oligosaccharides (from tetrasaccharides to octadecasaccharides) to determine the minimum size of heparin or heparan sulfate required to promote an efficient binding to proteins. An octasaccharide of heparin is sufficient to promote the binding of the ectodomain of $\alpha5\beta1$ integrin (Faye et al., 2009a), whereas a decasaccharide and an hexadecasaccharide of heparan sulfate are required to bind a fragment of collagen V (Ricard-Blum et al., 2006) and collagen XVIII (Ricard-Blum et al., 2004), respectively.

We have used the AutoDock program and molecular dynamic simulations to localize the binding site(s) of heparin on the 3D structure of endostatin (Ricard-Blum et al., 2004) and of the ectodomain of the $\alpha v\beta3$ integrin (Ballut et al., 2013) by molecular modeling. We have identified two further arginine residues of endostatin ($Arg^{47}$ and $Arg^{66}$) contributing to its interaction with heparin (Ricard-Blum et al., 2004) and we have shown that heparan sulfate could bind to different sites on RGD-dependent and RGD-independent integrins (Ballut et al., 2013).

The chemical groups and features of GAGs contributing to their binding to proteins or pathogens (*e.g.*, sulfate groups, size, and oligosaccharide sequences) are currently stored in MatrixDB as free text, which restricts searches within the database and data exchanges with other interaction databases. A disaccharide code, based on a 4-character descriptor (uronic acid, presence and location of sulfate groups, hexosamine and the *N*-substituent, pattern of sulfation on the hexosamine) has been proposed for GAGs by Lawrence et al. (2008) but it has not been widely adopted by the GAG community. We will use in the future a unifying sequence format for carbohydrates (GlycoCT, Herget et al., 2008) and/or the Web3 unique representation of carbohydrate structures (WURCS, Tanaka et al., 2014), a linear notation which can be used as a Uniform Resource Identifier, to represent heparin/heparan sulfate features responsible for protein or pathogen binding. These data are of crucial importance because the pattern of GAG post-synthetic modifications (sulfation, acetylation, epimerization) and their conformation modulate their molecular recognition, their interaction repertoire and hence their biological functions. Conformational changes occurring in proteins upon their binding to heparin/heparan sulfate may be demonstrated by circular dichroism (CD). The comparison of CD spectra of a fragment of collagen V in absence and in presence of heparin or heparan sulfate has shown that no significant structural change is induced in this protein fragment upon GAG binding (Ricard-Blum et al., 2006).

## Kinetics and affinity of protein—heparin/heparan sulfate interactions

Kinetic parameters (*i.e.*, association and dissociation rates) allow the discrimination of transient and stable interactions, whereas the affinity, which characterizes to the strength of the interaction, discriminates strong and weak interactions and is useful to prioritize interactions within networks. Kinetics and affinity of interactions are calculated by SPR binding assays performed in a Biacore T100 system, which is more sensitive than the Biacore Flexchip used for screening purpose. Biotinylated heparin or heparan sulfate is captured onto streptavidin, which is covalently immobilized *via* its amino groups onto a sensor chip functionalized with a carboxymethyl dextran layer (Ricard-Blum et al., 2004, 2006;

Please cite this article in press as: Ricard-Blum, S., Protein—glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004

**Table 1** Lifetime of the complexes formed by ECM proteins or fragments, secreted proteins and integrin with heparin and heparan sulfate. The lifetime of the complexes was calculated as the reciprocal of the dissociation rate determined by SPR binding assays. Two lifetimes were calculated for integrin—GAG interactions involving the formation of a first complex, which rearranges in a more stable complex, and for angiopoietin-like-4, which comprises two binding sites for heparan sulfate. The first value corresponds to the fixation of heparan sulfate on one binding site and the second value to the binding of heparan sulfate on the second binding site.

| Protein | GAG | Interaction model | Complex lifetime (1/kd, min) | Reference |
|---|---|---|---|---|
| Procollagen C-Proteinase Enhancer-1 | Heparin<br>Heparan sulfate | 1:1 model<br>1:1 model | ~44<br>~37 | Weiss et al. (2010) |
| Endostatin | Heparin<br>Heparan sulfate | 1:1 model | ~4<br><1 | Ricard-Blum et al. (2004) |
| Transglutaminase-2 | Heparin | 1:1 model | 34 | Scarpellini et al. (2009) |
| α5β1 integrin | Heparin<br><br>Heparan sulfate | 2-state (conformational change) | <1<br>~66<br>~1<br>~31 | Faye et al. (2009a) |
| Angiopoietin-like 4 | Heparan sulfate | Bivalent analyte model (2 binding sites for heparan sulfate on the protein) | ~2<br>~73 | Chomel et al. (2009) |

Weiss et al., 2010). Proteins are then injected at several concentrations over captured GAGs. Each sensor chip has four flow cells, the first one being a control flow cell used to evaluate the non-specific binding of injected proteins to the sensor chip surface. Kinetics and affinity are calculated using predetermined interaction models available in the Biaevaluation software (GE Healthcare). The association rate reflects the molecular recognition of both partners, whereas the dissociation rate reflects the stability of the complex resulting from the interaction of both partners. The reciprocal of the dissociation rate (kd) is an evaluation of the complex lifetime and consequently of the duration of the biological effects triggered by the interaction. It allows the comparison of the lifetime of complexes formed *via* different molecular mechanisms, reported in Table 1 as the interaction model fitting the experimental interaction data. The complexes formed by heparin/heparan sulfate with cell surface-associated proteins have very different lifetimes, ranging from 50 s to 70 min (Table 1).

## Relationship between kinetics, affinity molecular and biological functions

We have built a dataset of 125 protein—heparin/heparan sulfate interactions for which affinity and kinetic parameters are available in order to determine if there is a relationship between the molecular functions of heparin-binding proteins, their localization, the biological processes they are involved in, and the affinity or kinetics of their interactions with heparin/heparan sulfate. The dataset was comprised of data stored in MatrixDB database and of data manually curated from the literature (Peysselon and Ricard-Blum, 2014). Lipoproteins, growth factors and cytokines bind to heparin with a higher affinity than proteins playing a role in ECM assembly and organization. Enzymes interact with heparin with a higher affinity than the other proteins and form more stable complexes with heparin than the other proteins. In addition, heparin-binding proteins regulating angiogenesis have on average a higher affinity for heparin than other proteins. The association rates were higher for proteins regulating angiogenesis than for other proteins and the complexes formed with heparin by proteins regulating angiogenesis dissociated faster than those formed by other proteins (Peysselon and Ricard-Blum, 2014).

## Perspectives

Further experimental approaches are needed to address the limitations of those currently used to characterize protein—heparan sulfate interactions and the corresponding interaction networks. The vast majority of the studies carried out so far to characterize protein—heparan sulfate interactions in physiopathological contexts are performed with commercially available heparan sulfate chains, which may differ by their post-synthetic modifications from those present in tissue(s) or at the surface of cell(s) involved in the biological processes investigated. It would thus be more appropriate to study protein—heparan sulfate interactions with heparan sulfate chains purified from tissues and/or from cell surface to identify the molecular features contributing to their binding to proteins in a biologically relevant context. Other parameters of interest to take into account in order to delineate the molecular mechanisms underlying the biological effects triggered by

Please cite this article in press as: Ricard-Blum, S., Protein—glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004

protein—heparan sulfate interactions *in vivo* are the following:

i) The quantization of heparan sulfate at the cell surface and/or in tissues to correctly use the affinity of these interactions in prioritizing them in interaction networks. This should be possible with the methods recently developed for quantitative glycomics (Staples and Zaia, 2011; Mechref et al., 2013).
ii) The protein core presenting heparan sulfate, and other GAGs, to their protein partners. The characterization of protein—GAG interactions *in vitro* with free sulfated GAG chains do not recapitulate the *in vivo* situation since heparan sulfate chains are not free either in the ECM or at the cell surface. Syndecan-glycosaminoglycan intramolecular cross-talk matters for protein—heparan sulfate interactions in different cellular microenvironments. Indeed the number, size and types of GAG chains covalently linked to the four membrane proteoglycans syndecans vary according to the cell type and/or activation, which leads to the continuous rewiring of the interaction repertoires depending on the biological context (Eriksson and Spillmann, 2012). Binding assays performed with proteoglycans, in addition to those performed with free GAGs, will be useful to discriminate interactions involving the protein core of proteoglycans from those involving their GAG chain(s), and to refine the proteoglycan interaction network. The current draft of this network comprises 30 proteoglycans (73% of the known proteoglycans), 179 partners and 557 biomolecular interactions (Peysselon et al., 2012).
iii) The crowding at the cell surface, which is covered by a layer called glycocalyx, which is enriched in proteoglycans and GAGs. The endothelial cell surface for instance is highly crowded with numerous heparan and chondroitin sulfate chains linked to syndecans and glypicans and hyaluronan, a non-sulfated GAG. The composition and thickness of the glycocalyx vary, depending on cell type. The migration of ECM proteins or protein fragments (*e.g.*, endostatin) through this intricate 3D network enriched in GAGs to reach their specific cell-surface receptors such as integrins buried into the glycocalyx warrants further investigation. Heparan sulfate likely plays a role in this process as shown for the transport of fibroblast growth factor 2 in the pericellular matrix, which is controlled by the spatial distribution of its binding sites in heparan sulfate (Duchesne et al., 2012).
iv) The role of membrane lipids in the organization and distribution of syndecans, glypicans and part-time membrane or membrane-associated proteoglycans within the plasma membrane.

Another concern is to cross-reference MatrixDB with other glycan databases to improve exchanges of protein—GAG interaction data, not only with the IMEx interaction databases, which is already done, but also with other glycan databases such as Glyco3D, a portal for structural glycosciences providing the 3D features of monosaccharides, oligosaccharides, polysaccharides, and GAG-binding proteins (http://glyco3d.cermav.cnrs.fr, Pérez et al., 2015), and SugarBind, which provides information on carbohydrate sequences to which pathogens specifically adhere (http://sugarbind.expasy.org/, Shakhsheer et al., 2013).

## References

Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F.S., Ceol, A., Chautard, E., Dana, J.M., De Las Rivas, J., Dumousseau, M., Galeota, E., Gaulton, A., Goll, J., Hancock, R.E., Isserlin, R., Jimenez, R.C., Kerssemakers, J., Khadake, J., Lynn, D.J., Michaut, M., O'Kelly, G., Ono, K., Orchard, S., Prieto, C., Razick, S., Rigina, O., Salwinski, L., Simonovic, M., Velankar, S., Winter, A., Wu, G., Bader, G.D., Cesareni, G., Donaldson, I.M., Eisenberg, D., Kleywegt, G.J., Overington, J., Ricard-Blum, S., Tyers, M., Albrecht, M., Hermjakob, H., 2011. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat. Methods 8, 528—529, http://dx.doi.org/10.1038/nmeth.1637.

Ballut, L., Sapay, N., Chautard, É., Imberty, A., Ricard-Blum, S., 2013. Mapping of heparin/heparan sulfate binding sites on αvβ3 integrin by molecular docking. J. Mol. Recognit. 26, 76—85, http://dx.doi.org/10.1002/jmr.2250.

Bülow, H.E., Hobert, O., 2006. The molecular diversity of glycosaminoglycans shapes animal development. Annu. Rev. Cell Dev. Biol. 22, 375—407, http://dx.doi.org/10.1146/annurev.cellbio.22.010605.093433.

Chautard, E., Ballut, L., Thierry-Mieg, N., Ricard-Blum, S., 2009. MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. Bioinformatics 25, 690—691, http://dx.doi.org/10.1093/bioinformatics/btp025.

Chautard, E., Fatoux-Ardore, M., Ballut, L., Thierry-Mieg, N., Ricard-Blum, S., 2011. MatrixDB, the extracellular matrix interaction database. Nucleic Acids Res. 39 (Suppl. 1), D235—D240, http://dx.doi.org/10.1093/nar/gkq830.

Chomel, C., Cazes, A., Faye, C., Bignon, M., Gomez, E., Ardidie-Robouant, C., Barret, A., Ricard-Blum, S., Muller, L., Germain, S., Monnot, C., 2009. Interaction of the coiled-coil domain with glycosaminoglycans protects angiopoietin-like 4 from proteolysis and regulates its antiangiogenic activity. FASEB J. 23, 940—949, http://dx.doi.org/10.1096/fj.08-115170.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., D'Eustachio, P., 2014. The Reactome pathway knowledgebase. Nucleic Acids Res. 42 (D1), D472—D477, http://dx.doi.org/10.1093/nar/gkt1102.

Duchesne, L., Octeau, V., Bearon, R.N., Beckett, A., Prior, I.A., Lounis, B., Fernig, D.G., 2012. Transport of fibroblast growth factor 2 in the pericellular matrix is controlled by the spatial distribution of its binding sites in heparan sulfate. PLoS Biol. 10, e1001361, http://dx.doi.org/10.1371/journal.pbio.1001361.

Eriksson, A.S., Spillmann, D., 2012. The mutual impact of syndecan-1 and its glycosaminoglycan chains-a multivariable puzzle. J. Histochem. Cytochem. 60, 936—942, http://dx.doi.org/10.1369/0022155412460242.

Esko, J.D., Kimata, K., Lindahl, U., 2009. Proteoglycans and sulfated glycosaminoglycans. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (Eds.), Essentials of Glycobiology. , 2nd edition. Cold Spring Harbor Laboratory Press, Chapter 16.

Fatoux-Ardore, M., Peysselon, F., Weiss, A., Bastien, P., Pratlong, F., Ricard-Blum, S., 2014. Large-scale investigation of *Leishmania* interaction networks with host extracellular matrix by surface plasmon resonance imaging. Infect. Immun. 82, 594—606, http://dx.doi.org/10.1128/iai.01146-13.

Faye, C., Moreau, C., Chautard, E., Jetne, R., Fukai, N., Ruggiero, F., Humphries, M.J., Olsen, B.R., Ricard-Blum, S.,

2009a. Molecular interplay between endostatin, integrins, and heparan sulfate. J. Biol. Chem. 284, 22029—22040, http://dx.doi.org/10.1074/jbc.m109.002840.

Faye, C., Chautard, E., Olsen, B.R., Ricard-Blum, S., 2009b. The first draft of the endostatin interaction network. J. Biol. Chem. 284, 22041—22047, http://dx.doi.org/10.1074/jbc.m109.002964.

Gene Ontology Consortium, 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res. 43 (D1), D1049—D1056, http://dx.doi.org/10.1093/nar/gku1179.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C., 2013. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res 41 (D1), D456—D463, http://dx.doi.org/10.1093/nar/gks1146.

Heller, S., McNaught, A., Stein, S., Tchekhovskoi, D., Pletnev, I., 2013. InChI—the worldwide chemical structure identifier standard. J. Cheminf. 5, 7, http://dx.doi.org/10.1186/1758-2946-5-7.

Heller, S.R., McNaught, A., Pletnev, I., Stein, S., Tchekhovskoi, D., 2015. InChI, the IUPAC International Chemical Identifier. J. Cheminf. 7, 23, http://dx.doi.org/10.1186/s13321-015-0068-4.

Herget, S., Ranzinger, R., Maass, K., von der Lieth, C.-W., 2008. GlycoCT-a unifying sequence format for carbohydrates. Carbohydr. Res. 343, 2162—2171, http://dx.doi.org/10.1016/j.carres.2008.03.011.

Hill, D.P., Adams, N., Bada, M., Batchelor, C., Berardini, T.Z., Dietze, H., Drabkin, H.J., Ennis, M., Foulger, R.E., Harris, M.A., Hastings, J., Kale, N.S., de Matos, P., Mungall, C.J., Owen, G., Roncaglia, P., Steinbeck, C., Turner, S., Lomax, J., 2013. Dovetailing biology and chemistry: integrating the gene ontology with the ChEBI chemical ontology. BMC Genomics 14, 513, http://dx.doi.org/10.1186/1471-2164-14-513.

Iozzo, R.V., Schaefer, L., 2015. Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. Matrix Biol. 42, 11—55, http://dx.doi.org/10.1016/j.matbio.2015.02.003.

Jin, L., Wang, W., Fang, G., 2014. Targeting protein-protein interaction by small molecules. Annu. Rev. Pharmacol. Toxicol. 54, 435—456, http://dx.doi.org/10.1146/annurev-pharmtox-011613-140028.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2015. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44 (D1), D457—D462, http://dx.doi.org/10.1093/nar/gkv1070.

Khan, S., Fung, K.W., Rodriguez, E., Patel, R., Gor, J., Mulloy, B., Perkin, S.J., 2013. The solution structure of heparan sulfate differs from that of heparin: implications for function. J. Biol. Chem. 288, 27737—27751, http://dx.doi.org/10.1074/jbc.m113.492223.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J., Bryant, S.H., 2016. PubChem substance and compound databases. Nucleic Acids Res. 44 (D1), D1202—D1213, http://dx.doi.org/10.1093/nar/gkv951.

Launay, G., Salza, R., Multedo, D., Thierry-Mieg, N., Ricard-Blum, S., 2015. MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. Nucleic Acids Res. 43 (D1), D321—D327, http://dx.doi.org/10.1093/nar/gku1091.

Lawrence, R., Lu, H., Rosenberg, R.D., Esko, J.D., Zhang, L., 2008. Disaccharide structure code for the easy representation of constituent oligosaccharides from glycosaminoglycans. Nat. Methods 5, 291—292, http://dx.doi.org/10.1038/nmeth0408-291.

Mechref, Y., Hu, Y., Desantos-Garcia, J.L., Hussein, A., Tang, H., 2013. Quantitative glycomics strategies. Mol. Cell. Proteomics 12, 874—884, http://dx.doi.org/10.1074/mcp.r112.026310.

Moreno, P., Beisken, S., Harsha, B., Muthukrishnan, V., Tudose, I., Dekker, A., Dornfeldt, S., Taruttis, F., Grosse, I., Hastings, J., Neumann, S., Steinbeck, C., 2015. BiNChE: a web tool and library for chemical enrichment analysis based on the ChEBI ontology. BMC Bioinf. 16, 56, 10.1186/s12859-015-0486-3.

Naba, A., Hoersch, S., Hynes, R.O., 2012. Towards definition of an ECM parts list: an advance on GO categories. Matrix Biol. 31, 371—372, http://dx.doi.org/10.1016/j.matbio.2012.11.008.

Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., Bridge, A., Briganti, L., Brinkman, F.S.L., Cesareni, G., Chatr-aryamontri, A., Chautard, E., Chen, C., Dumousseau, M., Goll, J., Hancock, R.E.W., Hannick, L.I., Jurisica, I., Khadake, J., Lynn, D.J., Mahadevan, U., Perfetto, L., Raghunath, A., Ricard-Blum, S., Roechert, B., Salwinski, L., Stümpflen, V., Tyers, M., Uetz, P., Xenarios, I., Hermjakob, H., 2012. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat. Methods 9, 345—350, http://dx.doi.org/10.1038/nmeth.1931.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R.C., Meldal, B., Melidoni, A.N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., Hermjakob, H., 2014. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 42 (D1), D358—D363, http://dx.doi.org/10.1093/nar/gkt1115.

Pathan, M., Keerthikumar, S., Ang, C.-S., Gangoda, L., Quek, C.Y.J., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., Bacic, A., Hill, A.F., Stroud, D.A., Ryan, M.T., Agbinya, J.I., Mariadason, J.M., Burgess, A.W., Mathivanan, S., 2015. FunRich: an open access standalone functional enrichment and interaction network analysis tool. Proteomics 15, 2597—2601, http://dx.doi.org/10.1002/pmic.201400515.

Pérez, S., Sarkar, A., Rivet, A., Breton, C., Imberty, A., 2015. Glyco3D: a portal for structural glycosciences. Methods Mol. Biol. 1273, 241—258, http://dx.doi.org/10.1007/978-1-4939-2343-4_18.

Peysselon, F., Fatoux-Ardore, M., Ricard-Blum, S., 2012. From binary interactions of glycosaminoglycans and proteoglycans to interaction networks. In: Balazs, E.A. (Ed.), Structure and Function of Biomatrix. Control of Cell Behavior and Gene Expression, vol. 5. Matrix Biology Institute, Edgewater, New Jersey (USA), pp. 277—298, Chapter 4.

Peysselon, F., Ricard-Blum, S., 2014. Heparin—protein interactions: from affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis. Matrix Biol. 35, 73—81, http://dx.doi.org/10.1016/j.matbio.2013.11.001.

Pontius, J.U., Wagner, L., Schuler, G.D., 2003. UniGene: a unified view of the transcriptome. In: The NCBI Handbook. National Center for Biotechnology Information, Bethesda (MD).

Ricard-Blum, S., Féraud, O., Lortat-Jacob, H., Rencurosi, A., Fukai, N., Dkhissi, F., Vittet, D., Imberty, A., Olsen, B.R., van der Rest, M., 2004. Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling: role of divalent cations. J. Biol. Chem. 279, 2927—2936, http://dx.doi.org/10.1074/jbc.m309868200.

Ricard-Blum, S., Beraud, M., Raynal, N., Farndale, R.W., Ruggiero, F., 2006. Structural requirements for heparin/heparan sulfate binding to type V collagen. J. Biol. Chem. 281, 25195—25204, http://dx.doi.org/10.1074/jbc.m603096200.

Shakhsheer, B., Anderson, M., Khatib, K., Tadoori, L., Joshi, L., Lisacek, F., Hirschman, L., Mullen, E., 2013. SugarBind Database (SugarBindDB): a resource of pathogen lectins and corresponding glycan targets. J. Mol. Recognit. 26, 426—431, http://dx.doi.org/10.1002/jmr.2285.

Salza, R., Peysselon, F., Chautard, E., Faye, C., Moschcovich, L., Weiss, T., Perrin-Cocon, L., Lotteau, V., Kessler, E., Ricard-

Please cite this article in press as: Ricard-Blum, S., Protein—glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004

Blum, S., 2014. Extended interaction network of procollagen C-proteinase enhancer-1 in the extracellular matrix. Biochem. J. 457, 137–149, http://dx.doi.org/10.1042/bj20130295.

Scarpellini, A., Germack, R., Lortat-Jacob, H., Muramatsu, T., Billett, E., Johnson, T., Verderio, E.A.M., 2009. Heparan sulfate proteoglycans are receptors for the cell-surface trafficking and biological activity of transglutaminase-2. J. Biol. Chem. 284, 18411–18423, http://dx.doi.org/10.1074/jbc.m109.012948.

Staples, G.O., Zaia, J., 2011. Analysis of glycosaminoglycans using mass spectrometry. Curr. Proteomics 8, 325–336, http://dx.doi.org/10.2174/157016411798220871.

Su, G., Morris, J.H., Demchak, B., Bader, G.D., 2014. Biological network exploration with cytoscape 3. Curr. Protoc. Bioinformatics 47, http://dx.doi.org/10.1002/0471250953.bi0813s47, 8.13.1–8.13.24.

Tanaka, K., Aoki-Kinoshita, K.F., Kotera, M., Sawaki, H., Tsuchiya, S., Fujita, N., Shikanai, T., Kato, M., Kawano, S., Yamada, I., Narimatsu, H., 2014. WURCS: the Web3 unique representation of carbohydrate structures. J. Chem. Inf. Model. 54, 1558–1566, http://dx.doi.org/10.1021/ci400571e.

Vivès, R.R., Seffouh, A., Lortat-Jacob, H., 2014. Post-synthetic regulation of HS structure: the Yin and Yang of the sulfs in cancer. Front. Oncol. 3, 331, http://dx.doi.org/10.3389/fonc.2013.00331.

Weiss, T., Ricard-Blum, S., Moschcovich, L., Wineman, E., Mesilaty, S., Kessler, E., 2010. Binding of procollagen C-proteinase enhancer-1 (PCPE-1) to heparin/heparan sulfate: properties and role in PCPE-1 interaction with cells. J. Biol. Chem. 285, 33867–33874, http://dx.doi.org/10.1074/jbc.m110.141366.

Please cite this article in press as: Ricard-Blum, S., Protein–glycosaminoglycan interaction networks: Focus on heparan sulfate. Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.10.004