



HAL
open science

GLYCOSAMINOGLYCAN INTERACTION NETWORKS AND DATABASES

Sylvie Ricard-Blum, Serge Perez

► **To cite this version:**

Sylvie Ricard-Blum, Serge Perez. GLYCOSAMINOGLYCAN INTERACTION NETWORKS AND DATABASES. *Current Opinion in Structural Biology*, 2022, 74, pp.102355. 10.1016/j.sbi.2022.102355 . hal-04376356

HAL Id: hal-04376356

<https://hal.science/hal-04376356v1>

Submitted on 6 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GLYCOSAMINOGLYCAN INTERACTION NETWORKS AND DATABASES

Sylvie Ricard-Blum¹, Serge Perez²

¹Univ. Lyon, Université Lyon 1, ICBMS, UMR 5246, 69622 Villeurbanne, France.

²University Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France.

Corresponding author:

Professor Sylvie RICARD-BLUM

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

Institut de Chimie et Biochimie Moléculaires et Supramoléculaires

ICBMS, UMR 5246

Université Lyon 1, Bâtiment Lederer

43 Boulevard du 11 novembre 1918

F-69622 Villeurbanne, France

Telephone: +3(0)472448232, no fax

Author e-mails

sylvie.ricard-blum@univ-lyon1.fr

sperezperez@gmail.com

KEYWORDS: Glycosaminoglycans, Interactions, Networks, Structure, Databases

SHORT TITLE: Glycosaminoglycan interactomes

ABSTRACT

Glycosaminoglycans are complex polysaccharides exhibiting a large structural and conformational diversity. These key biological players organize the extracellular matrix, contribute to cell-matrix interactions, and regulate cell signaling. Natural and synthetic libraries of glycosaminoglycans have been spotted on microarrays to find glycosaminoglycan partners, and determine the size and the chemical groups promoting protein binding. Advances in glycosaminoglycan sequencing allow the characterization of glycosaminoglycan sequences interacting with proteins, and glycosaminoglycan-mediated pull-down proteomics can identify glycosaminoglycan binding-proteins at a proteome scale in various biological samples. The analysis of the glycosaminoglycan interaction networks generated using these data gives insights into the molecular and cellular mechanisms underlying glycosaminoglycan functions. These interactomes can also be used to design inhibitors targeting specific GAG interactions for therapeutic purpose.

INTRODUCTION

Glycosaminoglycans (GAGs) are complex linear polysaccharides comprised of disaccharide repeats made of glucose or galactose and a hexuronic acid (glucuronic or iduronic acid). Five out of the six mammalian GAGs, chondroitin sulfate (CS), dermatan sulfate (DS), heparan sulfate (HS), heparin (HP) and keratan sulfate (KS), are covalently attached to core proteins to form proteoglycans (PGs) [1], and undergo several modifications (e.g., sulfation, N-acetylation, epimerization). Hyaluronan (HA) is not sulfated and does not form PGs. GAGs play major biological roles in the extracellular matrix (ECM) and at the cell surface although HS has been found in the nucleus [2]. Throughout their interactions with numerous proteins GAGs contribute to ECM assembly and organization, and to cell-matrix interactions. They are involved in many diseases such as cancer and infections [1,3]. GAG interactions mediate their functions, influence each other at various scales *in vivo*, and form dynamic and context-dependent networks. These networks can be used to study the molecular and cellular mechanisms underlying their functions, and to determine how these functions are regulated. **Figure 1** displays the roadmap to build and analyze GAG interactomes. This review covers the large-scale identification of GAG-binding proteins (GAGBPs), and provides examples of GAG interactomes generated in various molecular and cellular contexts, and of integration of additional data. The perspectives highlight the major challenges to generate 3D specific GAG interactomes.

IDENTIFICATION OF GAG-BINDING PROTEINS AND GAG-BINDING FEATURES

A number of GAGBPs have been identified by individual binding assays carried out by affinity chromatography, gel mobility shift assays, biophysical techniques [4] such as surface plasmon resonance (SPR) [5–8], bio-layer interferometry (BLI) [5,9], isothermal titration calorimetry, quartz microbalance [10], capillary electrophoresis [11], nuclear magnetic resonance (NMR) [12], and computational approaches (*Samsonov et al., and Desai et al., this volume*). However, the building of comprehensive GAG interactomes requires large-scale, high-throughput, binding assays including GAG microarrays and affinity-proteomics.

Protein arrays can be probed with a single GAG to identify its protein partners as done for KS assayed with 8268 human proteins spotted on a commercially available array [13], or for HP, HS, DS, and CSC with *E. coli* proteome arrays comprised of 4256 proteins [14,15]. Alternatively, GAG arrays can be probed with a single protein to identify its GAG repertoire. Arrays comprised of the six GAGs and selectively desulfated HP have been used to investigate their binding to ECM proteins [5,16], monomer, oligomers, and fibrils of the amyloid A β 42 peptide [17] and live pathogens [18]. Natural and synthetic GAG libraries, reflecting the diversity of sizes, sequences and chemical modifications of the glycosaminoglycome (GAGome), are required to generate arrays. This is a challenge given that there are 202 unique GAG disaccharides [19], and that the GAGome composition depends on cell types, tissues, and physio-pathological contexts. GAG libraries are also prepared for combinatorial virtual library screening (CVLS). Two libraries of 400 GAG tetrasaccharides and one library of 1000 GAG hexasaccharides were generated for CVLS against HS 3-O-Sulfotransferase 1 [20], and a library of 1752 CS oligosaccharides (24 disaccharides, 192 tetrasaccharides, and 1536 hexasaccharides) was screened against transforming growth factor β 2 by CVLS [21].

The use of structurally defined oligosaccharides to generate GAG microarrays [22,23], and strategies to synthesize anomerically modified HA synthetic oligosaccharides with defined sulfation patterns have been recently reviewed [24]. Consequently, only some recent examples of GAG libraries of various size and structural diversity synthesized by chemical or chemoenzymatic methods, are listed in **Table 1**. CS hexasaccharides synthesized on a solid support with a photolabile linker were used to prepare CS arrays [25]. A combinatorial HS mimetic array can be prepared by the self-assembly of several ratios of two disaccharides with different physicochemical properties (lactose and sulfated lactose) to create a variety of modifications, charges, and topologies accounting for HS structural diversity. This array was able to discriminate the binding of a wild-type protein and its single-point mutants [26]. The molecular diversity of the array could be increased by additional disaccharides with other physicochemical properties. Two HS libraries comprising 27 hexasaccharides with or without 3-O-sulfate groups [27], and 47 tetrasaccharides with 3-O- and N-sulfation [28] have been synthesized to determine the role of these chemical groups in protein binding. The GAG size matters for protein interaction, and a HS library of 93 di-, tetra-, hexa-, and octasaccharides of various sequences and sulfation patterns has been generated to identify the minimal recognition sequence of the spike protein of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) [29]. GAG may be conjugated to bovine serum albumin (BSA) [30] or to neoglycolipids (NGLs) [31] before spotting on arrays.

Cells can be used to synthesize GAG libraries. Chondroitin, and heparosan (unsulfated HS) were prepared from *E. coli*, which can synthesize ¹³C-labeled polysaccharides to be used for NMR experiments when grown in presence of ¹³C-labeled glucose [9]. Genes encoding GAG biosynthetic enzymes have been edited, individually or in combination, in Chinese hamster ovary cells to produce a cell-based GAG library (GAGOme) with a large repertoire of modifications [32], and in mouse lung endothelial cells to generate a HS mutant mouse lung endothelial cell library [33]. In addition, two arrays made of 24 HS oligosaccharides and 46 GAG oligosaccharides (HA, CS, CSDS, HP, and HS oligosaccharides ranging from decasaccharides to eicosasaccharides) are commercially available (<https://www.zbiotech.com>).

Fluorescence is the most common detection method on GAG and protein arrays [14,27–31], but GAG arrays have been designed to detect binding events by a label-free method, SPR imaging (SPRi) [5,16,26,34], to calculate kinetics and affinity of the interactions. SPRi arrays can be interfaced with a MALDI mass spectrometer (MS) for on-chip analyses of the oligosaccharides captured by the proteins spotted on the arrays [34]. Proteins captured from biological media on any glycan arrays can be identified by MALDI-time-of-flight MS/MS after *in situ* proteolytic digestion [35]. Time of flight secondary ion mass spectrometry was used to analyze nanogram amounts of the six GAGs (CS, DS, HA, HP, HS and KS) spotted on a microarray [36]. Experiments performed with GAG arrays should be reported following the MIRAGE (Minimum Information Required for A Glycomics Experiment) guidelines [37].

Affinity proteomics-based approaches have been designed to identify GAGBPs in various cells and tissues, and to compare their patterns in health and disease (**Table 2**) [38–40]. An affinity-proteomics approach, named Limited Proteolysis in the absence of

denaturation, Heparin-Affinity chromatography, and high-resolution LC-MS/MS proteomics (LPHAMS), allows the enrichment of low-abundance HSBPs in several cell types, and the mapping of HP-binding protein domains [6]. Another glycan-mediated pull-down proteomics method based on affinity chromatography on immobilized GAGs coupled to nano-LC/MS/MS can evaluate the affinity of GAG-protein interactions by eluting GAGBPs with a salt step-gradient [38]. HSPGs have been immobilized onto nano-affinity columns (internal diameter: 75 μm) to screen potential partners [41]. A filter-entrapment enrichment pull-down assay with a membrane instead of an affinity column can be used to enrich ultra LMW HP sequences interacting with a protein before analysis by capillary zone electrophoresis MS/MS [8].

GAGBPs can be retrieved from interaction databases, and more specifically from the extracellular interaction database MatrixDB, which is the only database curating GAG-interactions on a regular basis (<http://matrixdb.univ-lyon1.fr/>) [42]. MatrixDB uses the curation interface of the IntAct database (<https://www.ebi.ac.uk/intact/home>) [43]. Both databases belong to the International Molecular Exchange Consortium (IMEx, <http://www.imexconsortium.org/>) providing an open access unified dataset of molecular interactions [44]. A step-by-step protocol has been set up to build protein-GAG interaction networks using MatrixDB data and its i-Navigator [45]. MatrixDB data have been integrated into several GAG interactomes [1,46,47]. The STRING database (Search Tool for the Retrieval of Interacting Genes, <https://string-db.org/>) [48] was used to build HPBP interactomes in normal and diseased pancreas [39], in cardiovascular diseases [49], in periodontitis [50], and the global HS interactome [51]. HP/HSBP networks can be generated *in silico* as shown for those involved in cardiovascular diseases, using a published dataset [46], and the Open Targets database (<https://www.opentargets.org>) to identify the HP/HSBPs related to four cardiovascular diseases [49], and in periodontal diseases using literature curation and the STRING database [50].

GAG INTERACTOMES

Interaction networks are analyzed to generate hypotheses, guide the design of new experiments, and define general principles of molecular, supramolecular, cellular, and tissular assembly, organization and functions. Computational tools and resources to analyze ECM networks including GAG interactomes have been recently reviewed [52]. Those used for perform functional enrichment analyses of GAG interactomes are the Database for Annotation, Visualization and Integrated Discovery (DAVID) [38,50,51], the Biological Network Gene Ontology tool (BiNGO) [46], the Functional Enrichment tool FunRich [47], and Metascape (<https://metascape.org>) [49]. Reactome [47] and KEGG [14,46] are specifically used for biological and metabolic pathways. The topology of GAG networks has been determined with STRING [39] or the Cytoscape apps Network analyzer [49,50], whereas the clusters and the most connected proteins have been identified with MCODE and cytoHubba apps respectively [49].

Table 2 lists the major GAG interactomes available. HP/HSBPs regulating angiogenesis have been identified by literature curation but not analyzed further [53]. The identification of

KSBPs by protein arrays has shown that 75 of them are kinases and 40 nerve-related proteins [13]. A systems biology approach has been used to investigate HP/HS interactomes [46,51] and the global GAG interactome including HA, CS, DS and KS in addition to HP/HS [47]. The HP/HS network is highly interconnected and enriched in biological processes such as cell-cell signaling, regulation of cell proliferation, system development, immune and inflammatory responses, and in pathways controlling biological processes characteristic of multicellular organisms [46]. The link between HS-binding and the evolution toward multicellularity has been confirmed on a larger HP/HS interactome, which was clustered into five classes (*i.e.*, chemokines and complement proteins, growth factors, ion sensing and transport, protease activity and coagulation, lipoprotein metabolism, and ECM assembly) [51]. The HPBPs identified in human microvascular endothelial cells are enriched in functions related to protein degradation and folding, glycan metabolism, inflammatory processes, apoptosis, cytoskeleton regulation, RNA recognition, splicing, and synthesis [38]. The cytosolic HPBPs isolated from human pulmonary fibroblasts are also associated with apoptosis regulation and RNA processing [38]. Signal transduction, cell communication and growth, immune response, ECM organization, hemostasis, development and metabolism of proteins are enriched in the global GAG interactome [47]. In this first draft of the GAG interactome, more than 80% of HP/HSBPs and KSBPs bind only to HP/HS and KS respectively, ~60% of HABPs interact with HA only and ~25% of CSBPs bind to CS only [47]. In contrast, all DSBPs interact with at least another GAG in addition to DS [47]. This low specificity of DS is consistent with the finding that only five proteins out of the 129 GAGBPs identified in human plasma bind only to DS [38]. Affinity purification-proteomics has been applied to several human proteomes to explore GAG-binding specificity and relative affinity [38], and to identify proteins binding to HP and sulfated synthetic HA [40].

GAG-protein interactions play also a role in diseases [3]. GAG-pathogen interactions have been investigated at a proteome scale in one strain of *E. coli* [14,15], and in 24 strains of intact, live, *Leishmania* parasites at the ECM scale [18]. *E. coli* HPBPs and HSBPs play a role in bacterial amino acid metabolism, and in enhanced lipopolysaccharide biosynthesis, whereas lipopolysaccharide biosynthesis and ribosome are enriched in the entire *E. coli* GAGBP dataset [14]. The interactome of the three IdoA-containing GAGs (HP, HS and DS), which can mediate host-bacteria interactions, has been generated using CSC, a non-IdoA-containing GAG, as a negative control, and shown to be enriched in functions related to protein synthesis and metabolism [15]. The 15 proteins interacting with the IdoA-containing GAGs share a conserved amino acid sequence [15]. Most *Leishmania* strains bind to two IdoA-containing GAGs, high molecular weight (HMW) HP (83%) and HS (62%), whereas less than 10% of them interact with DS, the third IdoA-containing GAG [18]. HP/HSBPs involved in cardiovascular diseases are associated with cell activation, movement, migration, and adhesion, and they upregulate the expression of ECM-affiliated proteins, ECM regulators, and secreted factors [49]. In acute pancreatitis, HPBPs are associated with acute phase response signaling, complement activation and mitochondrial dysfunction in addition to molecular transport and cellular movement found in normal pancreas [39]. The analysis of the interactome of HPBPs differentially expressed in periodontal diseases, including matrix metalloproteases, cytokines

and chemokines, highlights the importance a multi-targeted approach in periodontitis treatment [50].

INTEGRATION OF ADDITIONAL DATA IN GAG NETWORKS

The biological context promotes or abolishes interactions, which rewires the interactomes. GAG interactomes associated with specific proteomes can be built by experimentally targeting proteomes with protein arrays [13,14] and affinity-based methods [6,38,40]. Alternatively, context-specific networks can be generated by integrating GAGomic [54] and proteoglycanomic [55] data collected in various biological samples into GAG networks combining interaction data from different tissues, species, and methods. The molecular context also matters for interactions established by sulfated GAGs, which are attached to proteins *in vivo* to form PGs. The core protein of PGs may modulate protein affinity for HS as reported for an ECM bioactive fragment, endostatin [56]. The building of PG interactomes would thus provide a molecular context to discriminate GAG interactions mediated by protein cores from those mediated by their GAG chains as done for the interactome of the four membrane PG syndecans [57]. Affinity and kinetics of interactions can also be added to GAG networks. HP/HSBPs have been ranked according to their affinity in the endostatin network regulating angiogenesis at the endothelial cell surface [56]. Lipoproteins and non-ECM secreted proteins have the lowest dissociation rates, whereas matrisome-associated proteins and lipoproteins bind faster to HP than other proteins and have the highest affinity for HP [56].

Another critical area of investigation is the identification of GAG sequences interacting with proteins. There are 202 unique GAG disaccharides, including 48 HS disaccharides, which highlights the structural diversity of GAGs, and makes their sequencing challenging [19]. GAG patterns of chemical modifications form glycodes, which are cell-, tissue- or disease-specific. The identification of GAG sequences interacting with proteins is thus of crucial importance to decipher the glycode specificity. Few binding assays were performed with defined GAG sequences, but the availability of GAG libraries and their use to prepare GAG arrays will help to determine GAG sequences binding to proteins. Structural domains can be mapped in complex GAG mixtures using the GAGDoMa approach based on the enzymatic depolymerization of GAGs and on the analysis of the internal, terminating, and initiating GAG domains generated by LC-MS/MS [58]. The advances in MS and infrared spectroscopy should also ease the analysis of GAG sequences binding to proteins [59–61]. Another limitation to investigate the glycode specificity is that GAG sequences are reported in publications in heterogeneous formats, ranging from free-text and images of chemical formulae to sequences written in IUPAC, GlycoCT, and WURCS formats. This prevents GAG interaction data mining with computational tools. We have thus designed a pipeline consistent with the standards of the IMEx consortium and glycan repositories and databases such as GlyTouCan (<https://glytoucan.org/>) [62] and the Chemical Entities of Biological Interest database (ChEBI, <https://www.ebi.ac.uk/chebi/>) [63] to ease the conversion of the various GAG formats. MatrixDB is interoperable with these databases, and with the Glygen portal providing an interface to explore glycoscience data (<https://data.glygen.org>) [64]. A workflow has been

designed to improve the interoperability of the glycan-containing databases ChEBI, PubChem, and GlyGen [65].

The analysis of conserved basic amino acid sequences among 437 HP/HSBPs has shown that HP/HS-binding likely results from the 3D arrangement on the protein surface of small, independent basic amino acid sequences [66]. The integration of the 3D structures of GAGs and GAG-protein complexes in GAG networks is thus a prerequisite to build 3D GAG interactomes. GAGs have broad conformational landscapes, and conformational changes occur upon GAG binding to proteins. They contribute to the specificity of the binding but are still poorly investigated. The analysis of 75 protein–GAG complexes has shown that the glycosidic Φ and Ψ torsion angles of HP/HS and HA oligosaccharides complexed with proteins are clustered into two major areas. The sulfate groups have a minor influence on GAG conformational space whereas the conformation of iduronic acid induces a spatial restriction when located at the reducing end of disaccharides [19]. GAG structures are available in generic databases such as the Protein Data Bank (<https://www.rcsb.org/>), and the Cambridge Structural Database (<https://www.ccdc.cam.ac.uk/>). The dedicated database GAG-DB (<https://gagdb.glycopedia.eu/>) [67] covers the three-dimensional landscape of GAGs. It contains 123 entries reporting GAG composition, sequence, and topology, and provides the 3D visualization of contacts between GAGs and their partners using the protein-ligand interaction profiler [68]. GAG-DB also displays the structure adopted by GAGs in solid-state or in solution, alone or complexed with their partners. Experimental 3D structures of GAG-binding sequences are also on MatrixDB web site. The 3D models of these sequences can be generated on MatrixDB web site *via* the polysaccharide builder program POLYS [69] implemented in the database [42].

CONCLUSIONS AND PERSPECTIVES

The creation of GAG libraries by various approaches paves the way to the identification of additional GAGBPs and GAG oligosaccharides binding to proteins. The synthesis of CS by engineered *E. coli* provides an alternative to the purification of GAGs from animal tissues. In addition this approach can be used to prepare isotope-labeled GAGs for structural studies [70]. The recent advances in GAG sequencing by MS and infrared spectroscopy [59–61] (*Greis et al., this volume*) will ease the determination of GAG glycocodes for a variety of structural and functional protein families. The design of GAG arrays could be improved by mimicking the multivalent display occurring *in vivo*. Affinity-based techniques were successfully used to explore GAGBPs in various proteomes. They will be applied to further healthy and diseased proteomes in order to map GAGs at an organism level, and to determine disease-related changes in GAG interactomes, with a focus on GAG–pathogen interactomes, which are currently underexplored. GAG interaction datasets should be consistent with existing standard glycan formats [19] and experimental guidelines [37], and computational tools should be developed to explore these datasets. Integrating structural data in GAG interactomes is a prerequisite to generate a 3D interactome, and to investigate the relationships between glycocodes, involving both linear sequences and 3D motifs, and GABP

localization and molecular or biological functions. The structure of GAGs of various sizes and sequences and of GAG-protein complexes can be generated experimentally by X-ray and NMR, or by combining small angle X-ray scattering data with molecular modeling approaches [71] (*Samsonov et al., and Desai et al., this volume*). In absence of experimental data, 3D atomic-resolution models of non-sulfated GAGs with lengths up to 200 mers will be built using molecular dynamics [72]. Another topic worth of investigation is the study of GAG-GAG interactions, which should provide new insights into GAG functions. GAG self-interaction has been suggested to occur in tendon where adjacent rings formed by GAGs around collagen fibrils merge to form a planar network [73]. Integrating the above data in GAG interactomes generated from healthy and diseased cells and tissues will be of major importance to design inhibitors targeting specific GAG interactions for therapeutic purpose.

ACKNOWLEDGEMENTS

Sylvain D. Vallet (ICBMS, UMR 5246, Université Lyon, France) and Olivier Clerc (ICBMS, UMR 5246, Université Lyon, France) are acknowledged for their excellent contributions to the work made by the authors.

REFERENCES

1. Karamanos NK, Piperigkou Z, Theocharis AD, Watanabe H, Franchi M, Baud S, Brézillon S, Götte M, Passi A, Vigetti D, et al.: **Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics**. *Chem Rev* 2018, **118**:9152–9232.
2. Stewart MD, Sanderson RD: **Heparan sulfate in the nucleus and its control of cellular functions**. *Matrix Biol* 2014, **35**:56–59.
3. Shi D, Sheng A, Chi L: **Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease**. *Front Mol Biosci* 2021, **8**:64.
4. Zhang B, Chi L: **Chondroitin Sulfate/Dermatan Sulfate-Protein Interactions and Their Biological Functions in Human Diseases: Implications and Analytical Tools**. *Front Cell Dev Biol* 2021, **9**:693563.
5. Vallet SD, Deddens L, Vonarburg A, Salza R, Faye C, Aranyos A, Thierry-Mieg N, Ricard-Blum S: **Strategies for Building Protein–Glycosaminoglycan Interaction Networks Combining SPRI, SPR, and BLI**. In *Handbook of Surface Plasmon Resonance, 2nd Edition*. R.B.M. Schasfoort editor. The Royal Society of Chemistry; 2017:398–414.
6. Sandoval DR, Gomez Toledo A, Painter CD, Tota EM, Sheikh MO, West AMV, Frank MM, Wells L, Xu D, Bicknell R, et al.: **Proteomics-based screening of the endothelial heparan sulfate interactome reveals that C-type lectin 14a (CLEC14A) is a heparin-binding protein**. *J Biol Chem* 2020, **295**:2804–2821.

** A proteomics workflow based on limited proteolysis of live cells without denaturation and fixation for the identification and characterization of membrane-anchored and extracellular HS-binding proteins.

7. Yan L, Song Y, Xia K, He P, Zhang F, Chen S, Pouliot R, Weiss DJ, Tandon R, Bates JT, et al.: **Heparan sulfates from bat and human lung and their binding to the spike protein of SARS-CoV-2 virus.** *Carbohydr Polym* 2021, **260**:117797.
8. Yu Y, Zhang F, Renois-Predelus G, Amster IJ, Linhardt RJ: **Filter-entrapment enrichment pull-down assay for glycosaminoglycan structural characterization and protein interaction.** *Carbohydr Polym* 2020, **245**:116623.
* An oligosaccharide-protein pull-down assay that can serve as an analytical platform for the functional screening of low- and high-affinity GAG oligosaccharides binding to proteins.
9. Laguri C, Sadir R, Gout E, Vivès RR, Lortat-Jacob H: **Preparation and Characterization of Heparan Sulfate-Derived Oligosaccharides to Investigate Protein-GAG Interaction and HS Biosynthesis Enzyme Activity.** *Methods Mol Biol* 2022, **2303**:121–137.
10. Migliorini E, Thakar D, Kühnle J, Sadir R, Dyer DP, Li Y, Sun C, Volkman BF, Handel TM, Coche-Guerente L, et al.: **Cytokines and growth factors cross-link heparan sulfate.** *Open Biol* 2015, **5**:150046.
11. Liang A, Desai U: **Advances in Studying Glycosaminoglycan-Protein Interactions Using Capillary Electrophoresis.** *Methods Mol Biol* 2022, **2303**:365–387.
12. Pomin VH, Wang X: **Glycosaminoglycan-Protein Interactions by Nuclear Magnetic Resonance (NMR) Spectroscopy.** *Molecules* 2018, **23**:E2314.
13. Conrad AH, Zhang Y, Tasheva ES, Conrad GW: **Proteomic analysis of potential keratan sulfate, chondroitin sulfate A, and hyaluronic acid molecular interactions.** *Invest Ophthalmol Vis Sci* 2010, **51**:4500–4515.
14. Hsiao FS-H, Sutandy FR, Syu G-D, Chen Y-W, Lin J-M, Chen C-S: **Systematic protein interactome analysis of glycosaminoglycans revealed YcbS as a novel bacterial virulence factor.** *Sci Rep* 2016, **6**:28425.
15. Hsiao FS-H, Yang S-K, Lin J-M, Chen Y-W, Chen C-S: **Protein interactome analysis of iduronic acid-containing glycosaminoglycans reveals a novel flagellar invasion factor MbhA.** *J Proteomics* 2019, **208**:103485.
* Identification of the interactome of three IdoA-containing GAGs (HP, HS and DS) using E. coli proteome arrays.
16. Vallet SD, Miele AE, Uciechowska-Kaczmarzyk U, Liwo A, Duclos B, Samsonov SA, Ricard-Blum S: **Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners.** *Sci Rep* 2018, **8**:11768.
17. Salza R, Lethias C, Ricard-Blum S: **The Multimerization State of the Amyloid- β 42 Amyloid Peptide Governs its Interaction Network with the Extracellular Matrix.** *J Alzheimers Dis* 2017, **56**:991–1005.
18. Fatoux-Ardore M, Peysselon F, Weiss A, Bastien P, Pratlong F, Ricard-Blum S: **Large-scale investigation of Leishmania interaction networks with host extracellular matrix by surface plasmon resonance imaging.** *Infect Immun* 2014, **82**:594–606.

19. Clerc O, Mariethoz J, Rivet A, Lisacek F, Pérez S, Ricard-Blum S: **A pipeline to translate glycosaminoglycan sequences into 3D models. Application to the exploration of glycosaminoglycan conformational space.** *Glycobiology* 2019, **29**:36–44.
20. Sankaranarayanan NV, Bi Y, Kuberan B, Desai UR: **Combinatorial virtual library screening analysis of antithrombin binding oligosaccharide motif generation by heparan sulfate 3-O-Sulfotransferase 1.** *Comput Struct Biotechnol J* 2020, **18**:933–941.
21. Sankaranarayanan NV, Nagarajan B, Desai UR: **Combinatorial Virtual Library Screening Study of Transforming Growth Factor- β 2-Chondroitin Sulfate System.** *Int J Mol Sci* 2021, **22**:7542.
22. Pomin VH, Wang X: **Synthetic Oligosaccharide Libraries and Microarray Technology: A Powerful Combination for the Success of Current Glycosaminoglycan Interactomics.** *ChemMedChem* 2018, **13**:648–661.
23. Xu D, Arnold K, Liu J: **Using structurally defined oligosaccharides to understand the interactions between proteins and heparan sulfate.** *Curr Opin Struct Biol* 2018, **50**:155–161.
24. Schiller J, Lemmnitzer K, Dürig J-N, Rademann J: **Insights into structure, affinity, specificity, and function of GAG-protein interactions through the chemoenzymatic preparation of defined sulfated oligohyaluronans.** *Biol Chem* 2021, **402**:1375–1384.
25. Liang C-F, Hahm HS, Sabbavarapu NM, Seeberger PH: **Automated Synthesis of Chondroitin Sulfate Oligosaccharides.** *Methods Mol Biol* 2022, **2303**:319–327.
26. Genua M, Garçon L-A, Sergeeva YN, Saesen E, Musnier B, Buhot A, Billon M, Gout E, Sadir R, Lortat-Jacob H, et al.: **Discrimination of deletion to point cytokine mutants based on an array of cross-reactive receptors mimicking protein recognition by heparan sulfate.** *Anal Bioanal Chem* 2021, doi:10.1007/s00216-021-03516-z.
27. Chopra P, Joshi A, Wu J, Lu W, Yadavalli T, Wolfert MA, Shukla D, Zaia J, Boons G-J: **The 3-O-sulfation of heparan sulfate modulates protein binding and lyase degradation.** *Proc Natl Acad Sci U S A* 2021, **118**:e2012935118.
* A library of 27 HS oligosaccharides with and without 3-O sulfate printed as a glycan microarray to determine ligand requirements for proteins binding to HS modified by 3-O-sulfate groups.
28. Zong C, Venot A, Li X, Lu W, Xiao W, Wilkes J-SL, Salanga CL, Handel TM, Wang L, Wolfert MA, et al.: **Heparan Sulfate Microarray Reveals That Heparan Sulfate-Protein Binding Exhibits Different Ligand Requirements.** *J Am Chem Soc* 2017, **139**:9534–9543.
29. Liu L, Chopra P, Li X, Bouwman KM, Tompkins SM, Wolfert MA, de Vries RP, Boons G-J: **Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2.** *ACS Cent Sci* 2021, **7**:1009–1018.
30. Watanabe T, Takeda K, Hiemori K, Minamisawa T, Tateno H: **A glycosaminoglycan microarray identifies the binding of SARS-CoV-2 spike protein to chondroitin sulfate E.** *FEBS Lett* 2021, **595**:2341–2349.
31. Gyapon-Quast F, Goicoechea de Jorge E, Malik T, Wu N, Yu J, Chai W, Feizi T, Liu Y, Pickering MC: **Defining the Glycosaminoglycan Interactions of Complement Factor H-Related Protein 5.** *J Immunol* 2021, **207**:534–541.

32. Chen Y-H, Narimatsu Y, Clausen TM, Gomes C, Karlsson R, Steentoft C, Spleid CB, Gustavsson T, Salanti A, Persson A, et al.: **The GAGome: a cell-based library of displayed glycosaminoglycans.** *Nat Methods* 2018, **15**:881–888.
** A cell-based method to produce and display distinct a GAG library (GAGome) with a broad repertoire of modifications by editing, in combination or individually, GAG biosynthetic enzymes in Chinese hamster ovary cells.
33. Qiu H, Shi S, Yue J, Xin M, Nairn AV, Lin L, Liu X, Li G, Archer-Hartmann SA, Dela Rosa M, et al.: **A mutant-cell library for systematic analysis of heparan sulfate structure-function relationships.** *Nat Methods* 2018, **15**:889–899.
** Development of a HS mutant mouse lung endothelial cell library by deleting genes encoding HS biosynthetic and post-synthetic enzymes, providing a test platform to study HS structure-function e in a cellular context.
34. Przybylski C, Gonnet F, Saesen E, Lortat-Jacob H, Daniel R: **Surface plasmon resonance imaging coupled to on-chip mass spectrometry: a new tool to probe protein-GAG interactions.** *Anal Bioanal Chem* 2020, **412**:507–519.
*Analysis of protein-GAG interactions by coupling SPR imaging and MALDI ToF MS.
35. Gray CJ, Sánchez-Ruiz A, Šardžiková I, Ahmed YA, Miller RL, Reyes Martinez JE, Pallister E, Huang K, Both P, Hartmann M, et al.: **Label-Free Discovery Array Platform for the Characterization of Glycan Binding Proteins and Glycoproteins.** *Anal Chem* 2017, **89**:4444–4451.
36. Hook AL, Hogwood J, Gray E, Mulloy B, Merry CLR: **High sensitivity analysis of nanogram quantities of glycosaminoglycans using ToF-SIMS.** *Commun Chem* 2021, **4**:1–8.
37. Campbell MP, Abrahams JL, Rapp E, Struwe WB, Costello CE, Novotny M, Ranzinger R, York WS, Kolarich D, Rudd PM, et al.: **The minimum information required for a glycomics experiment (MIRAGE) project: LC guidelines.** *Glycobiology* 2019, **29**:349–354.
38. Gesslbauer B, Derler R, Handwerker C, Seles E, Kungl AJ: **Exploring the glycosaminoglycan-protein interaction network by glycan-mediated pull-down proteomics.** *Electrophoresis* 2016, **37**:1437–1447.
39. Nunes QM, Su D, Brownridge PJ, Simpson DM, Sun C, Li Y, Bui TP, Zhang X, Huang W, Rigden DJ, et al.: **The heparin-binding proteome in normal pancreas and murine experimental acute pancreatitis.** *PLoS One* 2019, **14**:e0217633.
*Identification of differentially expressed HPBPs in a murine model of acute pancreatitis, and analysis of the resulting interactome.
40. Großkopf H, Vogel S, Müller CD, Köhling S, Dürig J-N, Möller S, Schnabelrauch M, Rademann J, Hempel U, von Bergen M, et al.: **Identification of intracellular glycosaminoglycan-interacting proteins by affinity purification mass spectrometry.** *Biol Chem* 2021, **402**:1427–1440.
**Identification of intracellular and cell-associated proteins binding to at least one sulfated GAG (synthetic HAs and HP) by affinity purification coupled to mass spectrometry carried out in human bone marrow stromal cells, and analysis of the resulting interactome.
41. André C, Guillaume YC: **Development of nano affinity columns for the study of ligand (including SARS-CoV-2 related proteins) binding to heparan sulfate proteoglycans.** *Anal Methods* 2021, **13**:3050–3058.

42. Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N, Ricard-Blum S: **MatrixDB: integration of new data with a focus on glycosaminoglycan interactions.** *Nucleic Acids Res* 2019, **47**:D376–D381.
**An updated version of MatrixDB database focused on interactions established by ECM proteins and GAGs, including cross-references of GAGs to other databases, GAG sequences binding to proteins in various formats, and the GAG Builder online tool to build 3D models of GAG oligosaccharides
43. Del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, Perfetto L, How K, Ratan P, Shirodkar G, et al.: **The IntAct database: efficient access to fine-grained molecular interaction data.** *Nucleic Acids Res* 2021, doi:10.1093/nar/gkab1006.
44. Porras P, Barrera E, Bridge A, del-Toro N, Cesareni G, Duesbury M, Hermjakob H, Iannuccelli M, Jurisica I, Kotlyar M, et al.: **Towards a unified open access dataset of molecular interactions.** *Nat Commun* 2020, **11**:6144.
45. Berthollier C, Vallet SD, Deniaud M, Clerc O, Ricard-Blum S: **Building Protein-Protein and Protein-Glycosaminoglycan Interaction Networks Using MatrixDB, the Extracellular Matrix Interaction Database.** *Curr Protoc* 2021, **1**:e47.
46. Ori A, Wilkinson MC, Fernig DG: **A systems biology approach for the investigation of the heparin/heparan sulfate interactome.** *J Biol Chem* 2011, **286**:19892–19904.
47. Vallet SD, Clerc O, Ricard-Blum S: **Glycosaminoglycan-Protein Interactions: The First Draft of the Glycosaminoglycan Interactome.** *J Histochem Cytochem* 2021, **69**:93-104.
**The first comprehensive draft of the GAG interactome including 827 GAG-binding proteins. It will serve as a scaffold for integrating structural and expression data and for identifying druggable GAG-protein interactions.
48. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al.: **The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.** *Nucleic Acids Res* 2021, **49**:D605–D612.
49. Cai Y, Zhang X, Shen J, Jiang B, Hu D, Zhao M: **Heparin-Binding Protein: A Novel Biomarker Linking Four Different Cardiovascular Diseases.** *Cardiol Res Pract* 2020, **2020**:9575373.
50. Lackey B, Nunes QM, Higham SM, Fernig DG, Valappil SP: **In silico analyses of heparin binding proteins expression in human periodontal tissues.** *BMC Res Notes* 2016, **9**:53.
51. Gómez Toledo A, Sorrentino JT, Sandoval DR, Malmström J, Lewis NE, Esko JD: **A Systems View of the Heparan Sulfate Interactome.** *J Histochem Cytochem* 2021, **69**:105–119.
** The HS interactome comprising 530 HS-binding proteins grouped into five major functional classes by applying a networking approach.
52. Ricard-Blum S: **Extracellular Matrix Networks: From Connections to Functions.** In *Extracellular Matrix Omics*. Edited by Ricard-Blum S. Springer International Publishing; 2020:101–129.
53. Chiodelli P, Bugatti A, Urbinati C, Rusnati M: **Heparin/Heparan sulfate proteoglycans glycomic interactome in angiogenesis: biological implications and therapeutical use.** *Molecules* 2015, **20**:6342–6388.

54. Sethi MK, Zaia J: **Historical Overview of Integrated GAG-omics and Proteomics**. In *Extracellular Matrix Omics*. Edited by Ricard-Blum S. Springer International Publishing; 2020:83–99.
55. Koch CD, Apte SS: **Characterization of Proteoglycanomes by Mass Spectrometry**. In *Extracellular Matrix Omics*. Edited by Ricard-Blum S. Springer International Publishing; 2020:69–82.
56. Peysselon F, Ricard-Blum S: **Heparin-protein interactions: from affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis**. *Matrix Biol* 2014, **35**:73–81.
57. Gondelaud F, Ricard-Blum S: **Structures and interactions of syndecans**. *FEBS J* 2019, **286**:2994–3007.
58. Persson A, Vorontsov E, Larson G, Nilsson J: **Glycosaminoglycan Domain Mapping of Cellular Chondroitin/Dermatan Sulfates**. *Sci Rep* 2020, **10**:3506.
*An approach (GAGDoMa) for structural domain mapping of complex mixtures of GAGs based on enzymatic depolymerization of GAGs, reversed-phase ion-pairing chromatography with negative mode higher-energy collision dissociation tandem mass spectrometry.
59. Song Y, Zhang F, Linhardt RJ: **Analysis of the Glycosaminoglycan Chains of Proteoglycans**. *J Histochem Cytochem* 2021, **69**:121–135.
60. Gray CJ, Migas LG, Barran PE, Pagel K, Seeberger PH, Evers CE, Boons G-J, Pohl NLB, Compagnon I, Widmalm G, et al.: **Advancing Solutions to the Carbohydrate Sequencing Challenge**. *J Am Chem Soc* 2019, **141**:14463–14479.
61. Lettow M, Grabarics M, Greis K, Mucha E, Thomas DA, Chopra P, Boons G-J, Karlsson R, Turnbull JE, Meijer G, et al.: **Cryogenic Infrared Spectroscopy Reveals Structural Modularity in the Vibrational Fingerprints of Heparan Sulfate Diastereomers**. *Anal Chem* 2020, **92**:10228–10232.
62. Fujita A, Aoki NP, Shinmachi D, Matsubara M, Tsuchiya S, Shiota M, Ono T, Yamada I, Aoki-Kinoshita KF: **The international glycan repository GlyYouCan version 3.0**. *Nucleic Acids Res* 2021, **49**:D1529–D1533.
63. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C: **ChEBI in 2016: Improved services and an expanding collection of metabolites**. *Nucleic Acids Res* 2016, **44**:D1214–1219.
64. York WS, Mazumder R, Ranzinger R, Edwards N, Kahsay R, Aoki-Kinoshita KF, Campbell MP, Cummings RD, Feizi T, Martin M, et al.: **GlyGen: Computational and Informatics Resources for Glycoscience**. *Glycobiology* 2020, **30**:72–73.
65. Navelkar R, Owen G, Muthukrishnan V, Thiessen P, Cheng T, Bolton E, Edwards N, Tiemeyer M, Campbell MP, Martin M, et al.: **Enhancing the interoperability of glycan data flow between ChEBI, PubChem, and GlyGen**. *Glycobiology* 2021, doi:10.1093/glycob/cwab078.
66. Rudd TR, Preston MD, Yates EA: **The nature of the conserved basic amino acid sequences found among 437 heparin binding proteins determined by network analysis**. *Mol Biosyst* 2017, **13**:852–865.

67. Pérez S, Bonnardel F, Lisacek F, Imberty A, Ricard Blum S, Makshakova O: **GAG-DB, the New Interface of the Three-Dimensional Landscape of Glycosaminoglycans**. *Biomolecules* 2020, **10**:E1660.
**GAG-DB provides a family-based classification of GAGs, experimental structures of GAG and GAG-protein complexes, the visualization of contacts between GAGs and their protein ligands, and the evaluation of quaternary structure implemented via the Protein-Ligand Interaction Profiler.
68. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M: **PLIP: fully automated protein-ligand interaction profiler**. *Nucleic Acids Res* 2015, **43**:W443-447.
69. Engelsen SB, Hansen PI, Pérez S: **POLYS 2.0: An open source software package for building three-dimensional structures of polysaccharides**. *Biopolymers* 2014, **101**:733–743.
70. Badri A, Williams A, Awofiranye A, Datta P, Xia K, He W, Fraser K, Dordick JS, Linhardt RJ, Koffas MAG: **Complete biosynthesis of a sulfated chondroitin in Escherichia coli**. *Nat Commun* 2021, **12**:1389.
**Metabolically engineered E. coli cells capable of synthesizing intracellular chondroitin sulfate. This strategy can be extrapolated to develop microbial cells producing other GAGs.
71. Almond A: **Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges**. *Curr Opin Struct Biol* 2018, **50**:58–64.
72. Whitmore EK, Martin D, Guvench O: **Constructing 3-Dimensional Atomic-Resolution Models of Nonsulfated Glycosaminoglycans with Arbitrary Lengths Using Conformations from Molecular Dynamics**. *Int J Mol Sci* 2020, **21**:E7699.
73. Watanabe T, Kametani K, Koyama Y-I, Suzuki D, Imamura Y, Takehana K, Hiramatsu K: **Ring-Mesh Model of Proteoglycan Glycosaminoglycan Chains in Tendon based on Three-dimensional Reconstruction by Focused Ion Beam Scanning Electron Microscopy**. *J Biol Chem* 2016, **291**:23704–23708.

AUTHOR CONTRIBUTIONS

SRB wrote the original draft of the manuscript, reviewed and edited it. SP reviewed and edited the manuscript.

FUNDING SOURCES

The work made by the authors was supported by the “Fondation pour la Recherche Médicale”, France (Grant No. DBI20141231336 to SRB), the Institut Français de Bioinformatique (ANR-11-INBS-0013), Glycomatrix project, call 2015 to SRB and SP and the Groupement de Recherche (GDR) GagoSciences (CNRS, GDR 3739, Structure, Fonction et Régulation des Glycosaminoglycans to SRB, and SP) and by GlycoAlps (Université Grenoble Alpes) to SP.

FIGURE LEGENDS

Figure 1: The roadmap designed to generate and analyze glycosaminoglycan interactomes

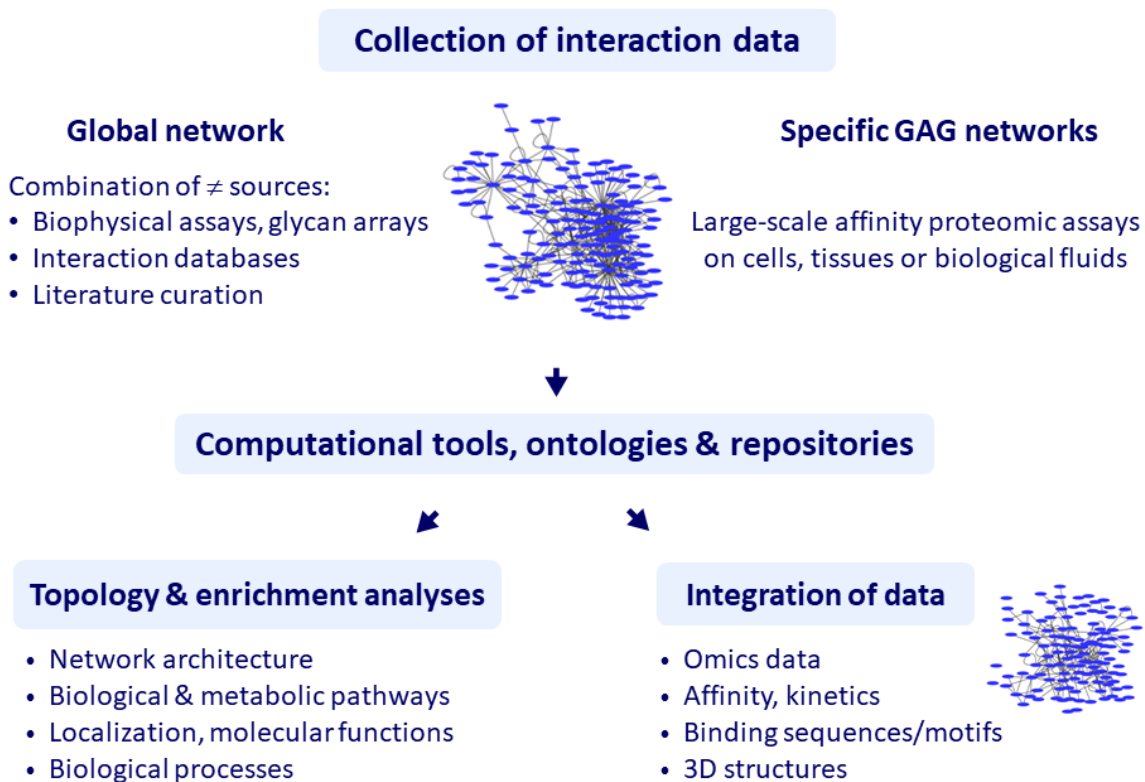


Table 1: Synthetic and natural GAG libraries recently developed (dp = degree of polymerization, *i.e.*, the number of monosaccharides, CHO: Chinese Hamster Ovary).

GAGs	GAG size and modifications	Applications	Proteins	References
47 dp4	Different backbones and chemical modifications (e.g., O- and N-sulfation)	Arrays	FGF2, CCL and CXCL chemokines	[28]
27 dp6	With and without O-sulfate groups	Arrays	Heparin lyases	[27]
HS 93 oligosaccharides (dp2, dp4, dp6, dp8)	Various sequences and sulfation patterns	Arrays	SARS-CoV-2 spike protein	[29]
CS dp6		Arrays		[25]
20 natural GAGs, including CSA, CSC, CSD, DS, HA, CSE, HP, HS, heparosan	Various Mw, and HP sulfation pattern conjugated with BSA	Arrays	SARS-CoV-2 spike protein	[30]
Chondroitin, and heparosan (<i>E. coli</i> K12 and K5 strains)	¹³ C-labeled polysaccharides	NMR spectroscopy experiments		[9]
CHO cell-based GAG libraries	Broad repertoire of GAG modifications (KO and KI of the 28 genes involved in GAG biosynthesis)	Cell-based assays Arrays Expression of recombinant PGs	VAR2CSA protein (<i>P. falciparum</i>) FGF2 CS-56 monoclonal antibody	[32]
A HS mutant mouse lung endothelial cell library	18 cell lines (systematic deletion of HS related genes)	Analysis of HS structure-function relationships	FGF2-FGFR1 signaling Anti-HS phage display antibody epitopes	[33]
63 lipid-GAGs (CSA, CSC, DS, HA, HP, HS)	Various GAG sizes	Array based on the NGL technology	Complement Factor H - Related Protein 5	[31]

Table 2 : Major GAG interaction datasets collected by different approaches. (CHO: Chinese Hamster Ovary, GO: Gene Ontology, LMW: low molecular weight)

GAGs	Biological sources	Techniques	Number of GAGBPs identified	References
CSA	Human proteins	Protein arrays	24	[13]
CSA	Nerve-related proteins (human, mouse, rat, chicken)	Protein arrays	9	[13]
HA	Human proteins	Protein arrays	6	[13]
Low sulfated polymeric HA (sHA1)	Human bone marrow stromal cells	AP-MS	125	[40]
High-sulfated polymeric hyaluronan (sHA3)	Human bone marrow stromal cells	AP-MS	126	[40]
Trideca-sulfo- HA dp6 (13sHA6-TAMRA)	Human bone marrow stromal cells	AP-MS	393	[40]
KS	Human proteins	Protein arrays	217	[13]
KS	Nerve-related proteins (human, mouse, rat, chicken)	Protein arrays	40	[13]
HP/HS	Plasma membrane enriched fractions from rat liver	Literature curation Databases GO terms UniProtKB keywords Affinity proteomics	435	[46]
HS	Human vein endothelial cells (HUVEC)	LPHAMS	32	[6]
HS	Murine brain microvascular endothelial cells	LPHAMS	12	[6]
HS	Murine lung microvascular endothelial cells	LPHAMS	9	[6]

LMW HP	Human microvascular endothelial cells	GAG-mediated pull-down proteomics	296 (269 novel)	[38]
LMW HP	Human pulmonary fibroblast	GAG-mediated pull-down proteomics	87 (75 novel)	[38]
HP	Human macrophage cell line U937	LPHAMS	20	[6]
HP	Murine pancreas	HP affinity proteomics	320 (extracellular)	[39]
HP	Murine pancreas (model of acute pancreatitis)	HP affinity proteomics	345 (extracellular)	[39]
HP	Murine plasma	HP affinity proteomics	161	[39]
HP	Murine pancreas (model of acute pancreatitis)	HP affinity proteomics	151	[39]
DS, LMW HP, HS	Human plasma	GAG-mediated pull-down proteomics	129 (71 novel)	[38]
HP	Human bone marrow stromal cells	AP-MS	43	[40]
HP (10-15 kDa)	<i>E. coli</i> proteins	Protein arrays	185	[14]
HS (10-15 kDa)	<i>E. coli</i> proteins	Protein arrays	62	[14]
DS (CSB) (10-15 kDa)	<i>E. coli</i> proteins	Protein arrays	98	[14]
CSC (10-15 kDa)	<i>E. coli</i> proteins	Protein arrays	101	[14]
IdoA-containing GAGs (DS, HP, HS)	<i>E. coli</i> proteins	Protein arrays	157	[15]
HP/HS	Angiogenesis	Literature curation	60	[53]
HP/HS	Combination of various sources and species	Literature curation	530	[51]
HP	Differentially expressed HBPs in periodontal disease	Literature curation + database	48	[50]

HP/HS	Heart diseases	[46] dataset + Open Targets Database	66 (atherosclerosis) 21 (myocarditis) 71 (myocardial infection) 34 (myocardial ischemia)	[49]
CS, DS, HA, HP/HS, KS	Combination of various sources and species	Literature curation, GO terms, UniProtKB keywords, databases, SPR, SPRi and BLI assays	827	[47]