



# INTERACTIONS BETWEEN PROTEINS AND BIOMACROMOLECULES: TOOLS AND APPLICATIONS

EDITED BY: Gunye Zhang, Lianli Chi and Fuming Zhang  
PUBLISHED IN: Frontiers in Molecular Biosciences



# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88971-121-5

DOI 10.3389/978-2-88971-121-5

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)

# INTERACTIONS BETWEEN PROTEINS AND BIOMACROMOLECULES: TOOLS AND APPLICATIONS

Topic Editors:

**Qunye Zhang**, Shandong University, China

**Lianli Chi**, Shandong University, China

**Fuming Zhang**, Rensselaer Polytechnic Institute, United States

**Citation:** Zhang, Q., Chi, L., Zhang, F., eds. (2021). Interactions Between Proteins and Biomacromolecules: Tools and Applications. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-121-5

# Table of Contents

- 04 Editorial: Interactions Between Proteins and Biomacromolecules: Tools and Applications**  
Qunye Zhang, Lianli Chi and Fuming Zhang
- 06 Structural Features of Heparin and Its Interactions With Cellular Prion Protein Measured by Surface Plasmon Resonance**  
So Young Kim, Fuming Zhang, David A. Harris and Robert J. Linhardt
- 15 Delayed Comparison and Apriori Algorithm (DCAA): A Tool for Discovering Protein–Protein Interactions From Time-Series Phosphoproteomic Data**  
Lianhong Ding, Shaoshuai Xie, Shucui Zhang, Hangyu Shen, Huaqiang Zhong Daoyuan Li , Peng Shi, Lianli Chi and Qunye Zhang
- 27 RNA-Binding Protein MSI2 Binds to miR-301a-3p and Facilitates Its Distribution in Mitochondria of Endothelial Cells**  
Qian Qian Guo, Jing Gao, Xiao Wei Wang, Xian Lun Yin, Shu Cui Zhang, Xue Li, Lian Li Chi, Xiao Ming Zhou, Zhe Wang and Qun Ye Zhang
- 38 Elucidating the Interactions Between Heparin/Heparan Sulfate and SARS-CoV-2-Related Proteins—An Important Strategy for Developing Novel Therapeutics for the COVID-19 Pandemic**  
Mingjia Yu, Tianji Zhang, Wei Zhang, Qianyun Sun, Hongmei Li and Jin-ping Li
- 51 Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease**  
Deling Shi, Anran Sheng and Lianli Chi
- 66 NMR Characterization of the Interactions Between Glycosaminoglycans and Proteins**  
Changkai Bu and Lan Jin
- 86 Macrophages Inhibit Ciliary Protein Levels by Secreting BMP-2 Leading to Airway Epithelial Remodeling Under Cigarette Smoke Exposure**  
Zhigang Wang, Wenzhang Liang, Cuiqing Ma, Jiachao Wang, Xue Gao and Lin Wei
- 101 Terminal Epitope-Dependent Branch Preference of Siglecs Toward N-Glycans**  
Shuaishuai Wang, Congcong Chen, Minhui Guan, Ding Liu, Xiu-Feng Wan and Lei Li
- 110 The Sulfation Code of Tauopathies: Heparan Sulfate Proteoglycans in the Prion Like Spread of Tau Pathology**  
Dylan Mah, Jing Zhao, Xinyue Liu, Fuming Zhang, Jian Liu, Lianchun Wang, Robert Linhardt and Chunyu Wang
- 122 Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer**  
Jingwen Yue, Weihua Jin, Hua Yang, John Faulkner, Xuehong Song, Hong Qiu, Michael Teng, Parastoo Azadi, Fuming Zhang, Robert J. Linhardt and Lianchun Wang





# Editorial: Interactions Between Proteins and Biomacromolecules: Tools and Applications

Qunye Zhang<sup>1\*</sup>, Lianli Chi<sup>2\*</sup> and Fuming Zhang<sup>3\*</sup>

<sup>1</sup>The Key Laboratory of Cardiovascular Remodeling and Function Research, Chinese Ministry of Education, Chinese National Health Commission and Chinese Academy of Medical Sciences, The State and Shandong Province Joint Key Laboratory of Translational Cardiovascular Medicine, Department of Cardiology, Qilu Hospital, Shandong University, Jinan, China, <sup>2</sup>National Glycoengineering Research Center, Shandong University, Qingdao, China, <sup>3</sup>Rensselaer Polytechnic Institute, Troy, NY, United States

**Keywords:** interaction, biomacromolecule, protein, glycosaminoglycan, SARS-CoV-2

## Editorial on the Research Topic

### Interactions Between Proteins and Biomacromolecules: Tools and Applications

## INTRODUCTION

As the protagonist in the symphony of life, proteins perform fundamental functions by interacting with other biomacromolecules including nucleic acids, and carbohydrates. Therefore, the interactions between proteins and biomacromolecules have always been the pivotal issue in biomedical research. In the past 2 decades, advances in technologies including mass spectrometry (MS), next-generation DNA sequencing, and bioinformatics give birth to a new field: interactomics (Luck et al., 2017).

In this Special Issue, we selected a series of articles that highlight technological advances for studying interactomics and the interactomics of protein-glycosaminoglycans (GAGs) and protein-miRNA. We hope that this Special Issue will instigate novel questions in the minds of our readers and will be helpful in facilitating the development of the field.

## TECHNOLOGICAL ADVANCES FOR STUDYING INTERACTOMICS

Characterization of interactions between proteins and biomacromolecules includes determining their selectivity or measuring their binding affinity. Selectivity is primarily screened by affinity chromatography, while binding affinity can be analyzed by a variety of technologies, such as isothermal titration calorimetry, surface plasmon resonance (SPR) and bio-layer interferometry. The advances of structural characterization technologies, such as MS and nuclear magnetic resonance (NMR) spectroscopy, make it possible to directly analyze the complexes formed by proteins and biomacromolecules. In this Special Issue, the technological advances for studying interactomics has been reviewed by Shi et al. A more specific review, focused on NMR characterization of GAGs and proteins, was provided by Bu and Jin. With the rapid accumulation of omics data, how to fully utilize these high-throughput data to uncover interactions between proteins and biomacromolecules has become a highly attractive research area (Hawe et al., 2019). Ding et al. presented a tool based on delayed comparison and Apriori algorithm to find protein-protein interactions according to the temporal information in time-series proteomic data. This tool is instrumental in fully utilizing the time-series omics data.

## OPEN ACCESS

### Edited and reviewed by:

Ray Luo,  
University of California, United States

### \*Correspondence:

Qunye Zhang  
wz.zhangqy@sdu.edu.cn  
Lianli Chi  
lianlich@sdnu.edu.cn  
Fuming Zhang  
zhangf2@rpi.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 11 May 2021

**Accepted:** 31 May 2021

**Published:** 11 June 2021

### Citation:

Zhang Q, Chi L and Zhang F (2021)  
Editorial: Interactions Between  
Proteins and Biomacromolecules:  
Tools and Applications.  
Front. Mol. Biosci. 8:708084.  
doi: 10.3389/fmolb.2021.708084

## INTERACTOMICS OF PROTEIN-GAG AND PROTEIN-MIRNA

GAGs are a family of highly negatively charged linear polysaccharides including heparin/heparan sulfate (HS), chondroitin sulfate (CS)/dermatan sulfate (DS), and keratan sulfate (KS). GAGs play vital roles in many pathological and physiological processes, such as embryonic development, extracellular matrix assembly, inflammation, cancer, and cardiovascular diseases, by interacting with numerous proteins. The study of protein-GAG interactions is an important theme in glycobiology, resulting in many therapeutic implications. Recently, the first comprehensive draft of GAG interactome was reported, which composes of 932 protein-GAG interactions (Vallet et al., 2021).

In this Special Issue, Shi et al. present an excellent review on the progress of protein-GAG interactions research. Bu and Jin review the application of NMR spectroscopy on the characterization of protein-GAG interactions. Since the COVID-19 pandemic, it was reported that the interactions between cellular HS and S-protein of SARS-CoV-2 are critical for the viral infection (Kim et al., 2020 and Clausen et al., 2020). Yue et al. review the latest advances in HS-protein interactions studies related to COVID-19. In a research article, Yue et al. confirmed HS facilitates S-protein-mediated SARS-CoV-2 host cell invasion. In Alzheimer's disease (AD) research, protein Tau (tau) and related tau pathology have been a hot research area since the interactions between tau and HS proteoglycans (HSPGs) are key facilitator in each stage of the prion-like propagation of pathology. Mah et al. review the sulfation code of HSPGs in tauopathies. Additionally, Kim et al. report the kinetics and structural features of heparin and its interactions with cellular prion protein measured by SPR.

## REFERENCES

- Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl Painter, J. C. D., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* 183, 1043–1057. doi:10.1016/j.cell.2020.09.033
- Ebert, M. S., and Sharp, P. A. (2012). Roles for microRNAs in Conferring Robustness to Biological Processes. *Cell* 149, 515–524. doi:10.1016/j.cell.2012.04.005
- Hawe, J. S., Theis, F. J., and Heinig, M. (2019). Inferring Interaction Networks from Multi-Omics Data. *Front. Genet.* 10, 535. doi:10.3389/fgene.2019.00535
- Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., et al. (2020). Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. *Antiviral Res.* 181, 104873. doi:10.1016/j.antiviral.2020.104873

MiRNAs are a class of RNA molecules with important regulatory functions (Ebert et al., 2012), but there are few studies on protein-miRNA interactions, especially about miRNA transport, distribution in organelles and secretion. In this Special Issue, Guo et al. discovered that MSI2 could bind to miR-301a-3p and facilitate its distribution in mitochondria using affinity purification and non-labeling proteomic techniques. This study provided valuable insight into the mechanism of mitochondrial distribution of miRNAs.

## FUTURE PERSPECTIVES

This collection of articles highlights the different approaches for studying interactomics and related new analytical techniques. Indeed, the breakthrough in analytical tools (ultra-highly sensitive and automatic level) and approaches (such as machine learning) during the last decade has facilitated a greater understanding of the importance of interactions between biomacromolecules and their roles in diseases. We expect the more outcomes from this interdisciplinary research will accelerate the identification of new biomarkers for diagnostics and drug discovery.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Luck, K., Sheynkman, G. M., Zhang, I., and Vidal, M. (2017). Proteome-Scale Human Interactomics. *Trends Biochemical Sciences* 42, 342–354. doi:10.1016/j.tibs.2017.02.006

Vallet, S. D., Clerc, O., and Ricard-Blum, S. (2021). Glycosaminoglycan-Protein Interactions: The First Draft of the Glycosaminoglycan Interactome. *J. Histochem. Cytochem.* 69, 93–104. doi:10.1369/0022155420946403

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhang, Chi and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Structural Features of Heparin and Its Interactions With Cellular Prion Protein Measured by Surface Plasmon Resonance

So Young Kim<sup>1,2</sup>, Fuming Zhang<sup>3\*</sup>, David A. Harris<sup>4</sup> and Robert J. Linhardt<sup>3,5\*</sup>

<sup>1</sup> Division of Pulmonary and Critical Care, Department of Medicine, University of California, San Diego, San Diego, CA, United States, <sup>2</sup> VA San Diego Healthcare System, Medical and Research Sections, San Diego, CA, United States, <sup>3</sup> Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, United States, <sup>4</sup> Department of Biochemistry, Boston University School of Medicine, Boston, MA, United States, <sup>5</sup> Department of Chemistry and Chemical Biology, Biological Science and Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, United States

## OPEN ACCESS

### Edited by:

Matthias Buck,  
Case Western Reserve University,  
United States

### Reviewed by:

Kurt Henry Piepenbrink,  
University of Nebraska-Lincoln,  
United States  
Andrew Benjamin Herr,  
Cincinnati Children's Hospital Medical  
Center, United States

### \*Correspondence:

Fuming Zhang  
zhangf2@rpi.edu  
Robert J. Linhardt  
linhar@rpi.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 13 August 2020

**Accepted:** 14 October 2020

**Published:** 26 November 2020

### Citation:

Kim SY, Zhang F, Harris DA and  
Linhardt RJ (2020) Structural Features  
of Heparin and Its Interactions With  
Cellular Prion Protein Measured by  
Surface Plasmon Resonance.  
Front. Mol. Biosci. 7:594497.  
doi: 10.3389/fmolb.2020.594497

Self-propagating form of the prion protein (PrP<sup>Sc</sup>) causes many neurodegenerative diseases, such as Creutzfeldt-Jakob disease (CJD) and Gerstmann-Straussler-Scheinker syndrome (GSS). Heparin is a highly sulfated linear glycosaminoglycan (GAG) and is composed of alternating D-glucosamine and L-iduronic acid or D-glucuronic acid sugar residues. The interactions of heparin with various proteins in a domain-specific or charged-dependent manner provide key roles on many physiological and pathological processes. While GAG-PrP interactions had been previously reported, the specific glycan structures that facilitate interactions with different regions of PrP and their binding kinetics have not been systematically investigated. In this study, we performed direct binding surface plasmon resonance (SPR) assay to characterize the kinetics of heparin binding to four recombinant murine PrP constructs including full length (M23–230), a deletion mutant lacking the four histidine-containing octapeptide repeats (M23–230 Δ59–90), the isolated N-terminal domain (M23–109), and the isolated C-terminal domain (M90–230). Additionally, we found the specific structural determinants required for GAG binding to the four PrP constructs with chemically defined derivatives of heparin and other GAGs by an SPR competition assay. Our findings may be instrumental in developing designer GAGs for specific targets within the PrP to fine-tune biological and pathophysiological activities of PrP.

**Keywords:** heparin, interaction, prion protein, surface plasmon resonance, glycosaminoglycan

## INTRODUCTION

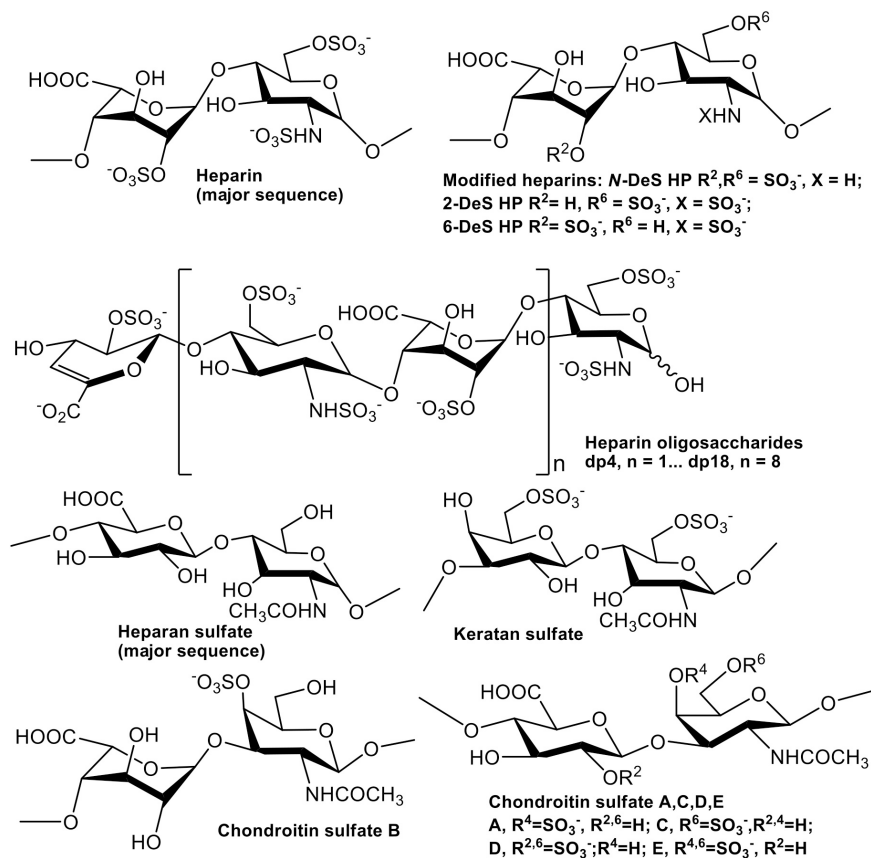
A group of neurodegenerative diseases, including Creutzfeldt-Jakob disease (CJD) and Gerstmann-Straussler-Scheinker syndrome (GSS), are caused by an infectious, self-propagating form of the prion protein, PrP<sup>Sc</sup> (Mercer et al., 2018). PrP<sup>Sc</sup> interacts with a normal, glycoposphatidylinositol (GPI) anchored cellular conformer, PrP<sup>C</sup>, on the neuronal surface and induces a conformational

change in PrP<sup>C</sup>, which leads, through an autocatalytic process, to accumulation of protease-resistant PrP<sup>Sc</sup> in the brain. As part of this process, PrP<sup>Sc</sup> also activates a PrP<sup>C</sup>-dependent signal transduction pathway that results in neurotoxicity (Le et al., 2019). This toxic pathway depends critically on the N-terminal domain of PrP<sup>C</sup> (Solomon et al., 2011; Westergard et al., 2011; Wu et al., 2017). Neurotoxicity can be prevented through interactions between N-terminal region of PrP<sup>C</sup> and several ligands, including sulfated glycosaminoglycans (GAGs) and copper ions (Pan et al., 2002; Warner et al., 2002). The PrP<sup>C</sup> molecule is comprised of two major structural domains: a flexible, natively unstructured N-terminal domain (residues 23–127), and a structured, globular C-terminal domain (residues 128–230) (Zahn et al., 2000).

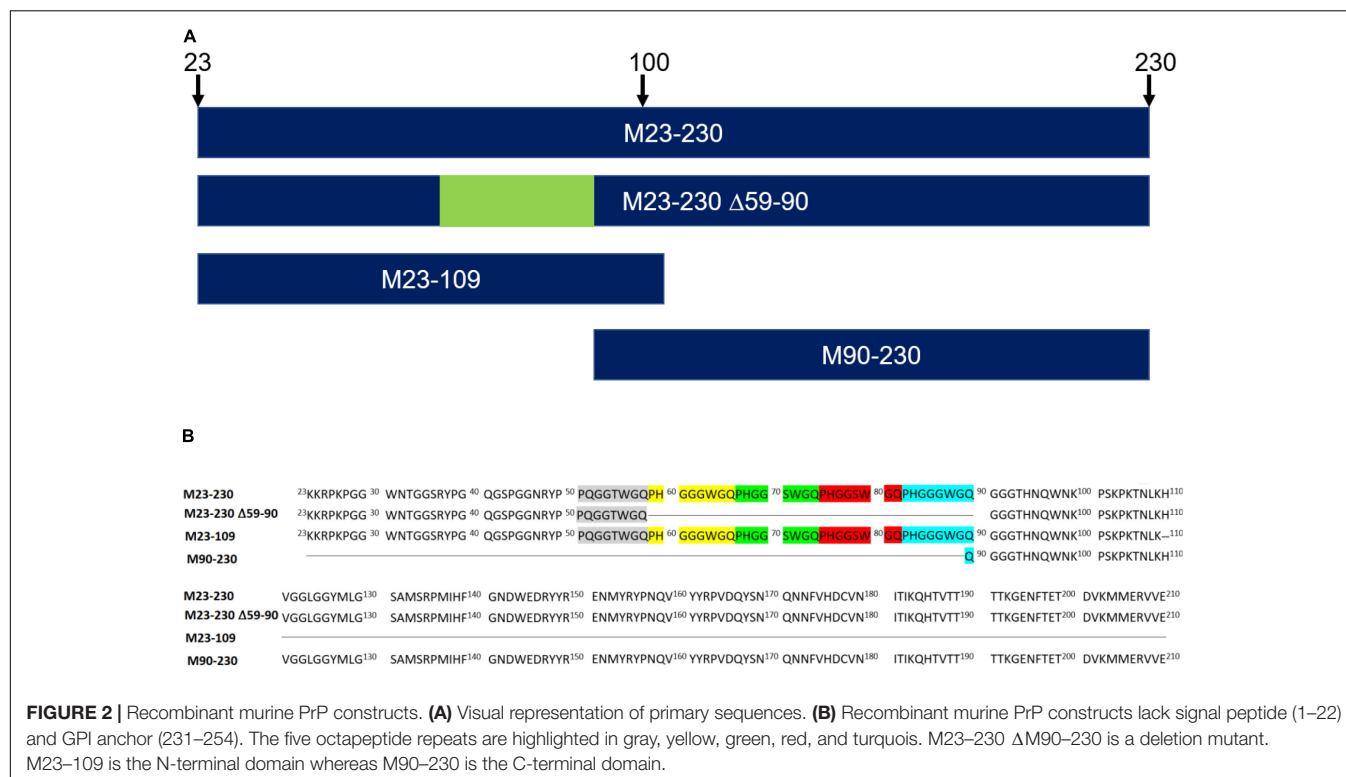
GAGs are anionic linear polysaccharides comprised of repeating disaccharide units (**Figure 1**) that are generally found covalently attached to core proteins as proteoglycans (PGs) (Linhardt and Toida, 2004). GAGs in both intracellular and extracellular spaces participate in biological processes such as cellular communication, as well as in the pathogenesis of diseases (Linhardt and Toida, 2004; Kim et al., 2018, 2020). GAG-PrP interactions had been previously reported using human, bovine, and murine PrP (Warner et al., 2002; Andrievskaia et al., 2007;

Veira et al., 2011). Three regions of PrP were identified as sufficient for binding of heparin (HP) and heparan sulfate (HS), including residues 23–52, 53–93, and 110–128 (Pan et al., 2002; Warner et al., 2002). GAGs also regulate the cellular localization of PrP<sup>C</sup> (Shyng et al., 1995; Taylor et al., 2009) and inhibit formation of PrP<sup>Sc</sup> in cells and in animal models (Caughey and Raymond, 1993; Doh-ura et al., 2004). Despite the importance of GAGs in prion biology, the specific glycan structures that interact with different regions of PrP, and the kinetics of these interactions, have not been systematically investigated. This information is important for understanding the normal function of PrP<sup>C</sup>, its transformation into PrP<sup>Sc</sup>, and how the latter process could be inhibited for therapeutic effect.

In this study, we employed direct binding surface plasmon resonance (SPR) assay to characterize the kinetics of heparin binding to four recombinant murine PrP constructs (**Figures 2A,B**): (1) full length (M23-230), (2) a deletion mutant lacking the four histidine-containing octapeptide repeats (M23-230  $\Delta$ 59-90), (3) the isolated N-terminal domain (M23-109), and (4) the isolated C-terminal domain (M90-230). Additionally, we identified the specific structural determinants required for GAG binding to the four PrP constructs using an SPR competition assay with chemically defined derivatives of heparin and other GAGs.



**FIGURE 1** | Chemical structures of heparin and heparin-derived oligosaccharides and GAGs.



## MATERIALS AND METHODS

### Recombinant PrP

Four different mouse PrP (moPrP) constructs (**Figure 2**) were prepared by Prof. David Harris's Group (Boston University). These included 1 mg of full length construct: M23–230 ( $MW = 23,061$  Da); 1 mg of the N-terminal domain: M23–109 ( $MW = 9,142$  Da); 0.5 mg of delta OR: M23–230 Δ59–90 ( $MW = 19,763$  Da); and 2.5 mg of the C-terminal domain: M90–230 ( $MW = 16,013$  Da). These proteins were prepared in *E. coli*, and purified as previously described, and characterized by SDS-PAGE and circular dichroism (CD) (Wu et al., 2017; McDonald et al., 2019).

### Glycosaminoglycans

The GAGs used were porcine intestinal heparin (16 kDa), low molecular weight heparin (LMWH) (5 kDa, enoxaparin, Sanofi-Aventis) and porcine intestinal heparan sulfate (12 kDa, Celsus Laboratories, Cincinnati, OH); chondroitin sulfate A (CSA, 20 kDa) from porcine rib cartilage (Sigma, St. Louis, MO), dermatan sulfate (also known as chondroitin sulfate B, CSB, 30 kDa, from porcine intestine, Sigma), dermatan disulfate (4,6-disulfo DS, 33 kDa, Celsus) prepared through the chemical 6-*O*-sulfonation of dermatan sulfate (Van Gorp et al., 1999), chondroitin sulfate C (CSC, 20 kDa, from shark cartilage, Sigma), chondroitin sulfate D (CSD, 20 kDa, from whale cartilage, Seikagaku, Tokyo, Japan), chondroitin sulfate E (CSE, 20 kDa from squid cartilage, Seikagaku) and keratan sulfate (KS, 14.3 kDa) was isolated from bovine cornea in Linhardt Lab

(Weyers et al., 2013). *N*-desulfated heparin (14 kDa) and 2-*O*-desulfated IdoA heparin (13 kDa) were all prepared based on Yates et al. (1996). 6-*O*-desulfated heparin (13 kDa) was kindly provided by Prof. Lianchun Wang from University of South Florida. Heparin oligosaccharides included tetrasaccharide (dp4), hexasaccharide (dp6), octasaccharide (dp8), decasaccharide (dp10), dodecasaccharide (dp12), tetradecasaccharide (dp14), hexadecasaccharide (dp16), and octadecasaccharide (dp18) and were prepared from controlled partial heparin lyase 1 treatment of bovine lung heparin (Sigma) followed by size fractionation. The chemical structures of these GAGs are shown in **Figure 1**.

### Preparation of Heparin Biochip

BIAcore 3000 SPR instrument and sensor SA chips were from GE Healthcare (Uppsala, Sweden). SPR measurements were performed on a BIAcore 3000 operated using BIAcore 3000 control and BIAevaluation software (version 4.0.1). Heparin was biotinylated by our previous protocol with minor modification (Kim et al., 2018). Heparin (2 mg) and 2 mg of amine-PEG<sub>3</sub>-Biotin (Thermo Fisher Scientific, Waltham, MA) were mixed with 10 mg of NaCNBH<sub>3</sub> in 200 μL of H<sub>2</sub>O for the initial reaction, which was performed at 70°C for 24 h, and then a further 10 mg of NaCNBH<sub>3</sub> was added to continue the reaction for another 24 h. Upon completion of this reaction, the mixture was desalted with a spin column (3000 molecular weight cut-off). The biotinylated heparin was immobilized to streptavidin (SA) chip based on the manufacturer's protocol. In brief, 20 μL solution of the heparin-biotin conjugate (0.1 mg/mL) in HBS-EP running buffer was injected over flow cell 2 (FC2) of the SA chip at a flow rate of 10 μL/min. The successful immobilization of heparin was confirmed



by the observation of a  $\sim 200$  resonance unit (RU) increase in the sensor chip. The control flow cell (FC1) was prepared by 1 min injection with saturated biotin.

## Measurement of Interaction Between Heparin and Prp Using Biacore

The PrP samples were diluted in HBS-EP buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% surfactant P20, pH 7.4). Different dilutions of PrP samples were injected at a flow rate of 30  $\mu\text{L}/\text{min}$ . At the end of the sample injection, the same buffer was flowed over the sensor surface to facilitate dissociation. After a 3 min dissociation time, the sensor surface was regenerated by injecting with 30  $\mu\text{L}$  of 2 M NaCl to get fully regenerated surface. The response was monitored as a function of time (sensorgram) at 25°C.

## Solution Competition Study Between Heparin on Chip Surface and Heparin-Derived Oligosaccharides in Solution Using SPR

PrP (63 or 125 nM) mixed with 1,000 nM of heparin oligosaccharides, including dp4, dp6, dp8, dp10, dp12, dp14, dp16, and dp18 in HBS-EP buffer were injected over heparin chip at a flow rate of 30  $\mu\text{L}/\text{min}$ , respectively. After each run, the dissociation and the regeneration were performed as described above. For each set of competition experiments on SPR, a control experiment (only protein without any heparin or oligosaccharides) was performed to make sure the surface was completely regenerated and that the results obtained between runs were comparable. Statistical analysis was conducted using a student's *t*-test.

## Solution Competition Study Between Heparin on Chip Surface and GAGs, Chemical Modified Heparin in Solution Using SPR

For testing of inhibition by other GAGs and chemical modified heparins of the PrP-heparin interaction, PrP at 63 or 125 nM was pre-mixed with 1,000 nM of GAG or chemical modified heparin and injected over the heparin chip at a flow-rate of 30  $\mu\text{L}/\text{min}$ . After each run, a dissociation period and regeneration protocol was performed as described above. Statistical analysis was conducted using a student's *t*-test.

## RESULTS

### Kinetics Measurements of Prp-Heparin Interactions

Kinetic curves calculated from sensorgrams fitted to a 1:1 Langmuir model from BIAevaluate 4.0.1 demonstrate binding affinity,  $K_D$ , values in the following order: full length PrP ( $1.1 \times 10^{-7}$  M), M23–109 PrP ( $K_D = 7.1 \times 10^{-7}$  M), and M23–230  $\Delta 59$ –90 PrP ( $3.3 \times 10^{-6}$  M) shown in **Table 1** and **Figures 3A–C**. M90–230 PrP showed negligible binding

to heparin even at the highest concentration used in this direct binding assay for all samples, 500 nM (**Figure 3D**). Similarly, association rate constant ( $k_a$ ) was greatest for full length PrP ( $1.6 \times 10^5$ ) followed by M23–109 PrP ( $3.2 \times 10^4$ ) and M23–230  $\Delta 59$ –90 PrP ( $2.1 \times 10^4$ ) (**Table 1** and **Figures 3A–C**). Of the previously identified, putative heparin binding motifs (23–52, 53–93, and 110–128) (Pan et al., 2002; Warner et al., 2002), 23–52 and 53–93 appear to contribute to binding affinity most significantly. The smaller and flexible conformation of M23–109 PrP may additionally facilitate tighter binding to heparin. For example, M23–109 PrP may be able to bind sub-populations of immobilized heparin (i.e., shorter chain length heparin) that full length PrP does not.

## Solution Competition Study on the Interactions Between the Immobilized Heparin With Prp Constructs to Heparin-Derived Oligosaccharides Using SPR

Solution/surface competition experiments were performed by SPR to examine the effect of the chain length of heparin on the heparin-PrP interactions. Different chain length heparin-derived oligosaccharides (from dp4 to dp18) at 1,000 nM were used in the competition study. LMWH ( $\sim 5$  kDa) and unfractionated heparin (12–15 kDa) at 1,000 nM were also tested for their ability to inhibit PrP-heparin interactions.

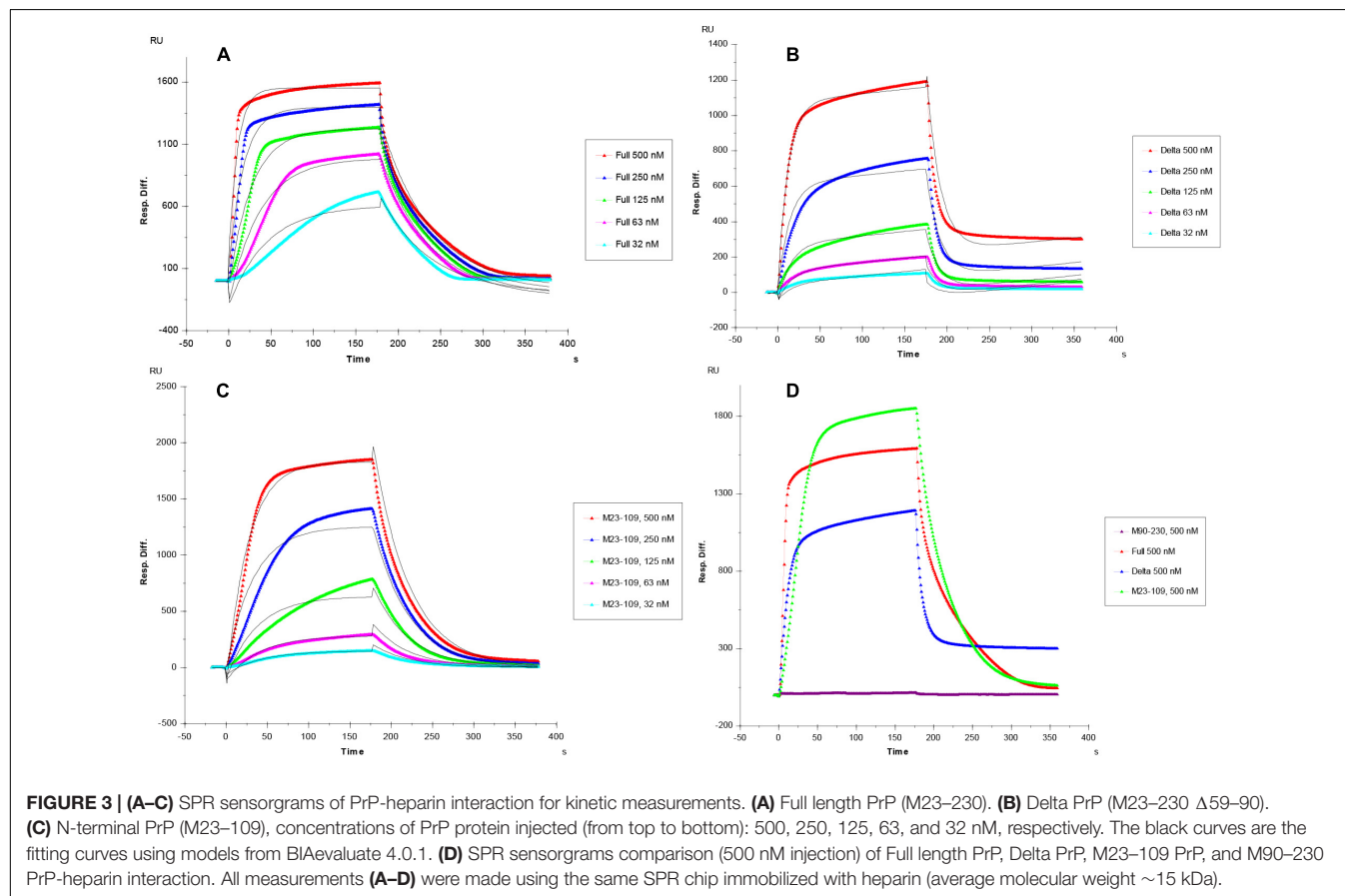
For the full length PrP, inhibition effects of heparin oligosaccharides, LMWH, and unfractionated heparin were chain-length-dependent (**Figure 4A**). Negligible competition was observed when 1,000 nM of oligosaccharides (dp 4 to dp 16) present in the full length PrP protein solution (**Figure 4A**). The longer chain length heparin oligosaccharide, dp18, however, inhibited the binding of full length PrP to the surface heparin by 40% (**Figure 4A**). LMWH and unfractionated heparin inhibited PrP-heparin interactions more effectively, by 60 and 80%, respectively (**Figure 4A**). These results demonstrate that full length PrP prefers bindings to longer heparin chains.

Similarly, longer chain length heparin and heparin oligosaccharides inhibited M23–230  $\Delta 59$ –90 PrP and heparin interactions more effectively (**Figure 4B**). However, the percent inhibition was greater for all compounds tested. Dp4-dp14 provided  $\sim 40\%$  inhibition to M23–230  $\Delta 59$ –90 PrP (**Figure 4B**), which could be reached starting at dp18 for the full length PrP (**Figure 4A**). Unfractionated heparin inhibited M23–230  $\Delta 59$ –90 PrP and heparin interactions by 90% (**Figure 4B**). M23–230  $\Delta 59$ –90 PrP has the same primary amino acid sequence as full length PrP except for deletion of majority of the octapeptide repeats (59–90) and lacks one putative heparin binding motif at residues 53–93 (**Figure 2**). This lack of this binding motif and the potential alteration on the three-dimensional structure/conformation in the absence of residues 59–90 may have weakened the interactions between delta PrP and immobilized heparin surface allowing greater inhibition by heparin and heparin oligosaccharides at same concentration.

**TABLE 1** | Summary of kinetic data of PrP protein- heparin interactions\*.

Interaction	$k_a$ (1/MS)	$k_d$ (1/S)	$K_D$ (M)
PrP Full length/Heparin	$1.6 \times 10^5 (\pm 4.1 \times 10^3)$	$0.017 (\pm 3.0 \times 10^{-4})$	$1.1 \times 10^{-7}$
PrP M23–230 $\Delta$ 59–90/Heparin	$2.1 \times 10^4 (\pm 1.5 \times 10^3)$	$0.069 (\pm 9.9 \times 10^{-4})$	$3.3 \times 10^{-6}$
PrP M23–109/Heparin	$3.2 \times 10^4 (\pm 996)$	$0.023 (\pm 2.3 \times 10^{-4})$	$7.1 \times 10^{-7}$

\*The data with ( $\pm$ ) in parentheses are the standard deviations (SD) from global fitting of five injections.

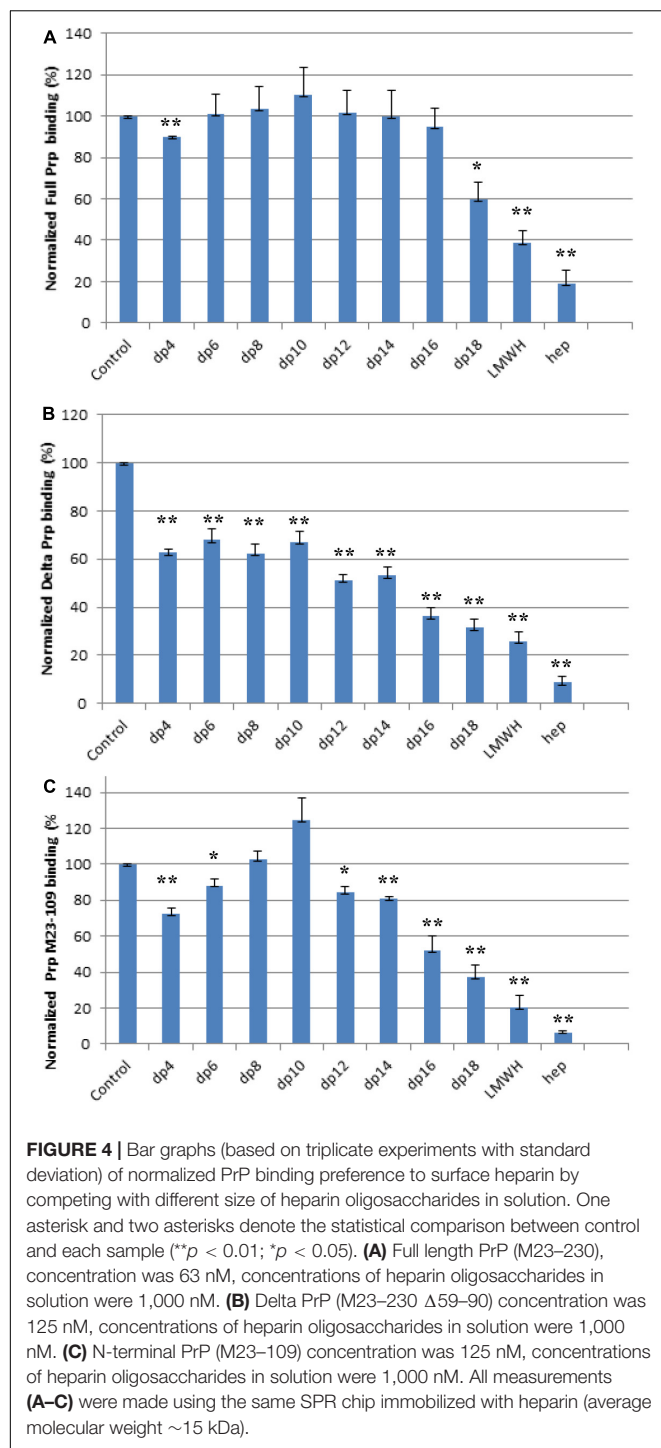


The N-terminal domain, M23–109 PrP demonstrated different mode of inhibition by heparin and heparin oligosaccharides (**Figure 4C**). dp4 inhibits M23–109 PrP and heparin interactions by  $\sim 20\%$ . However, this inhibition decreases with chain length up to dp 10, with the latter actually causing increased binding to surface heparin. From dp 12 to unfractionated heparin, the inhibition increases in a chain-length-dependent fashion. The N-terminal domain has two putative heparin binding motifs in 23–52 and 53–93 (Pan et al., 2002; Warner et al., 2002) and the 3-D conformation/folding may be altered from that of the full length PrP allowing heparin binding differently. For example, some of these regions may be exposed to the surface to more readily interact with shorter length heparin oligosaccharide, dp4. There is evidence that the N-terminal domain of PrP<sup>C</sup> physically interacts with the C-terminal domain (McDonald et al., 2019) and the absence of this interaction in M23–109 might also influence the heparin binding characteristics of the latter protein. Finally, it is possible

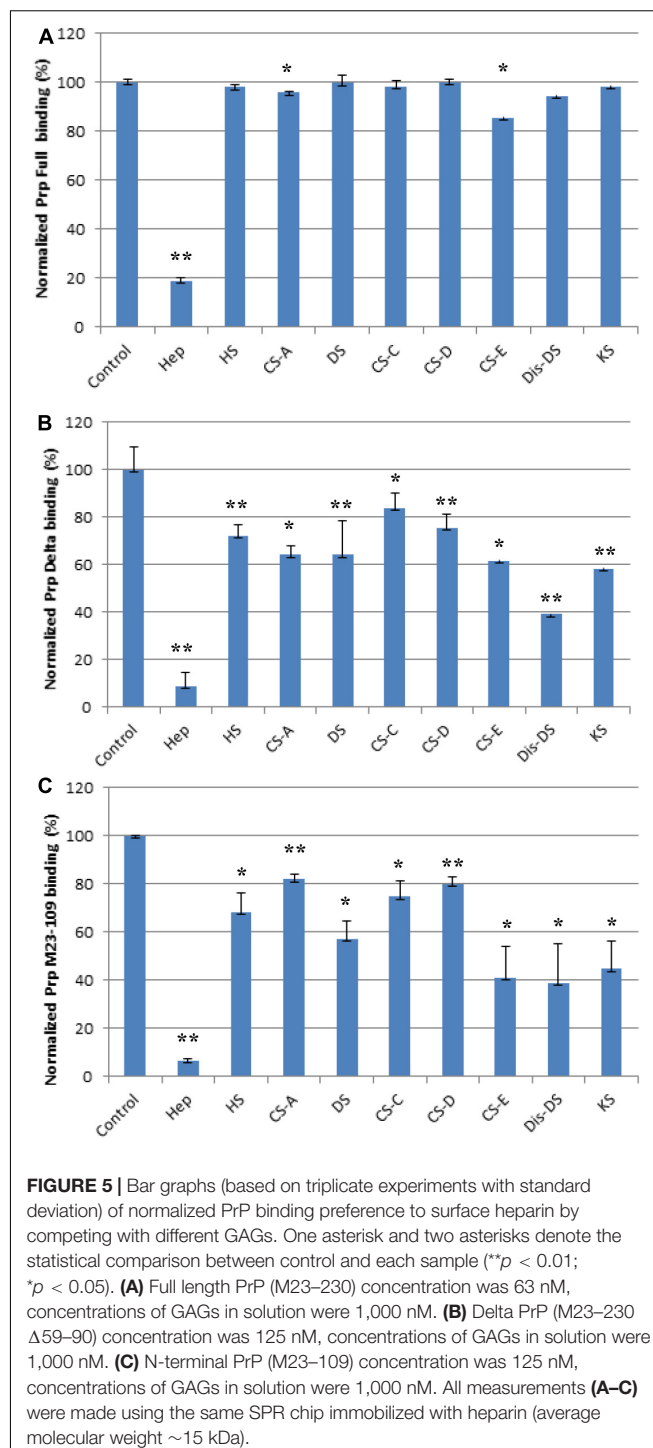
that shorter oligosaccharides (up to dp10) actually stabilize the structure of the N-terminal domain in such a way as to increase binding to surface heparin in the SPR experiments.

## SPR Solution Competition Study of Various GAGs

We screened inhibition capability of GAGs of different structures (**Figure 1**), including unfractionated heparin, HS, chondroitin sulfate type A (CS-A), CS-C, CS-D, CS-E, DS, disulfated DS (Dis-DS), and keratan sulfate (KS), against interactions between PrP constructs and immobilized heparin (**Figure 5**). All GAGs tested were used at 1,000 nM. For full length PrP, only unfractionated heparin was capable of inhibiting PrP-heparin interactions by 80% while the rest of GAGs showed negligible inhibition (**Figure 5A**). Unfractionated heparin inhibited M23–230  $\Delta$ 59–90 PrP and heparin interactions by  $\sim 90\%$  and varying degree of inhibition was observed by other GAGs ranging from 20 to



60% inhibition (Figure 5B). This reinforces the idea of weakened binding interaction to immobilized heparin due to lack of one putative heparin binding motif and potential change in 3-D structure as described above. Lastly, inhibition ranging from 20 to 90% was demonstrated by various GAGs for inhibiting M23–109 PrP-heparin binding (Figure 5C), however, the preferred structure of GAG was different from those of full length or M23–230  $\Delta$ 59–90 PrP, suggesting a different mode of binding then



was observed in competition assays utilizing varying chain length heparin oligosaccharides (Figure 4C).

## SPR Solution Competition Study of Chemically Modified Heparin Derivatives

Next, we determined if *N*-, 2-*O*-, 3-*O*-, and 6-*O*-sulfation on heparin were required for efficient binding to PrP constructs



using chemically modified heparin derivatives. Of these heparin derivatives, only 2-DeS hep inhibited full length PrP and heparin interactions by 20% (**Figure 6A**). *N*- and 6-*O* desulfated heparin derivatives, however, did not inhibit PrP and heparin interactions (**Figure 6A**). Unfractionated heparin has an additional 3-*O* sulfation, which may be responsible for forming electrostatic interactions with surface accessible basic residues on the putative heparin binding motifs on the full length PrP. For both M23–230  $\Delta$ 59–90 PrP and M23–109 PrP, however, all of the heparin derivatives inhibited PrP and heparin interactions

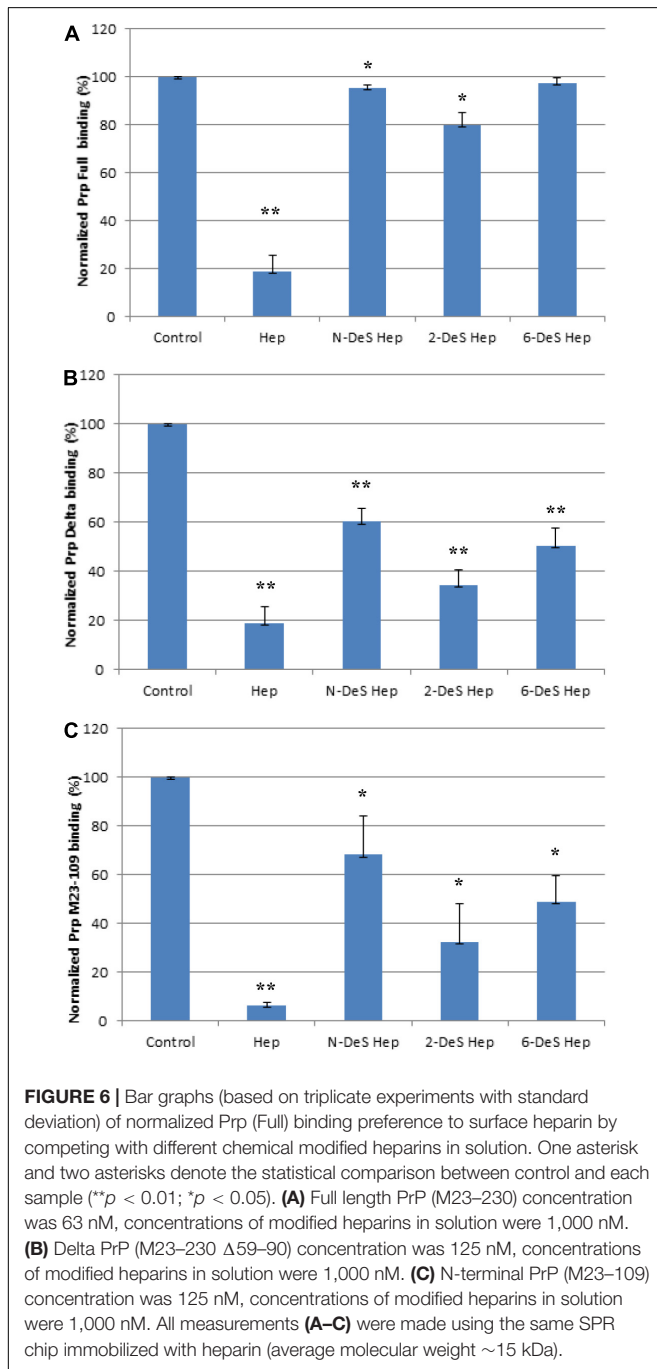
(**Figures 6B,C**). These findings further suggest the importance of presence of all three putative heparin binding motifs, which also allow native conformation of full length PrP, for efficient binding to heparin.

## DISCUSSION

Our investigation shows that full length PrP (M23–230) binds heparin with greatest binding affinity ( $K_D = 0.11 \mu\text{M}$ ) followed by the N-terminus region PrP (M23–109) ( $K_D = 0.71 \mu\text{M}$ ), and mutant PrP (M23–230  $\Delta$ 59–90) ( $K_D = 3.3 \mu\text{M}$ ) (**Figure 3** and **Table 1**). The C-terminus region PrP (M90–230) exhibited negligible binding. Comparable binding affinities between full length PrP and M23–109 PrP confirm that the major heparin binding sites are localized within the N-terminal region (23–52) and the octapeptide repeats (53–90), but not in the C-terminal region (Pan et al., 2002; Warner et al., 2002). The polybasic amino acid segment (residues 23–31) within the putative heparin binding motif (23–52) modulates ion channel activity of PrP, perhaps modulated by GAG binding (Le et al., 2019). The region containing the four histidine-containing octapeptide repeats [PHGG(G/S)WGQ] (53–93) was previously determined to possess an additional putative heparin binding motif (Pan et al., 2002; Warner et al., 2002). Lacking this region reduced heparin binding ability (**Table 1** and **Figures 3–6**). Copper ions binding to this region on the PrP<sup>C</sup> on the neuronal cell surface results in rapid clathrin-dependent endocytosis of PrP<sup>C</sup> (Hooper et al., 2008). Heparin binding to bovine PrP<sup>C</sup> is copper dependent (Andrievskaia et al., 2007). Copper and other metal ions interact with heparin or heparan sulfate to modulate heparin binding to proteins (Zhang et al., 2014; Kim et al., 2018). Binding of sulfated GAGs and copper ions to the N-terminal domain regulates ion channel activity and other toxic effects of PrP<sup>C</sup> (Wu et al., 2017; Le et al., 2019).

In the competition SPR binding assays, we determined structural preferences of PrP binding. Both full length PrP and M23–230  $\Delta$ 59–90 PrP show similar trend of preferred binding to heparin oligosaccharides with longer chain length (**Figures 4A,B**). By lacking four of the five octapeptide repeats, the M23–230  $\Delta$ 59–90 PrP shows lower binding affinity to immobilized heparin allowing heparin oligosaccharides to inhibit this interaction by a greater extent. The isolated N-terminal domain (M23–109), however, showed a different mode of competition with heparin, where both shorter (dp4 and dp6) or longer (dp12-unfractionated heparin) chain length heparin than dp8 and dp10 exhibited greater level of inhibition (**Figure 4C**). Lacking a heparin binding motif at 110–128 as well as intramolecular interactions with the C-terminal domain appears to have altered the original mode of heparin binding. While these are interesting trends, it is also possible that the varying levels of magnitude may be due to differences in the binding affinities against heparin.

The competition assay results of screening various types of GAGs further demonstrate that the positions of basic residues in PrP are important for GAG binding, likely by determining the spatial arrangement of electrostatic interactions with carboxylate



and sulfate groups on the GAG molecules. The last set of competition assays using chemically modified heparin derivatives has suggested that 3-O sulfation is most important for heparin binding to full length PrP, whereas N-, 2-O-, 3-O-, and 6-O-sulfation appears to be important for heparin interactions with M23–230  $\Delta$ 59–90 PrP or M23–109 PrP (**Figures 6B,C**). Overall 3-D structural changes in PrP lacking heparin-binding motifs (residues 53–90 or 110–128) alter types of GAGs and sulfation patterns of heparin it preferentially binds to; and this should be considered in developing designer GAGs as PrP therapeutic. Similarly, based on the results from the competition assay using heparin oligosaccharides (**Figure 4**), we conjecture that the varying levels of inhibition may be the result of varying heparin-binding strength for three PrP constructs. In human PrP<sup>C</sup>, 2-O-sulfate groups, but not 6-O-sulfate position, are required for heparin recognition (Warner et al., 2002).

In summary, we have characterized binding interactions between four different PrP constructs [full length (M23–230), M23–230  $\Delta$ 59–90, N- and C-terminal domains] and different forms of GAGs varying in their structures. By SPR direct binding assays, we determined the kinetics of these PrP-heparin interactions, and confirmed that previously identified, putative heparin binding motifs were essential for the binding. Competition assays utilizing varying chain length of heparin and heparin oligosaccharides revealed that full length and M23–230  $\Delta$ 59–90 PrP prefer binding longer chain length heparin, while the N-terminal domain of PrP had a different mode of binding. Binding of full length PrP to heparin was effectively inhibited only by unfractionated heparin. However, M23–230  $\Delta$ 59–90 and the N-terminal domain exhibited preferential binding to various types of GAGs, with Dis-DS being the best

inhibitor for both (besides heparin). Screening of chemically modified heparin derivatives in PrP-heparin competition assays demonstrated that 3-O sulfation is critical for full length PrP and heparin binding while M23–230  $\Delta$ 59–90 and N-terminal domain require all sulfation positions. Our findings on the structural requirements for efficient binding to these PrP constructs lays the foundation for designing tailored GAG inhibitors targeting different regions within the PrP molecule. Such inhibitors may be useful for controlling the biological and pathophysiological activities of PrP.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

SK and FZ performed SPR assays, analyzed the data, wrote, and revised the manuscript. RL conceived, designed the project, and revised the manuscript. DH prepared the PrP proteins and revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was funded through grants from the NIH (R01 NS065244 to DH) and (DK111958 and CA231074 to RL).

## REFERENCES

- Andrievskaia, O., Potetinova, Z., Balachandran, A., and Nielsen, K. (2007). Binding of bovine prion protein to heparin: a fluorescence polarization study. *Arch. Biochem. Biophys.* 460, 10–16. doi: 10.1016/j.abb.2007.02.001
- Caughey, B., and Raymond, G. J. (1993). Sulfated polyanion inhibition of scrapie-associated PrP accumulation in cultured cells. *J. Virol.* 67, 643–650. doi: 10.1128/jvi.67.2.643-650.1993
- Doh-ura, K., Ishikawa, K., Murakami-Kubo, I., Sasaki, K., Mohri, S., Race, R., et al. (2004). Treatment of transmissible spongiform encephalopathy by intraventricular drug infusion in animal models. *J. Virol.* 78, 4999–5006. doi: 10.1128/jvi.78.10.4999-5006.2004
- Hooper, N. M., Taylor, D. R., and Watt, N. T. (2008). Mechanism of the metal-mediated endocytosis of the prion protein. *Biochem. Soc. Trans.* 36, 1272–1276. doi: 10.1042/BST0361272
- Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., et al. (2020). Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. *Antiviral Res.* 181:104873. doi: 10.1016/j.antiviral.2020.104873
- Kim, S. Y., Zhang, F., Gong, W., Chen, K., Xia, K., Liu, F., et al. (2018). Copper regulates the interactions of antimicrobial piscidin peptides from fish mast cells with formyl peptide receptors and heparin. *J. Biol. Chem.* 293, 15381–15396. doi: 10.1074/jbc.RA118.001904
- Le, N. T. T., Wu, B., and Harris, D. A. (2019). Prion neurotoxicity. *Brain Pathol.* 29, 263–277. doi: 10.1111/bpa.12694
- Linhardt, R. J., and Toida, T. (2004). Role of glycosaminoglycans in cellular communication. *Acc. Chem. Res.* 37, 431–438. doi: 10.1021/ar030138x
- McDonald, A. J., Leon, D. R., Markham, K. A., Wu, B., Heckendorf, C. F., Schilling, K., et al. (2019). Altered domain structure of the prion protein caused by Cu2(binding and functionally relevant mutations: analysis by cross-linking, MS/MS, and NMR. *Structure* 27, 907–922.e5. doi: 10.1016/j.str.2019.03.008
- Mercer, R. C. C., McDonald, A. J., Bove-fenderson, E., Fang, C., Wu, B., and Harris, D. A. (2018). “Prion diseases,” in *The Molecular and Cellular Basis of Neurodegenerative Diseases: Underlying Mechanisms*, 1st Edn., ed. M. Wolfe (London: Academic Press), 23–56. Available online at: <https://www.elsevier.com/books/the-molecular-and-cellular-basis-of-neurodegenerative-diseases/wolfe/978-0-12-811304-2>
- Pan, T., Wong, B. S., Liu, T., Li, R., Petersen, R. B., and Sy, M. S. (2002). Cell-surface prion protein interacts with glycosaminoglycans. *Biochem. J.* 368, 81–90. doi: 10.1042/BJ20020773
- Shyng, S. L., Lehmann, S., Moulder, K. L., and Harris, D. A. (1995). Sulfated glycans stimulate endocytosis of the cellular isoform of the prion protein, PrP<sup>C</sup>, in cultured cells. *J. Biol. Chem.* 270, 30221–30229. doi: 10.1074/jbc.270.50.30221
- Solomon, I. H., Khatri, N., Biasini, E., Massignan, T., Huettnner, J. E., and Harris, D. A. (2011). An N-terminal polybasic domain and cell surface localization are required for mutant prion protein toxicity. *J. Biol. Chem.* 286, 14724–14736. doi: 10.1074/jbc.M110.214973
- Taylor, D. R., Whitehouse, I. J., and Hooper, N. M. (2009). Glypican-1 mediates both prion protein lipid raft association and disease isoform formation. *PLoS Pathog.* 5:1000666. doi: 10.1371/journal.ppat.1000666
- Van Gorp, C. L., Brister, S. J., Buchanan, M. R., and Linhardt, R. J. (1999). *Dermatan Disulfate, An Inhibitor of Thrombin Generation and Activation*.

- Patent No: US5922690A. Available online at: <https://patents.google.com/patent/US5922690A/en>
- Vieira, T. C. R. G., Reynaldo, D. P., Gomes, M. P. B., Almeida, M. S., Cordeiro, Y., and Silva, J. L. (2011). Heparin binding by murine recombinant prion protein leads to transient aggregation and formation of rna-resistant species. *J. Am. Chem. Soc.* 133, 334–344. doi: 10.1021/ja106725p
- Warner, R. G., Hundt, C., Weiss, S., and Turnbull, J. E. (2002). Identification of the heparan sulfate binding sites in the cellular prion protein. *J. Biol. Chem.* 277, 18421–18430. doi: 10.1074/jbc.M110406200
- Westergard, L., Turnbaugh, J. A., and Harris, D. A. (2011). A nine amino acid domain is essential for mutant prion protein toxicity. *J. Neurosci.* 31, 14005–14017. doi: 10.1523/JNEUROSCI.1243-11.2011
- Weyers, A., Yang, B., Solakyildirim, K., Yee, V., Li, L., Zhang, F., et al. (2013). Isolation of bovine corneal keratan sulfate and its growth factor and morphogen binding. *FEBS J.* 280, 2285–2293. doi: 10.1016/j.pestbp.2011.02.012. Investigations
- Wu, B., McDonald, A. J., Markham, K., Rich, C. B., McHugh, K. P., Tatzelt, J., et al. (2017). The N-terminus of the prion protein is a toxic effector regulated by the C-terminus. *eLife* 6:e23473. doi: 10.7554/eLife.23473
- Yates, E. A., Santini, F., Guerrini, M., Naggi, A., Torri, G., and Casu, B. (1996). <sup>1</sup>H and <sup>13</sup>C NMR spectral assignments of the major sequences of twelve systematically modified heparin derivatives. *Carbohydr. Res.* 294, 15–27. doi: 10.1016/s0008-6215(96)90611-4
- Zahn, R., Liu, A., Lühns, T., Riek, R., Von Schroetter, C., Garcia, F. L., et al. (2000). NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. U.S.A.* 97, 145–150. doi: 10.1073/pnas.97.1.145
- Zhang, F., Liang, X., Beaudet, J. M., Lee, Y., and Linhardt, R. J. (2014). The effects of metal ions on heparin/heparan sulfate-protein interactions. *J. Biomed. Technol. Res.* 1, 1–14. doi: 10.1016/j.physbeh.2017.03.040

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kim, Zhang, Harris and Linhardt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Delayed Comparison and Apriori Algorithm (DCAA): A Tool for Discovering Protein–Protein Interactions From Time-Series Phosphoproteomic Data

## OPEN ACCESS

### Edited by:

Emil Alexov,  
Clemson University, United States

### Reviewed by:

Grzegorz Wegrzyn,  
University of Gdansk, Poland  
Yasuteru Shigeta,  
University of Tsukuba, Japan

### \*Correspondence:

Qunye Zhang  
wz.zhangqy@sdu.edu.cn  
Lianli Chi  
lianlichis@sdu.edu.cn  
Peng Shi  
pshi@ustb.edu.cn

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 15 September 2020

**Accepted:** 02 November 2020

**Published:** 10 December 2020

### Citation:

Ding L, Xie S, Zhang S, Shen H,  
Zhong H, Li D, Shi P, Chi L and  
Zhang Q (2020) Delayed Comparison  
and Apriori Algorithm (DCAA): A Tool  
for Discovering Protein–Protein  
Interactions From Time-Series  
Phosphoproteomic Data.  
Front. Mol. Biosci. 7:606570.  
doi: 10.3389/fmolb.2020.606570

Lianhong Ding<sup>1†</sup>, Shaoshuai Xie<sup>2†</sup>, Shucui Zhang<sup>3</sup>, Hangyu Shen<sup>4</sup>, Huaqiang Zhong<sup>4</sup>,  
Daoyuan Li<sup>2</sup>, Peng Shi<sup>4\*</sup>, Lianli Chi<sup>2\*</sup> and Qunye Zhang<sup>3\*</sup>

<sup>1</sup> School of Information, Beijing Wuzi University, Beijing, China, <sup>2</sup> National Glycoengineering Research Center, Shandong University, Qingdao, China, <sup>3</sup> The Key Laboratory of Cardiovascular Remodeling and Function Research, Chinese Ministry of Education, Chinese National Health Commission and Chinese Academy of Medical Sciences, Qilu Hospital of Shandong University, Jinan, China, <sup>4</sup> National Center for Materials Service Safety, University of Science and Technology Beijing, Beijing, China

Analysis of high-throughput omics data is one of the most important approaches for obtaining information regarding interactions between proteins/genes. Time-series omics data are a series of omics data points indexed in time order and normally contain more abundant information about the interactions between biological macromolecules than static omics data. In addition, phosphorylation is a key posttranslational modification (PTM) that is indicative of possible protein function changes in cellular processes. Analysis of time-series phosphoproteomic data should provide more meaningful information about protein interactions. However, although many algorithms, databases, and websites have been developed to analyze omics data, the tools dedicated to discovering molecular interactions from time-series omics data, especially from time-series phosphoproteomic data, are still scarce. Moreover, most reported tools ignore the lag between functional alterations and the corresponding changes in protein synthesis/PTM and are highly dependent on previous knowledge, resulting in high false-positive rates and difficulties in finding newly discovered protein–protein interactions (PPIs). Therefore, in the present study, we developed a new method to discover protein–protein interactions with the delayed comparison and Apriori algorithm (DCAA) to address the aforementioned problems. DCAA is based on the idea that there is a lag between functional alterations and the corresponding changes in protein synthesis/PTM. The Apriori algorithm was used to mine association rules from the relationships between items in a dataset and find PPIs based on time-series phosphoproteomic data. The advantage of DCAA is that it does not rely on previous knowledge and the PPI database. The analysis of actual time-series phosphoproteomic data showed that more than 68% of the

protein interactions/regulatory relationships predicted by DCAA were accurate. As an analytical tool for PPIs that does not rely on a priori knowledge, DCAA should be useful to predict PPIs from time-series omics data, and this approach is not limited to phosphoproteomic data.

**Keywords:** protein–protein interactions, phosphoproteomics, delayed comparison, Apriori, DCAA

## INTRODUCTION

Protein–protein interactions (PPIs) are the basis and prerequisite for protein functions. Proteins *in vivo* are part of complex regulatory networks involving sophisticated interactions to coordinately regulate various biological processes and functions under different spatiotemporal conditions. The PPIs in living organisms are more complex than one might imagine. Therefore, PPI is one of the most critical issues in biomedical research (Braun and Gingras, 2012). Many experimental techniques and equipment for studying PPIs have been developed, such as immunoprecipitation, biolayer interferometry, and surface plasmon resonance (Douzi, 2017; Lin and Lai, 2017; Wu et al., 2017). Although these techniques are reliable and widely used, they are time- and cost-consuming and low-throughput. To reveal the mechanisms underlying physiological and pathological processes, high-throughput methods for studying PPIs are urgently needed. To date, some high-throughput experimental methods for detecting PPIs have been reported, such as yeast two-hybrid, tandem affinity purification, phage display, and protein chip methods (Gavin et al., 2002; Rao et al., 2014; Sundell and Ivarsson, 2014; Mehla et al., 2015; Huang et al., 2017; Viala and Bouveret, 2017; Woloschuk et al., 2020). However, these methods also have many drawbacks, including complexity, required time, and high cost. Therefore, computational methods could be useful supplements to high-throughput experimental methods (Lei et al., 2016; Sun et al., 2017; Zhang et al., 2017).

In the past two decades, high-throughput omics technologies, including genomes, transcriptomes and proteomes, have developed rapidly (Mortazavi et al., 2008; Consortium, 2012; Liu et al., 2018; Jiang et al., 2019). Likewise, many tools have been developed to analyze these omics data and obtain useful information about protein/gene interactions. For example, BindML+ can predict PPIs using an amino acid substitution model, and PIC (Protein Interaction Calculator) is a web tool to compute intra- and interprotein interactions (Tina et al., 2007; La et al., 2013). Time-series omics data are a series of omics data points indexed in time order and normally contain more abundant information about the interactions between biological macromolecules than static omics data. Therefore, many tools and websites have been proposed for discovering PPIs based on these data. To date, the algorithms used in the reported tools include learning vector quantization (LVQ), profile-kernel support vector machine, random forest classifier, semantic-based regularization (a machine learning framework), feature extraction, and deep learning (Planas-Iglesias et al., 2013; Yousef and Moghadam Charkari, 2013; Saccà et al., 2014; Hamp and Rost, 2015; Sun et al., 2017; Zeng et al., 2020). However,

many of the reported tools require substantial amounts of supporting data in addition to omics data, such as protein structure, protein/gene sequence, gene functional similarity, and protein–protein interaction databases (Planas-Iglesias et al., 2013; Sun et al., 2017; Zeng et al., 2020).

Phosphorylation is the most common posttranslational modification (PTM) of proteins for functional regulation (Cohen, 2000; Ardito et al., 2017). Thus, phosphoproteomic data might suggest possible changes in protein function. Analysis of time-series phosphoproteomic data should provide more meaningful information about protein interactions. However, although many algorithms, databases, and websites have been developed to analyze omics data, the tools dedicated to discovering molecular interactions from time-series omics data, especially from time-series phosphoproteomic data, are still scarce. More importantly, protein interactions and PTMs are a series of events that undergo sequential and dynamic alterations. There are lags between functional alterations and the corresponding changes in protein synthesis/PTM. However, the most widely reported tools ignore these lags and are highly dependent on previous knowledge, resulting in high false-positive rates and difficulties in finding newly discovered PPIs. In addition, the false-positive rates of PPIs predicted by many tools from static omics data are very high.

In this study, considering the aforementioned lags, we propose a novel method for predicting PPIs combining delayed comparison and the Apriori algorithm (DCAA), which does not rely on previous knowledge. High-throughput dynamic phosphoproteomic data from human umbilical vein endothelial cells treated with oxidized low-density lipoprotein (ox-LDL) were used to verify this method. By not relying on previous knowledge and the PPI database, DCAA could discover PPIs from dynamic phosphoproteomic data with a relatively low false-positive rate. Moreover, DCAA should also be applied to other time-series omics data.

## MATERIALS AND METHODS

### Cell Treatment and Protein Digestion by Trypsin

EA.hy926 cells were purchased from the American Tissue Culture Collection (Manassas, USA) and cultured in Dulbecco's modified essential medium containing 10% fetal bovine serum (FBS), 100 U/ml penicillin G, and 100 µg/ml streptomycin. After treatment with 50 µg/ml ox-LDL for 0, 0.5, 1, 1.5, 2, 4, 6, 8, 12, 18, 24, 36, 48, and 72 h, all cells were harvested and lysed in lysis buffer (8 M urea, 50 mM Tris–HCl, 10% isopropyl alcohol, 12.5% isobutyl alcohol containing complete protease inhibitor cocktail



and PhosSTOP phosphatase inhibitor cocktail). The common control samples were produced by mixing equal amounts of all 14 samples. All samples (2 common controls and 14 time point samples, 400 µg/sample) were reduced with dithiothreitol for 1 h after alkylation by iodoacetamide for 1 h in the dark. After replacing the solvent with 50 mM triethylammonium bicarbonate using ultracentrifugal filtration units (MWCO 10 kDa), all samples were digested by trypsin at 37°C for 18 h with a 50:1 protein-to-protease ratio.

## Phosphopeptide Enrichment and iTRAQ Labeling

The tryptic digests of all 16 samples were labeled with 8-Plex iTRAQ (SCIEX, MA, USA) according to the manufacturer's instructions. Then, the 16 samples were equally divided into two sample pools. Each sample pool contained equal amounts of proteins from a common control (labeled with iTRAQ 113) and seven time point samples (from 0 to 6 h, respectively labeled with iTRAQ 114–iTRAQ 121; or from 8 to 72 h, respectively labeled with iTRAQ 114–iTRAQ 121). After desalting and drying, the peptides from two sample pools were dissolved in 5 ml of TiO<sub>2</sub> loading buffer (1.25 M glycolic acid, 80% ACN, 1% TFA) and incubated with 16 mg of Titanosphere TiO<sub>2</sub> (5 µm; GL Science, Tokyo, Japan) for 30 min. The TiO<sub>2</sub> beads were washed sequentially with TiO<sub>2</sub> loading buffer, 1% TFA in 80% aqueous ACN, and 0.1% TFA in 2% aqueous ACN. Then, the beads were eluted sequentially with 8% NH<sub>4</sub>OH and 50 mM phosphate buffer (pH 12.0). Finally, the eluates were combined and immediately neutralized with 10% FA.

## Peptide Fractionation Using Basic Reversed-Phase Liquid Chromatography

After neutralization, each sample pool was dissolved in 80 µl of buffer A (2% ACN, 15 mM NH<sub>4</sub>COOH, pH 10.0) and separated using basic reversed-phase liquid chromatography at 0.14 ml/min with a Kinetex EVO C18 column (2.6 µm particles, 100 Å, 15 cm × 2.1 mm; Phenomenex, Torrance, USA). For separation, a step gradient of 2% B (80% ACN, 15 mM NH<sub>4</sub>COOH, pH 10.0), 0–8 min; 2–28% B, 8–68 min; 28–40% B, 68–78 min; 40–100% B, 78–83 min; and 100% B, 83–93 min was used. The eluates were collected at 1-min intervals and then pooled into 16 fractions. After desalting and drying, the fractions were dissolved in 0.1% FA for LC-MS/MS analysis.

## Nano-LC-MS/MS Analysis

Each sample pool was loaded on a ReproSil-Pur C18 precolumn (3 cm × 100 µm, 5 µm, 120 Å; Dr Maisch, Germany) at 5 µl/min using an Easy nLC-1000 nano-LC system (Thermo Scientific, San Jose, CA, USA). For separation, mobile phase A was 0.1% FA in 2% ACN, and mobile phase B was 0.1% FA in 98% ACN. A step gradient of 2–8% B, 0–5 min; 8–22% B, 5–85 min; 22–30% B, 85–105 min; 30–90% B, 105–110 min; and 90% B, 110–120 min was used at 300 nl/min. Data-dependent MS/MS was performed using an Orbitrap Fusion mass spectrometer in positive ion mode with the following parameters: 2.2 kV spray voltage, 275°C capillary temperature, 55% S-lens level, 350–1,550 mass acquisition range, and 120,000 resolution for MS analysis.

Each precursor ion scan was followed by a 4-s top speed data-dependent HCD MS/MS at 35% normalized collision energy. The resolution for MS/MS analysis was 30,000. The quadrupole isolation width was 2 *m/z*. The dynamic exclusion time was 60 s with a ±10 ppm exclusion mass width. The raw data were processed using Proteome Discoverer version 1.4.0.28 and the UniProt database. The PhosphoRS 3.0 algorithm was used to evaluate the localization probabilities of phosphorylation sites.

## Clustering Process of the DCAA Method

To reduce the data volume for better processing, the original data consisting of the relative expression levels of phosphopeptides were first clustered according to shape similarity, namely, based on the well-known fact that trends of greater data similarity indicate closer relationships between data. Three clustering methods, including full trend clustering, angle clustering, and Pearson clustering, were used in the present study. Full trend clustering classifies the original data according to the changing trend (increasing, decreasing, or unchanging) of every line segment, which is constructed with two close time points of the relative expression data of phosphopeptides. The phosphopeptides with the same changing trends are classified into one cluster. The angle clustering classifies the original data considering both the speed and trend of the changes in the relative expression of phosphopeptides. The change speeds are measured by the angle between two-line segments constructed as mentioned previously. The phosphopeptides with angles less than a given value are classified into one cluster. Pearson clustering classifies the original data based on the Pearson correlation coefficient of the relative expression of phosphopeptides, which indicates the degree of linear correlation. The Pearson correlation coefficient between vectors *X* and *Y* is defined as follows:

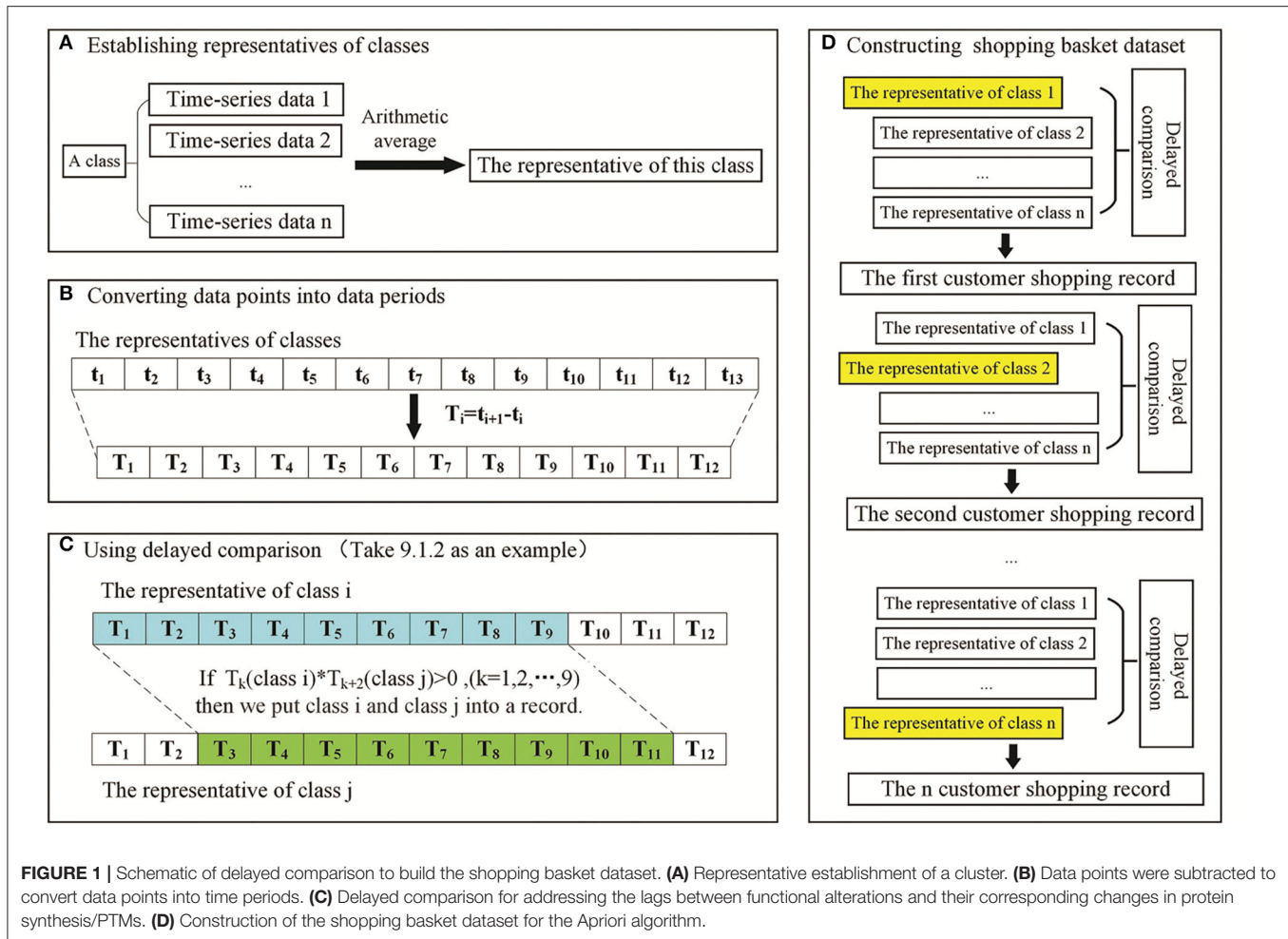
$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\delta_X \delta_Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right) \left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}} \quad (1)$$

where *N* is the dimension number of vectors *X* and *Y*.

## The Delayed Comparison Processes of the DCAA Method

Delayed comparison and shopping basket dataset construction were used to reflect the lags between upstream and downstream events of PPIs. The shopping basket dataset is named to reflect the concept of shopping in supermarkets in the data mining area. The data selection process is analogized as a good selection process in supermarkets. The shopping basket dataset is the set containing the selected data from all candidate data. There were 13 different time-points of phosphoprotein change data compared with the data at 0 h. Normally, the lags between upstream and downstream PPIs did not last very long. Therefore, we fixed nine time periods to build sliding time windows and control the delayed time periods within three time periods. The steps of delayed comparison are as follows:

- Establish representatives of classes: the representative of a class was the arithmetic average of all the data in this class.



**FIGURE 1 |** Schematic of delayed comparison to build the shopping basket dataset. **(A)** Representative establishment of a cluster. **(B)** Data points were subtracted to convert data points into time periods. **(C)** Delayed comparison for addressing the lags between functional alterations and their corresponding changes in protein synthesis/PTMs. **(D)** Construction of the shopping basket dataset for the Apriori algorithm.

- Convert data points into time periods: two adjacent data points were subtracted ( $T_1 = t_2 - t_1, T_2 = t_3 - t_2, \dots, T_{12} = t_{13} - t_{12}$ ), and the values of time periods were obtained ( $T_1, T_2, \dots, T_{12}$ ).
- Use delayed comparison: to fix time periods, a start time and delayed time periods were chosen. For example, 9.1.2 means there are 9 time periods, the start time is 1 and there are 2 delayed time periods. Sliding windows of six groups (9.1.1, 9.1.2, 9.1.3, 9.2.1, 9.2.2, and 9.3.1.) will cover all delayed scenarios.
- Construct the shopping basket dataset: one cluster of original data was compared with another cluster of original data after sliding to achieve delayed comparison, thereby producing one shopping basket data item. Six shopping basket datasets obtained by delayed comparison were aggregated into one dataset to build the experimental dataset for the Apriori algorithm.

## The Processes of the Apriori Algorithm in the DCAA Method

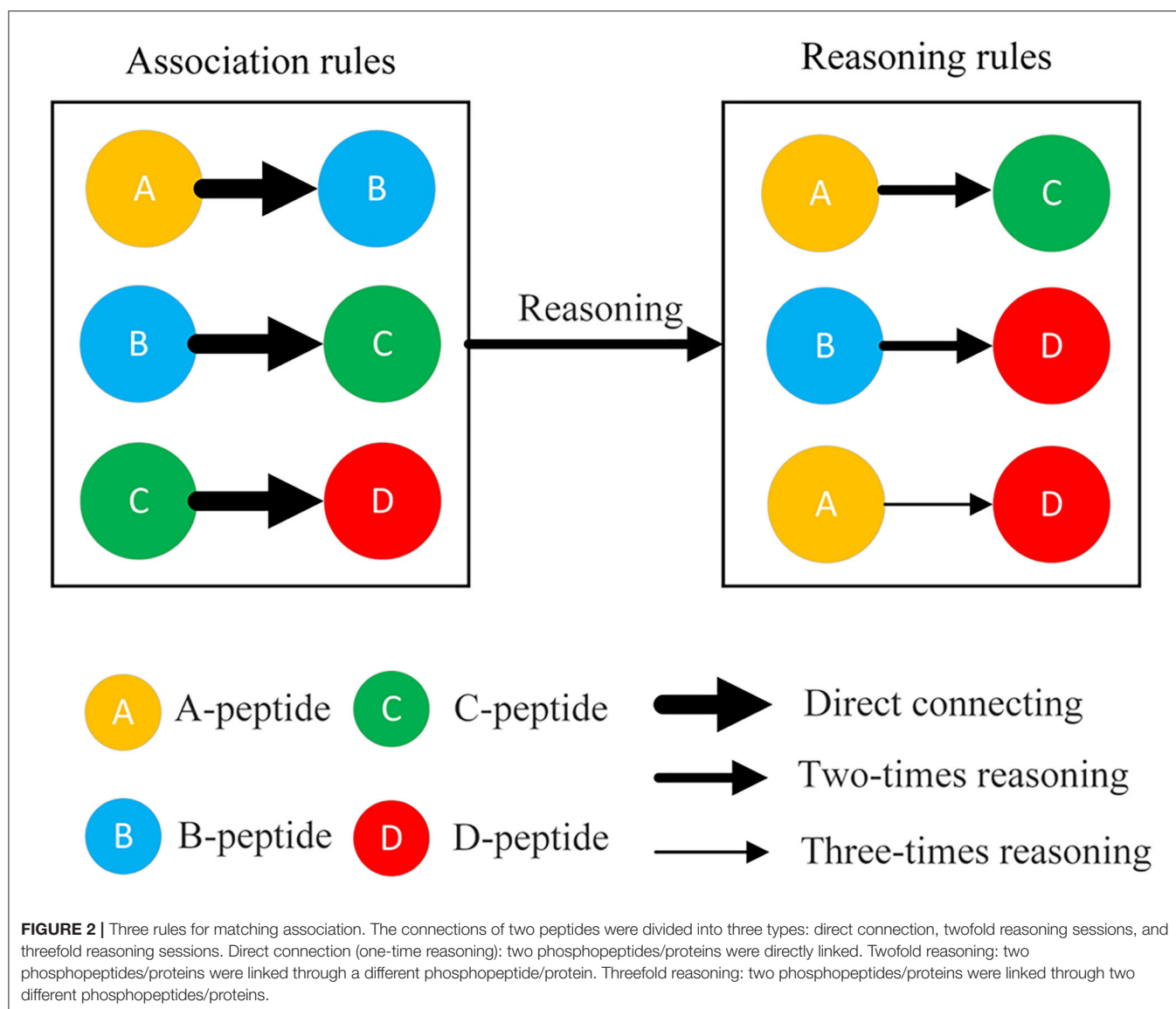
The Apriori algorithm is a kind of machine learning algorithm, and its core purpose is to mine frequent sets of customer

shopping records and excavate association rules (Agrawal et al., 1993). The nonempty subset of a frequent item set must be a frequent item set. The Apriori algorithm first generates one frequent item sets and then uses one frequent item sets to generate two frequent item sets. Next, three frequent item sets are generated from two frequent item sets. Finally, all frequent item sets are generated. Then, the association rules are found from these frequent item sets. The Apriori algorithm is outlined in **Appendix 1**. The confidence and minimum support of the Apriori algorithm are set up for obtaining interclass inference results. The support degree and confidence degree of an association rule between  $X$  and  $Y$  are, respectively, as follows:

$$\text{Support}(X, Y) = P(XY) = \frac{\text{number}(XY)}{\text{num}(\text{All Samples})} \quad (2)$$

$$\text{Confidence}(X \Leftarrow Y) = P(X|Y) = P(XY)/P(Y) \quad (3)$$

An example of building a shopping basket dataset with a delayed comparison of 9.1.2 is shown in **Figure 1**. A customer shopping record was constructed as follows:  $T_k(\text{class } i) * T_{k+2}(\text{class } j) > 0$  or  $T_k(\text{class } i) = 0$  and  $T_{k+2}(\text{class } j) = 0$  ( $k = 1, 2, \dots, 9$ ), class  $i$  and class  $j$  are put into one record. After applying the sliding window treatment described previously once,  $n$  ( $n \geq 0$ ) customer



shopping records can be produced. Six shopping basket datasets are obtained using six kinds of sliding windows. They are combined into one dataset as the experimental dataset of the Apriori algorithm.

### Matching Process of the DCAA Method

The association rules among phosphopeptides were discovered from the delayed comparison and Apriori algorithm results using different reasoning methods, including direct relationship (one-time reasoning), two-times reasoning, and three-times reasoning (Figure 2). One-time reasoning means that two phosphopeptides/proteins are directly linked. Two-times reasoning means that two phosphopeptides/proteins are linked through one different phosphopeptide/protein. Three-times reasoning means that two phosphopeptides/proteins are linked through two different phosphopeptides/proteins, but the existing relationships in one-time reasoning and two-time reasoning

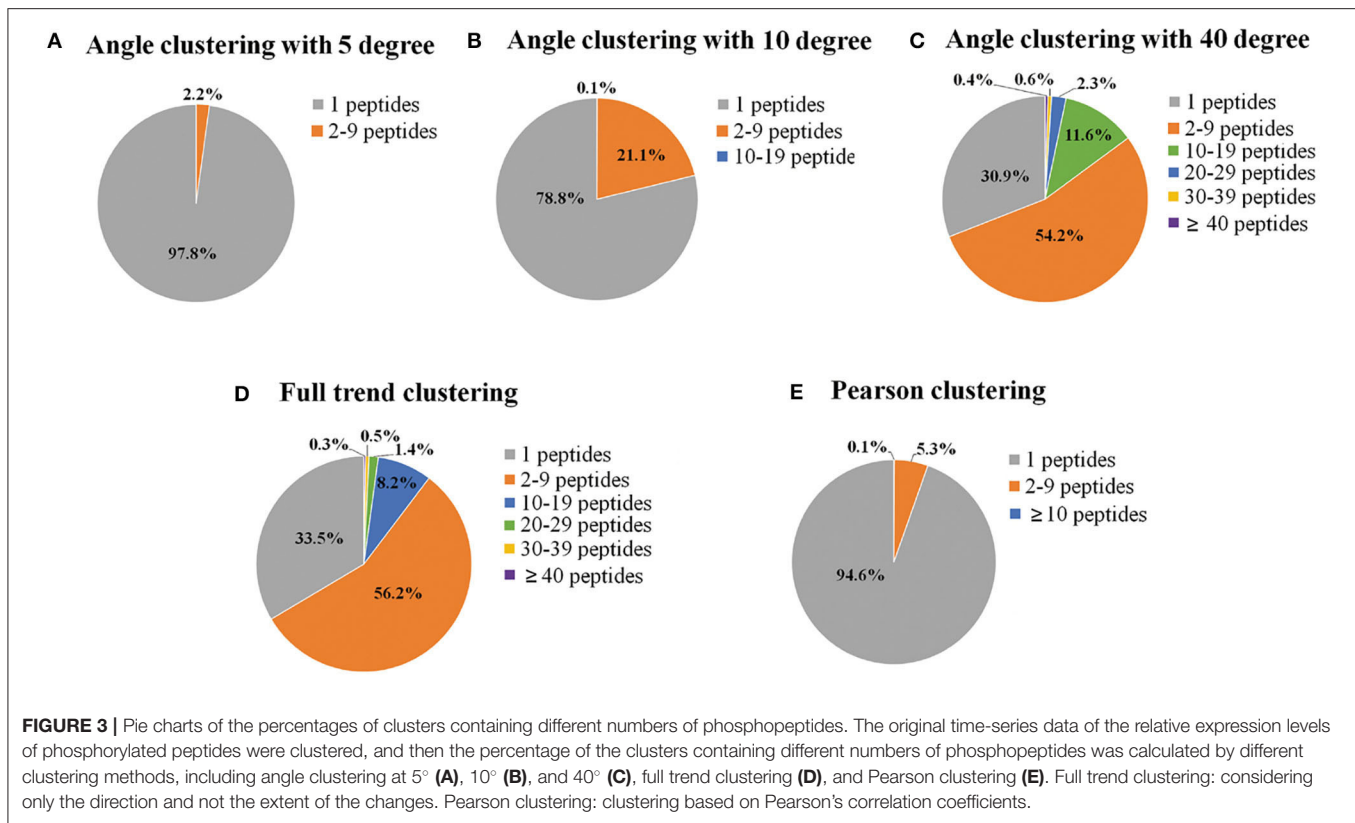
were excluded. Finally, a database of interacting proteins (DIP) consisting of the reported PPIs was used to evaluate the relationships discovered by DCAA by matching the relationships discovered by DCAA and the relationships recorded in the DIP (Xenarios et al., 2002).

## RESULTS AND DISCUSSION

### Different Clustering Methods Reflect Similar Patterns of Change in Data

In the present study, endothelial cells were seeded in T-175 flasks and treated with ox-LDL for different time periods to investigate the changes in protein phosphorylation. Four-hundred-microgram samples of proteins were taken from each time point to map phosphorylation sites using nano-LC-MS/MS. A total of 17,287 phosphorylation sites were identified on 15,037 phosphopeptides from 4,539 proteins. The dataset was used for



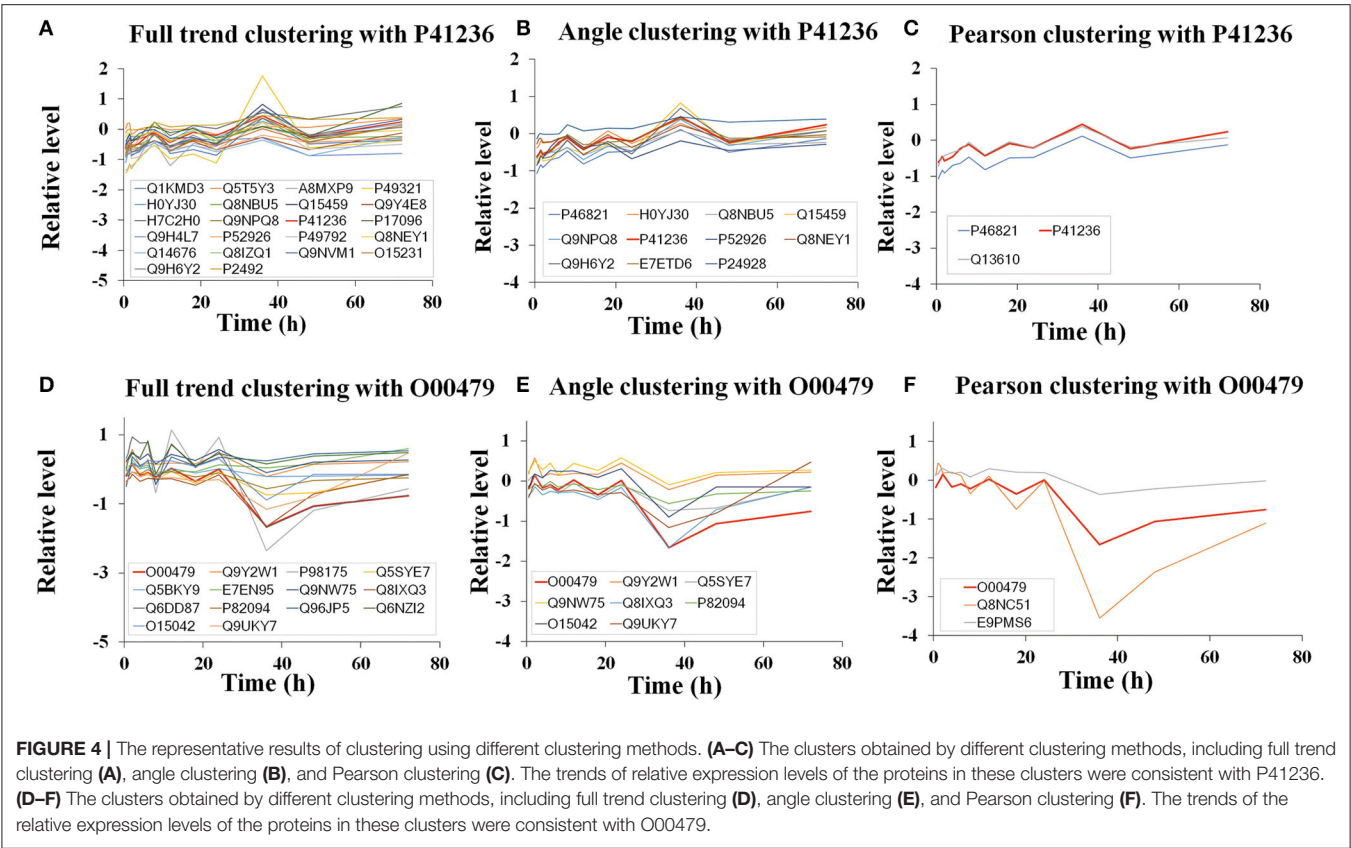


mining the protein–protein interaction network. Each group contained 13 data points at different times (0.5, 1, 1.5, 2, 4, 6, 8, 12, 18, 24, 36, 48, and 72 h). To reduce the computational load of subsequent data processing, clustering was the first process of DCAA. Three different clustering methods, named full trend clustering, angle clustering, and Pearson clustering, were used to classify the original data set (Figure 3). The clustering algorithm in the paper can be regarded as the direct angle threshold method. For one group of data, 13 points can produce 12 line segments connected by two adjacent points. The 12 angles of the corresponding lines from the two groups of data were compared. If all of the angles are not larger than the threshold, they are classified into two classes. Physically, this indicates that the changes in every specific period between the two groups of data are similar. Angle clustering clustered the data groups with the same changing trend and similar changing speed. Therefore, angle clustering should divide the clusters of full trend clustering into smaller clusters. When the angle was set to 5°, 97.8% of clusters only contained one element class (only one peptide in the class), which implies failure of clustering (Figure 3A). When the angle was set to 10°, the percentage of clusters containing one peptide was significantly decreased (Figure 3B). When the angle was set to 40°, the percentages of clusters containing different amounts of peptides were similar to those of full trend clustering (Figures 3C,D). For Pearson clustering, different thresholds of the Pearson correlation coefficient ( $R$ ) were tested on the original dataset. Most peptides with obvious relativities were put into one

cluster when the threshold was set to 0.97. Therefore, we classify the data with  $|R| > 0.97$  as one cluster (Figure 3E). In sum, we obtained 3,494 clusters by full trend clustering, 12,300 clusters by angle clustering at 10°, and 13,686 clusters by Pearson clustering.

The indicators for evaluation of clustering algorithms included external and internal standards. External standards required knowing the previous distribution of samples. However, angle clustering at 10°, full trend clustering, and Pearson clustering are unsupervised clustering methods, so we evaluated them by internal standards. The internal standards were mainly based on the principle of inner-class distance and interclass distance. As Figure 4 shows, compared with angle clustering at 10° and Pearson clustering, full trend clustering had a higher degree of aggregation in the clustering of time-series data. The actual hit rates of interclass inference rules were obtained under different conditions for the aforementioned three clustering methods (Table 1). The average hit rate of full trend clustering was 68.7%, and full trend clustering was better than the other two clustering methods.

As examples, the clustering results for proteins P41236 and O00497 using the three clustering methods are shown in Figure 4. For protein P41236, 22, 11, and 3 peptides were placed into one cluster by full trend clustering, angle clustering at 10°, and Pearson clustering, respectively (Figures 4A–C). For protein O00497, 14, 8, and 3 peptides were placed into one cluster by full trend clustering, angle clustering at 10°, and Pearson clustering, respectively (Figures 4D–F).



**TABLE 1 |** The numbers of interclass inference rules using different minimum supporting degrees and confidence degrees.

Full trend clustering			Angle clustering at 10°			Pearson clustering		
MSD	MCD	Rule numbers	MSD	MCD	Rule numbers	MSD	MCD	Rule numbers
0.0054	0.9	0	0.026	0.9	0	0.020	0.9	2
0.0052	0.9	28	<b>0.024</b>	<b>0.9</b>	<b>528</b>	0.018	0.9	2
<b>0.0050</b>	<b>0.9</b>	<b>106</b>	0.022	0.9	566	0.016	0.9	78
0.0048	0.9	1,111	0.020	0.9	1,768	<b>0.014</b>	<b>0.9</b>	<b>210</b>
0.0046	0.9	2,252	0.018	0.9	2,094	0.012	0.9	427
0.0044	0.9	4,290	0.016	0.9	3,950	0.010	0.9	960
0.0042	0.9	11,136	0.014	0.9	5,168	0.008	0.9	3,263
0.0040	0.9	37,692	0.012	0.9	6,932	0.006	0.9	6,982
0.0038	0.9	74,941	0.010	0.9	11,223	0.004	0.9	17,641
0.0036	0.9	129,345	0.008	0.9	14,453	0.002	0.9	36,150

MSD, minimum support degree; MCD, minimum confidence degree.  
The bold and italic values indicates the most suitable MSD and MCD degree for different clustering.

### Evaluation of Different Clustering Methods of DCAA

To evaluate different clustering methods used in DCAA, the DCAA hit rates for the association rules in the DIP were calculated by comparing the association rules based on each interclass inference rule obtained from DCAA with the records of association rules in the DIP. In addition, the same number of protein pairs as that in the association rules obtained from DCAA were randomly selected from the original dataset containing

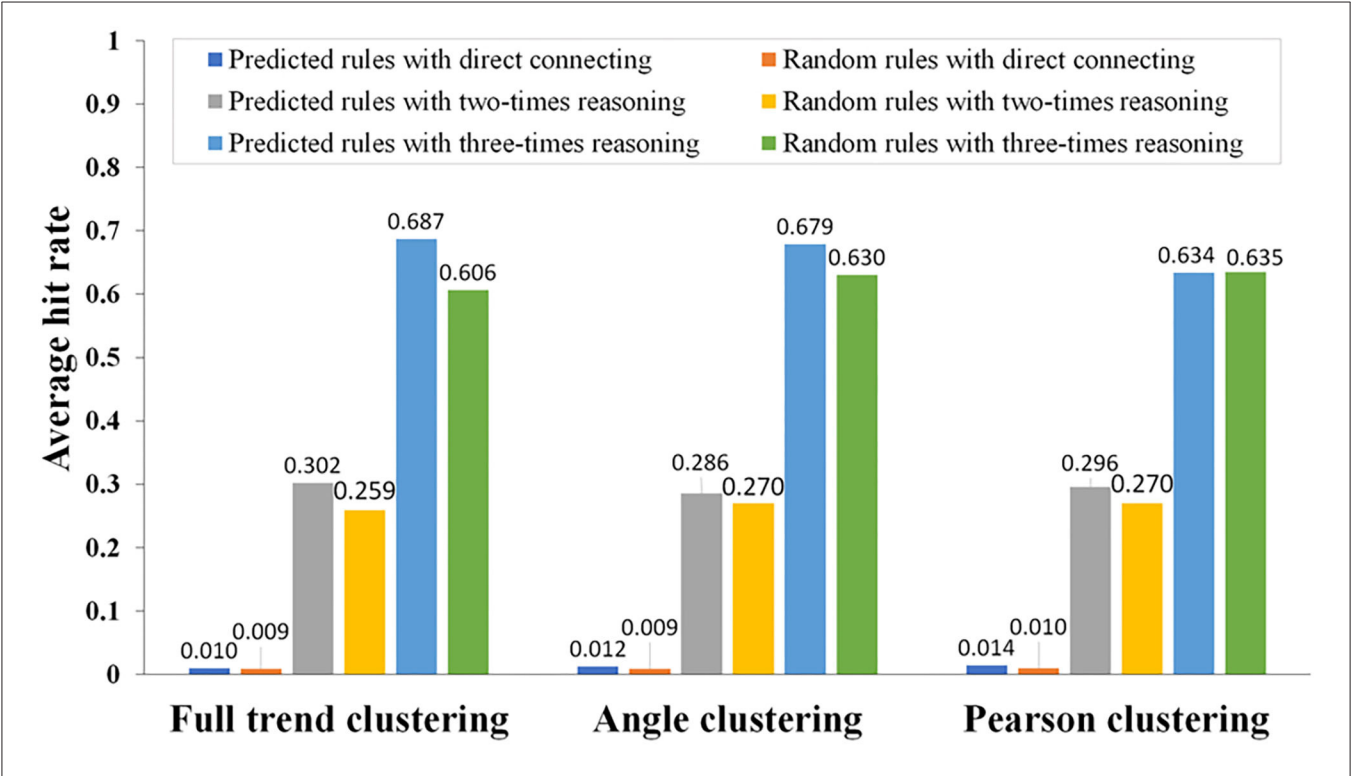
over 4,000 proteins, and the random association rules were created by randomly associating pairs of these proteins. Then, the random hit rates of association rules in DIP were calculated by comparing the random association rules with the records of association rules in DIP. The operation was repeated 100 times to obtain the average hit rate of random matching. Next, the DCAA hit rates of association rules in DIP were compared with the random hit rates of association rules. The calculation formulas of hit rates of DCAA and random matching were

as follows:

$$R = \frac{m}{M} \tag{4}$$

$$R' = \frac{n}{M'} \tag{5}$$

where  $m$  is the number of association rules in the DIP predicted by DCAA and  $n$  is the number of randomly predicted association rules in the DIP.  $M$  and  $M'$  are the number of predicted association rules and random association rules, respectively. The results showed that there were significant differences between

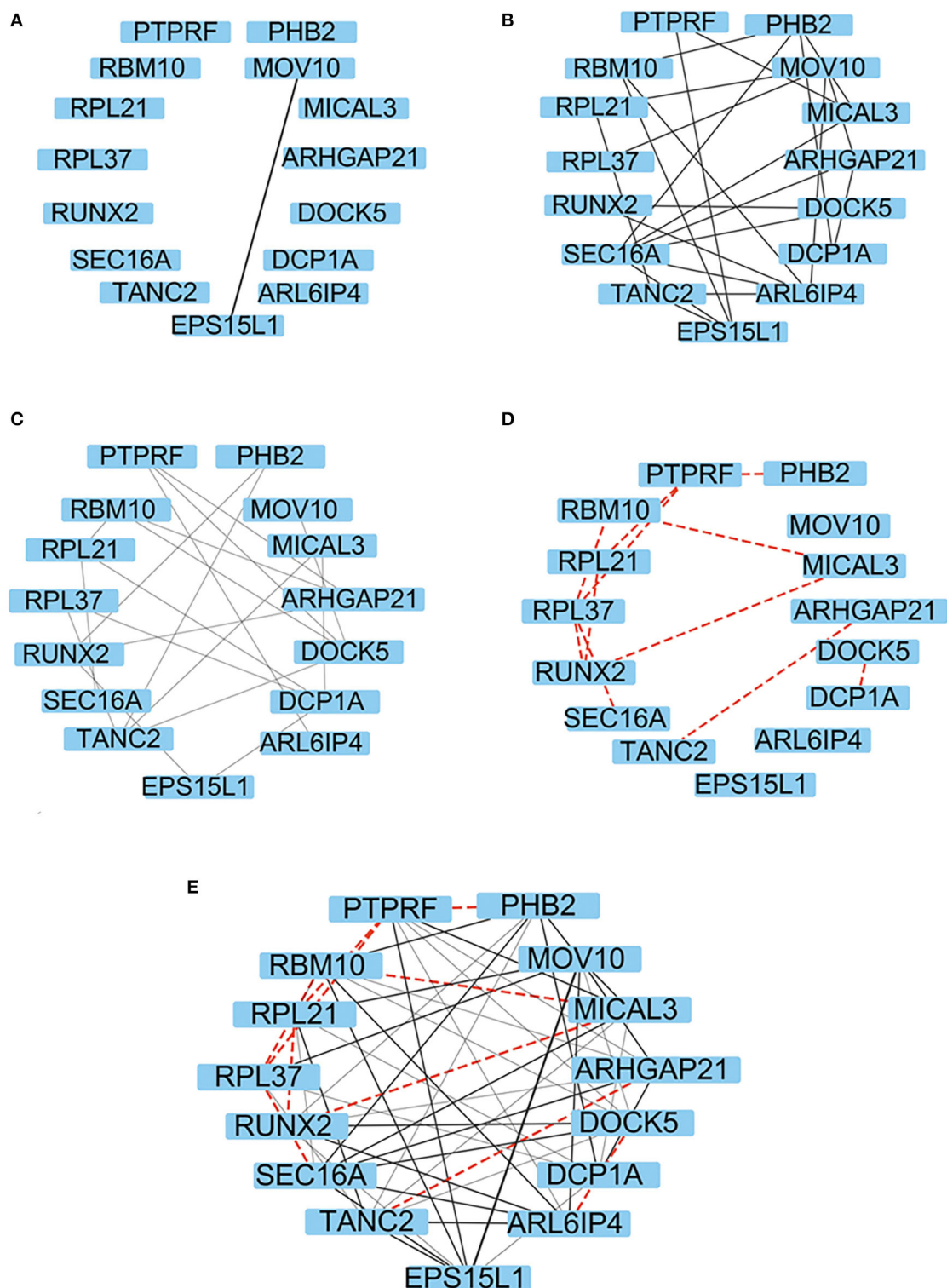


**FIGURE 5 |** Comparison of the association rules predicted by DCAA and the random association rules using different reasoning modes. Using different clustering methods (direct connection, twofold reasoning, and threefold reasoning), the hit rates of association rules were calculated by comparing the predicted association rules with the association rules recorded in the DIP. Hit rates of DCAA and random matching were compared to judge whether there was a significant difference.

**TABLE 2 |** Examples of sliding windows with different strategies.

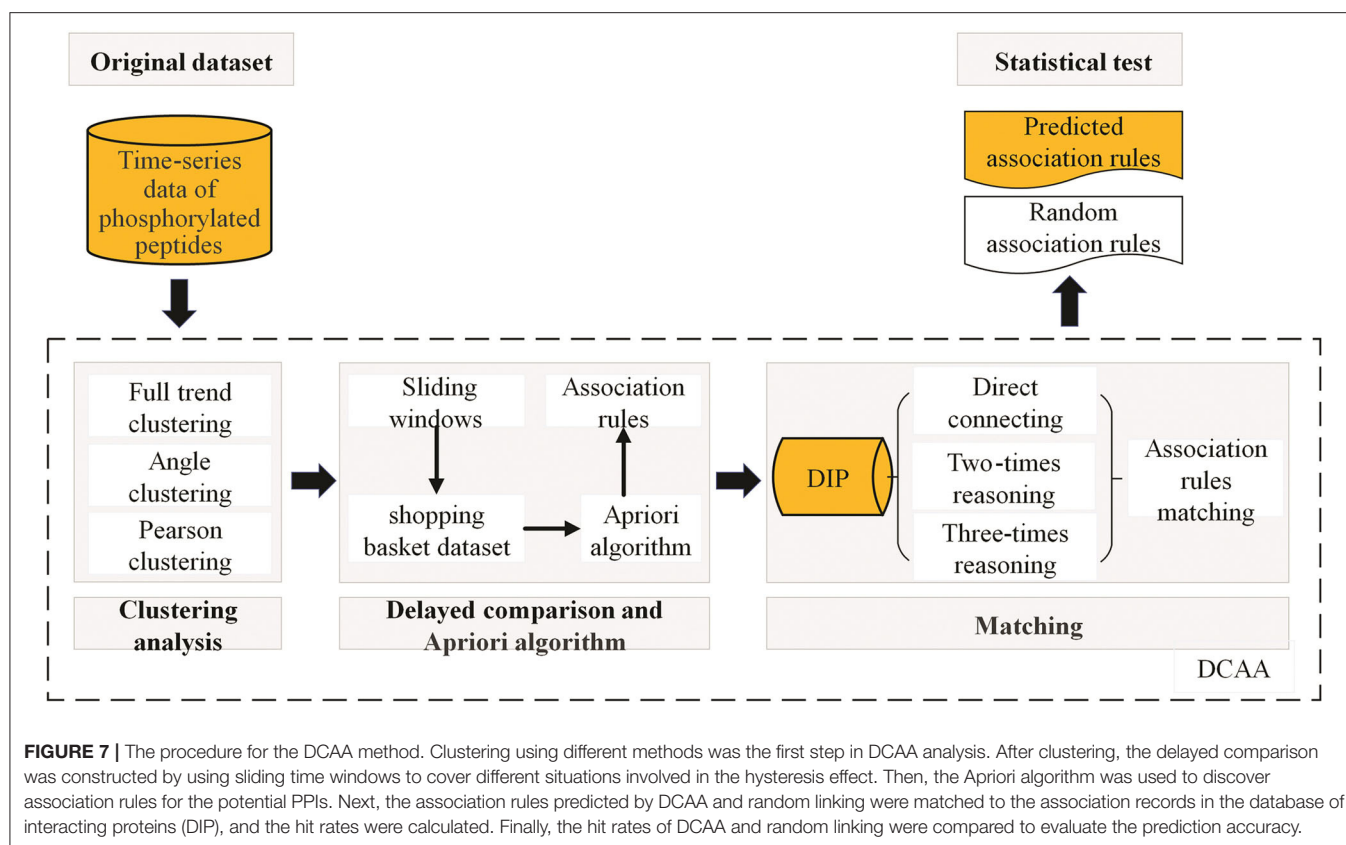
Type	Compared class	Time (hours)												
		0.5	1	1.5	2	4	6	8	12	18	24	36	48	72
9.1.1	Class one	−0.11	−0.21	−0.33	−0.05	−0.01	−0.27	0.06	0.54	0.07	−0.05	−0.21	−0.37	−1.05
	Class two	−0.15	−0.11	−0.08	−0.12	−0.24	−0.25	−0.18	−0.21	−0.32	−0.54	−0.68	−0.28	−0.19
9.1.2	Class one	−0.11	−0.21	−0.33	−0.05	−0.01	−0.27	0.06	0.54	0.07	−0.05	−0.21	−0.37	−1.05
	Class two	−0.15	−0.11	−0.08	−0.12	−0.24	−0.25	−0.18	−0.21	−0.32	−0.54	−0.68	−0.28	−0.19
9.1.3	Class one	−0.11	−0.21	−0.33	−0.05	−0.01	−0.27	0.06	0.54	0.07	−0.05	−0.21	−0.37	−1.05
	Class two	−0.15	−0.11	−0.08	−0.12	−0.24	−0.25	−0.18	−0.21	−0.32	−0.54	−0.68	−0.28	−0.19
9.2.1	Class one	−0.11	−0.21	−0.33	−0.05	−0.01	−0.27	0.06	0.54	0.07	−0.05	−0.21	−0.37	−1.05
	Class two	−0.15	−0.11	−0.08	−0.12	−0.24	−0.25	−0.18	−0.21	−0.32	−0.54	−0.68	−0.28	−0.19
9.2.2	Class one	−0.11	−0.21	−0.33	−0.05	−0.01	−0.27	0.06	0.54	0.07	−0.05	−0.21	−0.37	−1.05
	Class two	−0.15	−0.11	−0.08	−0.12	−0.24	−0.25	−0.18	−0.21	−0.32	−0.54	−0.68	−0.28	−0.19
9.3.1	Class one	−0.11	−0.21	−0.33	−0.05	−0.01	−0.27	0.06	0.54	0.07	−0.05	−0.21	−0.37	−1.05
	Class two	−0.15	−0.11	−0.08	−0.12	−0.24	−0.25	−0.18	−0.21	−0.32	−0.54	−0.68	−0.28	−0.19

By comparing the situation before and after sliding, a sliding window with different strategies could reflect the lags between upstream and downstream events of PPIs. The same color indicates the classes to be compared.



**FIGURE 6 |** The representation of the association rules/PPIs revealed by DCAA. **(A–C)** The representation of the PPIs recorded in the DIP revealed by DCAA using different reasoning modes, namely, direct connection **(A)**, twofold reasoning **(B)**, and threefold reasoning **(C)**. **(D)** Representation of the PPIs that were predicted by DCAA but were not recorded in the DIP. The black solid lines represent the PPIs recorded in the DIP. The red dashed lines represent the PPIs predicted by DCAA. The thickness represents the distance of the relationship; the thicker the line is, the closer the relationship is **(E)** Representation of PPIs that were combined with the results of **(A–D)**.





the DCAA hit rates and the random hit rates ( $p < 0.01$ ; **Figure 5**).

The three clustering methods used in this study were based on the shape similarity of time-series data, whereas popular approaches, including *k*-means and hierarchical clustering, measure the relevance between two clusters by distance. The methods based on distance ignore similar changing trends. However, the changing trends are important features of PPIs. Therefore, many peptides with similar changing trends but long distances are divided into different categories by clustering methods based on distance. Compared with these methods, full trend clustering, angle clustering, and Pearson clustering might be more suitable for PPI analysis.

### Types of Sliding Windows to Obtain Interclass Inference Rules

In the present study, a sliding time window was used for delayed comparison to reflect the lags between upstream and downstream events of PPIs. The size of the time window represents the affected time of delayed influence. Because the time of delayed influence in the test was not certain, many delayed comparisons with different time windows were executed. All possible delayed influences were attained by these tests. The type of delayed comparison was denoted as a combination of three integers connected by two dots. The first number indicated the length of data to be compared (the size of the sliding window). The start time and the length of the delayed time period were denoted by the second and third numbers, respectively. Some

examples of delayed sliding window comparisons are shown in **Table 2**. The delayed comparison resulted in six datasets, which were merged into one dataset to construct the dataset for the Apriori algorithm. After processing with the Apriori algorithm, 3,494, 12,213, and 13,138 customer shopping records were obtained by full trend clustering, angle clustering, and Pearson clustering, respectively.

Because too many rules could drastically increase the computational load, and too few rules may not support sufficient reasoning, the number of rules should be controlled at a proper level. In this paper, the number of rules was controlled to a few hundred. As the confidence degree mainly affected the probability of rule occurrence, we fixed the confidence degree at 0.9 and only adjusted the value of the support degree. Equally spaced values of support degree were tested, and the numbers of obtained rules are shown in **Table 1**. When using full trend clustering, 106 interclass inference rules were produced at a minimum support degree of 0.005. When using angle clustering at  $10^\circ$ , 528 interclass inference rules were obtained at a minimum support degree of 0.024. When using Pearson clustering, 210 interclass inference rules were obtained at a minimum support degree of 0.014 (**Supplementary Table 1**).

### The Matching Process of DCAA and Its Comparison With PPI Records in the DIP

The association rules identified by DCAA were compared with the reported association rules in the DIP. **Figure 6** shows an example of the matching results. **Figures 6A–C** represents the

inference results identified from the DIP under the conditions of direct connection, two-times reasoning, and three-times reasoning, respectively. **Figure 6D** shows the association rules that were predicted by DCAA but were not recorded in the DIP. These newly discovered PPIs should be helpful for designing biological experiments.

## CONCLUSIONS

In the present study, a novel tool, DCAA, was developed to discover PPIs from time-series phosphoproteomic data. The basic idea of DCAA was to classify the peptides with similar changing trends into one class by clustering and then identify the association rules among different classes by delayed comparison and the Apriori algorithm. DCAA consists of three main steps, namely, clustering, delayed comparison, and the Apriori algorithm, as well as matching (**Figure 7**). In DCAA, the lags between upstream and downstream events of PPIs were considered. Therefore, DCAA can find novel association rules of proteins with relatively lower false-positive rates without previous knowledge and databases. DCAA should be useful to predict PPIs from time-series omics data, which is not limited to phosphoproteomic data.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

## REFERENCES

- Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.* 22, 207–216. doi: 10.1145/170036.170072
- Ardito, F., Giuliani, M., Perrone, D., Troiano, G., and Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *Int. J. Mol. Med.* 40, 271–280. doi: 10.3892/ijmm.2017.3036
- Braun, P., and Gingras, A. C. (2012). History of protein-protein interactions: from egg-white to complex networks. *Proteomics*. 12, 1478–1498. doi: 10.1002/pmic.201100563
- Cohen, P. (2000). The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem. Sci.* 25, 596–601. doi: 10.1016/S0968-0004(00)01712-6
- Consortium, G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632
- Douzi, B. (2017). Protein-protein interactions: surface plasmon resonance. *Methods Mol Biol.* 1615, 257–275. doi: 10.1007/978-1-4939-7033-9\_21
- Gavin, A. C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141. doi: 10.1038/415141a
- Hamp, T., and Rost, B. (2015). Evolutionary profiles improve protein-protein interaction prediction from sequence. *Bioinformatics* 31, 1945–1950. doi: 10.1093/bioinformatics/btv077
- Huang, B. Y., Chen, P. C., Chen, B. H., Wang, C. C., Liu, H. F., Chen, Y. Z., et al. (2017). High-throughput screening of sulfated proteins by using a genome-wide proteome microarray and protein tyrosine sulfation system. *Anal. Chem.* 89, 3278–3284. doi: 10.1021/acs.analchem.6b02853
- Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., et al. (2019). Proteomics identifies new therapeutic targets of early-stage

## AUTHOR CONTRIBUTIONS

LD, HS, and HZ developed the algorithm, SX, SZ, and DL completed the experiment, PS, LC, and QZ designed the project, interpreted the data, and wrote the paper. All authors participated in writing the paper.

## FUNDING

This work was supported by the National Natural Science Foundation of China (No. 91439111, 81570712, and 81670247) and the Natural Science Outstanding Youth Foundation of Shandong Province (No. JQ201519), Major Science and Technology Innovation Project of Shandong Province (No. 2018CXGC1218), Jinan Clinical Medical Science and Technology Innovation Program (No. 201805055), and the Science and Technology Plan General Program of Beijing Municipal Education Commission (No. KM201910037186).

## ACKNOWLEDGMENTS

The authors would like to thank Xiaowei Wang and Xinjie Zhang for their help in the experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.606570/full#supplementary-material>

hepatocellular carcinoma. *Nature* 567, 257–261. doi: 10.1038/s41586-019-0987-8

- La, D., Kong, M., Hoffman, W., Choi, Y. I., and Kihara, D. (2013). Predicting permanent and transient protein-protein interfaces. *Proteins*. 81, 805–818. doi: 10.1002/prot.24235
- Lei, X., Wang, F., Wu, F. X., Zhang, A., and Pedrycz, W. (2016). Protein complex identification through Markov clustering with firefly algorithm on dynamic protein-protein interaction networks. *Inf. Sci.* 329, 303–316. doi: 10.1016/j.ins.2015.09.028
- Lin, J. S., and Lai, E. M. (2017). Protein-protein interactions: co-immunoprecipitation. *Methods Mol Biol.* 1615, 211–219. doi: 10.1007/978-1-4939-7033-9\_17
- Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 173, 400–416. doi: 10.1158/1538-7445.AM2018-3287
- Mehla, J., Caufield, J. H., and Uetz, P. (2015). The yeast two-hybrid system: a tool for mapping protein-protein interactions. *Cold Spring Harb. Protoc.* 2015, 425–430. doi: 10.1101/pdb.top083345
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*. 5, 621. doi: 10.1038/nmeth.1226
- Planas-Iglesias, J., Marin-Lopez, M. A., Bonet, J., Garcia-Garcia, J., and Oliva, B. (2013). iLoops: a protein-protein interaction prediction server based on structural features. *Bioinformatics*. 29, 2360–2362. doi: 10.1093/bioinformatics/btt401
- Rao, V. S., Srinivas, K., Sujini, G. N., and Kumar, G. N. (2014). Protein-protein interaction detection: methods and analysis. *Int. J. Proteomics*. 2014:147648. doi: 10.1155/2014/147648

- Saccà, C., Teso, S., Diligenti, M., and Passerini, A. (2014). Improved multi-level protein-protein interaction prediction with semantic-based regularization. *BMC Bioinformatics* 15:103. doi: 10.1186/1471-2105-15-103
- Sun, T., Zhou, B., Lai, L., and Pei, J. (2017). Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC Bioinform.*, 18:277. doi: 10.1186/s12859-017-1700-2
- Sundell, G. N., and Ivarsson, Y. (2014). Interaction analysis through proteomic phage display. *Biomed. Res. Int.* 2014, 176172. doi: 10.1155/2014/176172
- Tina, K. G., Bhadra, R., and Srinivasan, N. (2007). PIC: protein interactions calculator. *Nucleic Acids Res.* 35, W473–W476. doi: 10.1093/nar/gkm423
- Viala, J., and Bouveret, E. (2017). Protein-protein interaction: tandem affinity purification in bacteria. *Methods Mol Biol.* 1615, 221–232. doi: 10.1007/978-1-4939-7033-9\_18
- Woloschuk, R. M., Reed, P. M. M., McDonald, S., Uppalapati, M., and Woolley, G. A. (2020). Yeast two hybrid screening of photo-switchable protein-protein interaction libraries. *J Mol Biol.* 432, 3113–3126. doi: 10.1016/j.jmb.2020.03.011
- Wu, F. L., Liu, Y., Jiang, H. W., Luan, Y. Z., Zhang, H. N., He, X., et al. (2017). The Ser/Thr protein kinase protein-protein interaction map of *M. tuberculosis*. *Mol Cell Proteomics*. 16: 1491–1506. doi: 10.1074/mcp.M116.065771
- Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M., and Eisenberg, D. (2002). DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 30: 303–305. doi: 10.1093/nar/30.1.303
- Yousef, A., and Moghadam Charkari, N. (2013). A novel method based on new adaptive LVQ neural network for predicting protein-protein interactions from protein sequences. *J. Theor. Biol.* 336, 231–239. doi: 10.1016/j.jtbi.2013.07.001
- Zeng, M., Zhang, F., Wu, F. X., Li, Y., Wang, J., and Li, M. (2020). Protein-protein interaction site prediction through combining local and global features with deep neural networks. *Bioinformatics*. 36, 1114–1120. doi: 10.1093/bioinformatics/btz699
- Zhang, C., Freddolino, P. L., and Zhang, Y. (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res.* 45, W291–W299. doi: 10.1093/nar/gkx366

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ding, Xie, Zhang, Shen, Zhong, Li, Shi, Chi and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# RNA-Binding Protein MSI2 Binds to miR-301a-3p and Facilitates Its Distribution in Mitochondria of Endothelial Cells

Qian Qian Guo<sup>1,2,3,4</sup>, Jing Gao<sup>1,2,3</sup>, Xiao Wei Wang<sup>1,2,3</sup>, Xian Lun Yin<sup>1,2,3</sup>, Shu Cui Zhang<sup>1,2,3</sup>, Xue Li<sup>5</sup>, Lian Li Chi<sup>5</sup>, Xiao Ming Zhou<sup>6,7</sup>, Zhe Wang<sup>8,9\*</sup> and Qun Ye Zhang<sup>1,2,3,4\*</sup>

<sup>1</sup> The Key Laboratory of Cardiovascular Remodeling and Function Research, Chinese Ministry of Education and Chinese National Health Commission, Shandong University, Jinan, China, <sup>2</sup> Department of Cardiology, Cheeloo College of Medicine, Qilu Hospital, Shandong University, Jinan, China, <sup>3</sup> The State and Shandong Province Joint Key Laboratory of Translational Cardiovascular Medicine, Jinan, China, <sup>4</sup> Cardiovascular Disease Research Center of Shandong First Medical University, Central Hospital Affiliated to Shandong First Medical University, Jinan, China, <sup>5</sup> National Glycoengineering Research Center, Shandong University, Qingdao, China, <sup>6</sup> Institute of Endocrinology, Shandong Academy of Clinical Medicine, Jinan, China, <sup>7</sup> Shandong Clinical Medical Center of Endocrinology and Metabolism, Jinan, China, <sup>8</sup> Division of Endocrinology and Metabolism, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China, <sup>9</sup> Division of Geriatrics, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Grzegorz Węgrzyn,  
University of Gdansk, Poland  
Anirban Chakraborty,  
University of Texas Medical Branch at  
Galveston, United States

### \*Correspondence:

Qun Ye Zhang  
wz.zhangqy@sdu.edu.cn  
Zhe Wang  
18615206628@163.com

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 24 September 2020

**Accepted:** 11 December 2020

**Published:** 21 January 2021

### Citation:

Guo QQ, Gao J, Wang XW, Yin XL,  
Zhang SC, Li X, Chi LL, Zhou XM,  
Wang Z and Zhang QY (2021)  
RNA-Binding Protein MSI2 Binds to  
miR-301a-3p and Facilitates Its  
Distribution in Mitochondria of  
Endothelial Cells.  
Front. Mol. Biosci. 7:609828.  
doi: 10.3389/fmolb.2020.609828

Numerous miRNAs have been detected in mitochondria, which play important roles in many physiological and pathophysiological processes. However, the dynamic changes of miRNA distribution in mitochondria and their mechanisms in reactive oxygen species (ROS)-induced endothelial injury remain unclear. Therefore, miRNA levels in whole cells and mitochondria of H<sub>2</sub>O<sub>2</sub>-treated endothelial cells were analyzed by small RNA sequencing in the present study. The results showed that H<sub>2</sub>O<sub>2</sub> significantly reduced the relative mitochondrial distribution of dozens of miRNAs in human umbilical vein endothelial cells (HUVECs). Among the high-abundance miRNAs, miR-301a-3p has the most significant changes in the redistribution between cytosol and mitochondria confirmed by absolute quantitative polymerase chain reaction (qPCR). To unravel the mechanism of miR-301a-3p distribution in mitochondria, RNA pull-down followed by label-free quantitative proteomic analysis was performed, and RNA-binding protein Musashi RNA binding protein 2 (MSI2) was found to specifically bind to miR-301a-3p. Western blotting and immunofluorescence colocalization assay showed that MSI2 was located in mitochondria of various cell types. H<sub>2</sub>O<sub>2</sub> significantly downregulated MSI2 expression in whole endothelial cells, promoted the distribution of MSI2 in cytosol and decreased its distribution in the mitochondria. Moreover, overexpression of MSI2 increased the mitochondrial distribution of miR-301a-3p, whereas inhibition of MSI2 decreased its distribution in mitochondria. Thus, MSI2 might be responsible for the distribution of miR-301a-3p between cytosol and mitochondria in endothelial cells. Our findings revealed for the first time that MSI2 was involved in the regulation of miRNA distribution in mitochondria and provided valuable insight into the mechanism of mitochondrial distribution of miRNAs.

**Keywords:** mitochondrial miRNA distribution, RNA-binding protein, Musashi RNA binding protein 2, reactive oxygen species, endothelial cell injury



## INTRODUCTION

It is believed that miRNAs regulate mRNA stability and translation primarily in the cytosol (Gebert and MacRae, 2019). However, a growing body of evidence indicated that miRNAs are also distributed in various subcellular regions, such as nucleus, processing bodies, endoplasmic reticulum and mitochondria (Leung, 2015; Trabucchi and Mategot, 2019). Subcellular distribution of miRNAs is indicative of regulatory role of miRNA in various cellular processes and it profoundly influenced many pathophysiological processes (Trabucchi and Mategot, 2019). To date, hundreds of mature miRNAs have been detected in mitochondria of various types of cells and tissues (Barrey et al., 2011; Sripada et al., 2012a). Many miRNAs located in mitochondria have been found to regulate many general and mitochondrial-specific biological processes (Borrallho et al., 2014). Moreover, the expression profile of miRNAs in mitochondria was dynamically changed under different pathophysiological conditions (Wang et al., 2015).

It is well-known that the genome of human mitochondria only encodes 13 proteins, 22 tRNAs, and 2 rRNAs (Anderson et al., 1981). Numerous nucleus-encoded proteins and ncRNAs must be imported into mitochondria via sophisticated transport systems to maintain proper mitochondrial function. To date, several proteins have been found to contribute to the transport of nucleus-encoded non-coding RNAs (ncRNAs) including tRNAs, rRNA, and RNA component of mitochondrial RNA processing endoribonuclease (MRP RNA) into mitochondria (Kim et al., 2017). The polynucleotide phosphorylase (PNPase) is also associated with the import of miRNA-378 into mitochondria (Shepherd et al., 2017). Considering the complexity of mitochondrial miRNAs and their transport system, it should be a reasonable assumption that many unknown players apart from PNPase must be involved in the mitochondrial distribution of miRNAs. However, the mechanisms regulating the distribution of miRNAs in mitochondria are still poorly understood.

In the present study, we found that H<sub>2</sub>O<sub>2</sub> noticeably reduced the levels of many miRNAs in mitochondria. In particular, the redistribution of miR-301a-3p between the cytosol and mitochondria was further confirmed by absolute quantitative polymerase chain reaction (qPCR). Moreover, our results demonstrated that Musashi RNA Binding Protein 2 (MSI2) could specifically bind to miR-301a-3p and facilitate its distribution in mitochondria. MSI2 downregulation may be responsible for the redistribution of miR-301a-3p in mitochondria of H<sub>2</sub>O<sub>2</sub>-treated endothelial cells.

## MATERIALS AND METHODS

### Cell Culture and Transfection

Human umbilical vein endothelial cells (HUVECs) were cultured in complete endothelial cell medium (#1001, ScienCell, Carlsbad, CA, USA) supplemented with 5% fetal bovine serum (FBS), 100 U/mL penicillin, 100 µg/mL streptomycin, and 1% endothelial cell growth supplement (ECGS) at 37°C in a humid atmosphere with 5% CO<sub>2</sub>. HepG2 cells and HEK293T cells were cultured in Dulbecco's modified Eagle's medium (#0030034DJ,

Gibco, Rockville, MD, USA) containing 10% FBS (#04-001-1A, Biological Industries, Kibbutz Beit Haemek, Israel), 100 U/mL penicillin, and 100 µg/mL streptomycin (#15070063, Gibco). Human arterial smooth muscle cells (HASMCs) were cultured in smooth muscle cell medium (SMCM), 1% SMCM growth supplements, 2% FBS, 100 U/mL penicillin, and 100 µg/mL streptomycin (#1101, ScienCell). The siRNAs targeting MSI2 gene were purchased from Ribobio (Guangzhou, China) and transfected at 50 nM using Lipofectamine RNAiMAX reagent (#13778030, Life Technologies, Carlsbad, CA, USA). The pCMV3-MSI2 expression plasmid was purchased from Sino Biological (#HG13069-NF, Beijing, China). Plasmid transfection was performed using Lipofectamine 3000 reagent (#L3000001, Life Technologies).

### Subcellular Fractionation

The mitochondrial and cytosolic fractions of HUVECs were isolated as described previously (Clayton and Shadel, 2014a,b). First, the cell pellet was washed twice in cold phosphate-buffered saline (PBS) and then resuspended in hypotonic buffer. Next, the cell suspensions were homogenized by a motor-driven Potter-Elvehjem homogenizer followed by centrifugation twice to collect the debris and nuclei. To collect the crude mitochondria, the supernatants were further centrifuged at 15,000 × g for 15 min at 4°C. The supernatants were collected as cytosolic fractions. Next, the crude mitochondria were layered on a discontinuous sucrose gradient (1.0M and 1.5M) and ultra-centrifuged at 60,000 × g for 40 min at 4°C. Then, the purified mitochondria were obtained from the interface between the two sucrose cushions. Next, the purified mitochondria were treated with RNase I (#AM2295, Life Technologies) to decontaminate the cytosolic RNA, followed by stopping the activity of RNase I using a ribonuclease inhibitor (RNasin, #N2615, Promega, Madison, WI, USA). The final mitochondrial pellet was resuspended in 100 µL RNAlater (#AM7020, Life Technologies) for miRNA quantitative real-time polymerase chain reaction (qRT-PCR) analysis and small RNA sequencing.

### Immunoprecipitation and Western Blotting

For immunoprecipitation, purified mitochondria from HUVECs were lysed in NP-40 buffer on ice for 20 min followed by centrifugation at 12,000 × g for 25 min. The mitochondrial lysate was then incubated with 2 µg anti-MSI2 antibody (#10770-1-AP, Proteintech, Wuhan, China) or anti-AGO2 antibody (#10686-1-AP, Proteintech) and 40 µL Protein A/G PLUS-Agarose beads (#sc-2003, Santa Cruz Biotechnology, Dallas, Texas, USA) overnight at 4°C with gentle rotation. The beads were washed and boiled in 2X SDS polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer. Non-immune IgG (#12-370, Millipore, Burlington, MA, USA) was used as a negative control. For western blotting, total and mitochondrial proteins were extracted and loaded on a 10% gel followed by gel electrophoresis. Subsequently, the separated proteins were transferred to a 0.2 µm polyvinylidene difluoride membrane (#1620177, Bio-Rad, Berkeley, CA, USA). Next, the membrane was blocked with 5% non-fat milk and incubated with primary antibodies overnight at 4°C with gentle rotation. After

washing, the membrane was incubated with the corresponding secondary antibodies at room temperature for 1 h followed by visualization using a ChemiDoc XRS+ imaging system (Bio-Rad). All antibodies used in this study were listed in **Supplementary Table 1**.

### Quantitative Real-Time PCR (qRT-PCR)

Total RNA in the whole cell, mitochondrial fraction, or cytosolic fraction was extracted using TRIzol reagent (#15596026, Life Technologies). To evaluate mRNA expression levels, cDNA synthesis and PCR amplification were performed using PrimeScript RT Master Mix and TB Green Premix Ex Taq II (#RR036A, RR820B, Takara, Kusatsu, Japan). The miR-301a-3p primer was purchased from Takara and its levels were determined using the Mir-X miRNA First-Strand Synthesis Kit and Mir-X miRNA qRT-PCR TB Green Kit (#638313, 638314, Clotech, Kusatsu, Japan) according to the manufacturer's instructions. MiRNA levels in the mitochondria were normalized to those of 12S rRNA. To absolutely quantify the amount of miR-301a-3p, a serial dilution of synthetic miR-301a-3p single-stranded RNA oligonucleotides (Ribobio) was used to generate the standard curve by qRT-PCR. The primers for mRNAs and miRNAs used in our study were listed in **Supplementary Table 2**.

### Small RNA Sequencing and Data Analysis

Small RNA sequencing was performed on an Illumina HiSeq X Ten (Illumina, San Diego, CA, USA; SE50 model) according to the manufacturer's instructions. After removing adapters and filtering low-quality reads, the high-quality clean data were mapped to the human genome sequence (hg19) and the unmapped reads were filtered. Then the mapped reads were aligned to the miRNA database using the R package bowtie2 (version 2.0.6). The sequencing data have been deposited in ArrayExpress (E-MTAB-9851).

### Mitotracker and Immunofluorescence Staining

Cells were treated with MitoTracker Red CMXRos (#M7512, Thermo Fisher, Waltham, MA, USA) to label mitochondria according to the manufacturer's instructions. After fixation and permeabilization, the cells were successively incubated with anti-MSI2 antibody (#10770-1-AP, Proteintech) and Alexa Fluor 488-conjugated secondary antibody (#SA00013-2, Proteintech). The nuclei were stained using 4',6-diamidino-2-phenylindole (DAPI, #10236276001, Sigma, China). Immunofluorescent images were obtained using a Zeiss LSM 710 confocal microscope (Zeiss, Oberkochen, Germany) and colocalization analysis was performed using Image-Pro Plus 6.0 software (Media Cybernetics, Rockville, MD, USA).

### RNA Immunoprecipitation

RNA immunoprecipitation was performed using the Magna RIP RNA-Binding Protein Immunoprecipitation Kit (#17-700, Millipore) following the manufacturer's instructions. Briefly, HUVECs were lysed in the ice-cold lysis buffer supplemented with protease inhibitors and recombinant RNase inhibitors. Five micrograms of anti-MSI2 antibody (#10770-1-AP, Proteintech)

or control Rabbit IgG (#12-370, Millipore) and 50  $\mu$ l magnetic beads were incubated in 0.5 ml wash buffer for 30 min to prepare magnetic beads for immunoprecipitation. Then prepared magnetic beads with 100  $\mu$ l cell lysate were suspended in 900  $\mu$ l immunoprecipitation buffer to immunoprecipitate RNA-binding protein-RNA complexes. After rotation at 4°C overnight, the beads containing RNA-binding protein-RNA complexes were washed total six times with 500  $\mu$ L of cold wash buffer. Then the proteinase K buffer was added into the tube and incubated at 55°C for 30 min with shaking to digest the protein, followed by the purification of RNA which was carried out by phenol, chloroform, Salt Solution I, Salt Solution II, Precipitate Enhancer and absolute ethanol. Finally, coimmunoprecipitated miRNAs were determined using the Mir-X miRNA First-Strand Synthesis Kit and Mir-X miRNA qRT-PCR TB Green<sup>®</sup> Kit (#638313, 638314, Clotech).

### Biotinylated miRNA Pull-Down Assay

HUVECs were collected and incubated with 100  $\mu$ l hypotonic buffer (10 mM Tris-Cl [pH 7.5], 20 mM KCl, 1.5 mM MgCl<sub>2</sub>, 5 mM DTT, 0.5 mM EGTA, 5% glycerol, 0.5% NP-40) supplemented with RNase inhibitors (#N2615, Promega) and protease inhibitors (#5892970001, Roche Applied Science, Penzberg, Germany) on ice for 5–10 min. Then, a homogenizer was used to mechanically disrupt the cell membrane. The cell homogenization was centrifuged at 2,000  $\times g$  at 4°C for 10 min to remove the nucleus, and then at 10,000  $\times g$  for 10 min at 4°C to further purify the supernatant. The protein concentration was determined using the BCA method. Then, 6  $\mu$ g of biotinylated RNA oligonucleotides ( $\sim 1$  nM) in 200  $\mu$ l high-salt wash buffer (HS-WB: 20 mM HEPES, pH 8.0, 300 mM KCl, 10 mM MgCl<sub>2</sub>, 0.01% NP-40, 1 mM DTT) was bound to 20  $\mu$ l streptavidin-agarose beads (#S1638, Millipore), previously blocked with 1 mg/mL yeast tRNA (#10109495001, Roche Applied Science), during a 3 h incubation at 4°C on a turning wheel. Then, streptavidin-agarose beads carrying 6  $\mu$ g of oligonucleotides were collected and washed three times in HS-WB buffer followed by incubation with 100  $\mu$ l cell lysate supernatant on a slowly rotating turning wheel for 30 min at room temperature, and then at 4°C for 2 h. The streptavidin-agarose beads were then washed four times with HS-WB containing increasing amounts of KCl (0.6, 0.8, 1.2, and 2 M) to reduce background noise. The specifically bound proteins were eluted with 100  $\mu$ l of 6 M urea with gentle shaking at room temperature for 30 min. The protein concentration of the elution was then determined by the Bradford method and the elution was freeze-dried for preservation until MS analysis was performed. The biotinylated RNA oligonucleotides used in the present study were shown in **Supplementary Table 3**.

### Label-Free Quantitative (LFQ) Proteomic Analysis

Protein samples were prepared for label-free quantitative (LFQ) proteomic analysis by nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS) using the filter-aided sample preparation (FASP) protocol as previously described with minor modifications (Wiśniewski et al., 2009). Briefly, after reduction

and alkylation, the protein samples were transferred to Microcon YM-10 filter units (#1602002vs, Sartorius Stedim Biotech, Göttingen, Germany) and centrifuged to remove detergent. Then, they were digested by 1:25 (w/w) trypsin (#T2600000, Sigma) at 37°C overnight. MS/MS analysis was performed on an EASY-nLC 1000 liquid chromatograph coupled with a Q Exactive Plus mass spectrometer (Thermo Scientific). Raw MS data were analyzed using Proteome Discoverer software 2.2 (Thermo Scientific) for protein identification and quantification according to the manual of this software. The detailed procedure was described in the **Supplemental Methods**.

### Bio-Layer Interferometry (BLI)

ForteBio Octet RED 96 (Forte Bio, Fremont, CA, USA) was used to perform the BLI assay. Briefly, streptavidin-coated biosensors (Forte Bio) were bound to the biotinylated RNA oligonucleotides and incubated with protein samples at different concentrations to measure the binding kinetics between the MSI2 protein and biotinylated RNA oligonucleotides. The detailed procedure was described in the **Supplemental Methods**.

### Annexin V/Propidium Iodide and EdU Staining

The apoptosis of HUVECs treated or untreated with hydrogen peroxide ( $H_2O_2$ ) was evaluated using the Annexin V/propidium iodide apoptosis detection kit (#556547, BD Pharmingen, San Jose, CA, USA) following the manufacturer's protocol. Briefly, HUVECs were seeded on 6-wells plates and treated with 100  $\mu M$   $H_2O_2$ . After treatment with  $H_2O_2$ , HUVECs were washed twice with ice-cold PBS and trypsinized to detach cells. Then the cells were resuspended and 5  $\mu l$  FITC Annexin V with 5  $\mu l$  propidium iodide were added. After incubating for 15 min at room temperature in the dark, the cells were analyzed by flow cytometry. In addition, an EdU staining assay (#ab219801, Abcam) was performed to assess cell proliferation. Briefly, after treatment with  $H_2O_2$ , 20 mM 5-ethynyl-2'-deoxyuridine (EdU) was added to the culture medium and incubated for 6 h at 37°C in the dark. The cells were then fixed with 4% paraformaldehyde (PFA), washed with 3% bovine serum albumin (BSA) and permeabilized using 0.5% Triton X-100. Then, EdU reaction cocktail, which was prepared according to the manufacturer's procedure was added and incubated for 30 min. Flow cytometry was used to analyze EdU staining.

### MTT Assay

The effect of glucose oxidase (GO) on endothelial cell viability was assessed by MTT assay. After treating HUVECs with different concentrations of GO for 24 h, MTT solution (5 mg/mL, #C0009S, Beyotime, Shanghai, China) was added to the culture medium and incubated for a further 4 h to generate an insoluble formazan crystal. Then the formazan crystal was solubilized in dimethyl sulfoxide (DMSO, #D8418, Sigma). The absorbance was measured at 570 nm in a microplate reader (SynergyH1 Hybrid Multi-Mode Reader, Biotek, Winooski, VT, USA).

### Statistical Analysis

Data normality was evaluated using Shapiro–Wilk test. The Wilcoxon rank-sum test was used for the comparison of two-group of non-normally distributed data. Two-tailed Student's *t*-tests and one-way analysis of variance (ANOVA) were used for comparison of two-group and multiple-group of normally distributed data, respectively. Statistical analysis was performed using GraphPad Prism v8.1.  $P < 0.05$  was considered statistically significant. All experiments were repeated independently at least three times.

## RESULTS

### ROS Leads to miR-301a-3p Redistribution Between Cytosol and Mitochondria in HUVECs

In order to evaluate the dynamic changes in the distribution of miRNAs in mitochondria during endothelial injury, HUVECs were treated with 100  $\mu M$   $H_2O_2$  for 12 and 18 h. Then, the miRNA profiling of mitochondria and the whole fraction of HUVECs was detected using small RNA sequencing. As shown in **Figure 1**,  $H_2O_2$  impaired HUVECs by promoting apoptosis and inhibiting proliferation (**Figure 1**). To ensure the reliability of the sequencing results, the purity of mitochondria was evaluated at the protein and mRNA levels. The results showed that a mitochondrial protein, translocase of outer mitochondrial membrane 40 (Tomm40), was strongly enriched while two cytosolic proteins,  $\beta$ -actin and ribosomal protein S9 (RPS9) were undetected in the mitochondrial fraction (**Supplementary Figure 1A**). In addition, the ratios of mtRNAs (*MT-ND4* and *12S rRNA*) to cytosolic (*GAPDH*, *18S rRNA*) or nuclear (*NEAT1*) RNAs in mitochondria were significantly higher than those in whole cells, which confirmed the high purity of the isolated mitochondria (**Supplementary Figure 1B**). Fifty-six percent of the mitochondria sequencing reads and 37% of the whole cell sequencing reads were aligned to miRNAs, respectively (**Figure 2A**). Total 193 high-abundance miRNAs (sequence reads  $> 100$  in all samples) were detected in both mitochondrial and total fractions from  $H_2O_2$ -treated HUVECs for 0, 12, and 18 h. Compared with untreated HUVECs, the abundances of many miRNAs in total and mitochondrial fractions were significantly changed in  $H_2O_2$ -treated HUVECs (**Figure 2B**). To illustrate the dynamic distribution changes of miRNAs in mitochondria, the relative mitochondrial distributions (Rm/t, ratio of miRNA levels in mitochondria to those in whole cells) of these 193 miRNAs in each group were calculated. The results showed that  $H_2O_2$  treatment significantly reduced the relative mitochondrial distribution of dozens of miRNAs in HUVECs (**Figure 2C**). Among them, miR-301a-3p was focused on because of its high abundance in mitochondria and the significant changes in its mitochondrial distribution. The results of qPCR further verified the decreased mitochondrial distribution of miR-301a-3p in HUVECs treated with  $H_2O_2$  for 12 and 18 h (**Figure 2D**). Moreover, the level of miR-301a-3p in mitochondria and cytosol of HUVECs treated or untreated with  $H_2O_2$  was absolutely quantified. The results demonstrated



that the amount of miR-301a-3p was markedly decreased in the mitochondria, whereas it was significantly increased in the cytosol of H<sub>2</sub>O<sub>2</sub>-treated HUVECs (**Figures 2E,F**). To further validate the ROS-induced redistribution of miR-301a-3p between the cytosol and mitochondria, an alternative ROS-generating reagent, glucose oxidase (GO) (Mueller et al., 2009), was used to treat HUVECs. The results showed that GO could induce a concentration-dependent damage to HUVECs (**Figure 2G**). Compared to the untreated HUVECs, GO significantly decreased the level of miR-301a-3p in the mitochondria and increased its cytosolic level (**Figures 2H,I**). These results indicated that miR-301a-3p was redistributed between cytosol and mitochondria during ROS-induced endothelial cell injury.

### RNA-Binding Protein (RBP) MSI2 Specifically Binds to miR-301a-3p

To reveal the mechanism underlying the mitochondrial distribution of miR-301a-3p, biotinylated miRNA pull-down was performed (**Figure 3A**). To clearly distinguish between the non-specific and specific binding of proteins to miR-301a-3p, a label-free quantitative proteomic analysis was carried out to compare the relative amount of proteins bound to miR-301a-3p and a random sequence after biotinylated miRNA pull-down. Several RNA-binding proteins were identified to specifically bind to miR-301a-3p in endothelial cells (**Supplementary Table 4**). Among these proteins, MSI2 was the most enriched protein because its relative amount in miR-301a-3p group was over 122 times compared to that in the control group (random small sequence) (**Figures 3B,C**). Moreover, MSI2 has been reported to bind to pri-miR-7-1 (Choudhury et al., 2013), implying the possibility of MSI2 binding to miR-301a-3p and facilitating its transport to mitochondria. Therefore, MSI2 was further studied by western blot analysis. The results confirmed that MSI2 was enriched in the product of miR-301a-3p pull-down, whereas it was undetectable in the control group (**Figure 3D**). To further validate the endogenous interaction between MSI2 and miR-301a-3p, RNA immunoprecipitation analysis was carried out using anti-MSI2 or normal IgG antibodies. The results showed that the level of miR-301a-3p in immunoprecipitation products using anti-MSI2 antibody was significantly higher than that in immunoprecipitation products using normal IgG, demonstrating the interaction between endogenous MSI2 and miR-301a-3p (**Figure 3E**). Moreover, the biolayer interferometry assay, which can quantitatively measure the strength of interactions between biomolecules, also showed a very low equilibrium dissociation constant (K<sub>d</sub>) of MSI2 and miR-301a-3p (25 nM), indicating a strong binding between them (**Figure 3F**). Therefore, these findings indicated that MSI2 specifically bound to miR-301a-3p in HUVECs.

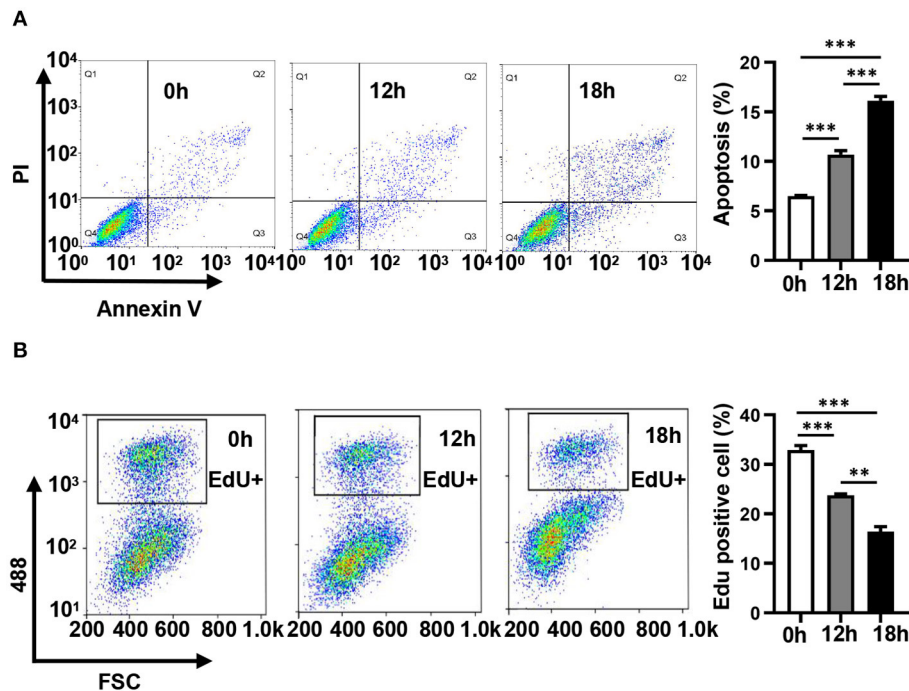
### MSI2 Localizes in Mitochondria and Facilitates the Mitochondrial Distribution of miR-301a-3p

To uncover the roles of MSI2 in the mitochondrial distribution of miRNAs, the localization of MSI2 in HUVECs was analyzed by immunoblot and immunofluorescence colocalization assays. The

results of immunoblot showed that MSI2 protein was abundant in mitochondria of HUVECs and its amount was comparable to that of cytochrome c (**Figure 4A**). Immunofluorescence colocalization also demonstrated high overlap coefficients of MSI2 and mitochondria in a variety of cells, indicating that MSI2 was also present in the mitochondria of these cells (**Figure 4B**). From these results, it is evident that the abundant mitochondrial distribution of MSI2 is common and not a cell-type-specific phenomenon. To further investigate the role of MSI2 in miR-301a-3p mitochondrial distribution, HUVECs were transfected, respectively, with MSI2 siRNAs, MSI2 expression vector or corresponding negative control sequences (control group) (**Supplementary Figures 1C,D**). The results showed that, compared to control group, the level of miR-301a-3p in the mitochondria was significantly reduced while its cytosolic level was significantly increased in HUVECs transfected with MSI2 siRNAs (**Figures 4C,D**). Moreover, MSI2 overexpression increased the level of miR-301a-3p in the mitochondria and reduced its level in the cytosol (**Figures 4E,F**). These findings demonstrated that MSI2 could facilitate the mitochondrial distribution of miR-301a-3p. Additionally, H<sub>2</sub>O<sub>2</sub> or GO treatment significantly decreased MSI2 levels in whole HUVECs and mitochondria but increased its cytosolic level, implying that MSI2 may be responsible for the H<sub>2</sub>O<sub>2</sub>- or GO-induced redistribution of miR-301a-3p in HUVECs (**Figures 4G,H**). Ago2 is an essential component of miRISC that has the ability to bind miRNAs, and has been detected in the mitochondria of many types of cells (Macgregor-Das and Das, 2018). Given the potential roles of Ago2 in mitochondrial distribution of miRNAs, Ago2 may interact with MSI2 in mitochondria. However, immunoprecipitation using the anti-Ago2 antibody or anti-MSI2 antibody in the mitochondrial pellet from HUVECs showed no interaction between Ago2 and MSI2 in mitochondria (**Figures 4I,J**). Together, the above results supported the conclusion that MSI2 could facilitate the distribution of miR-301a-3p in mitochondria.

## DISCUSSION

Although several studies have established the subcellular location of miRNAs in the mitochondria (Latronico and Condorelli, 2012; Sripada et al., 2012b), the underlying mechanism is still unclear. Polynucleotide phosphorylase (PNPase) located in the mitochondrial intermembrane space was reported to facilitate the trafficking of miR-378 to the mitochondrial matrix. Overexpressing PNPase significantly increased miR-378 levels in the mitochondria of HL-1 cells (Shepherd et al., 2017). In the present study, we did not detect PNPase in the pull-down materials using biotinylated miR-301a-3p. This suggested that PNPase might not be involved in trafficking nucleus-encoded miR-301a-3p to mitochondria. We found that MSI2 could specifically bind to miR-301a-3p and facilitate its mitochondrial distribution. However, it is still unknown whether MSI2 and PNPase synergistically participate in miRNA mitochondrial distribution.



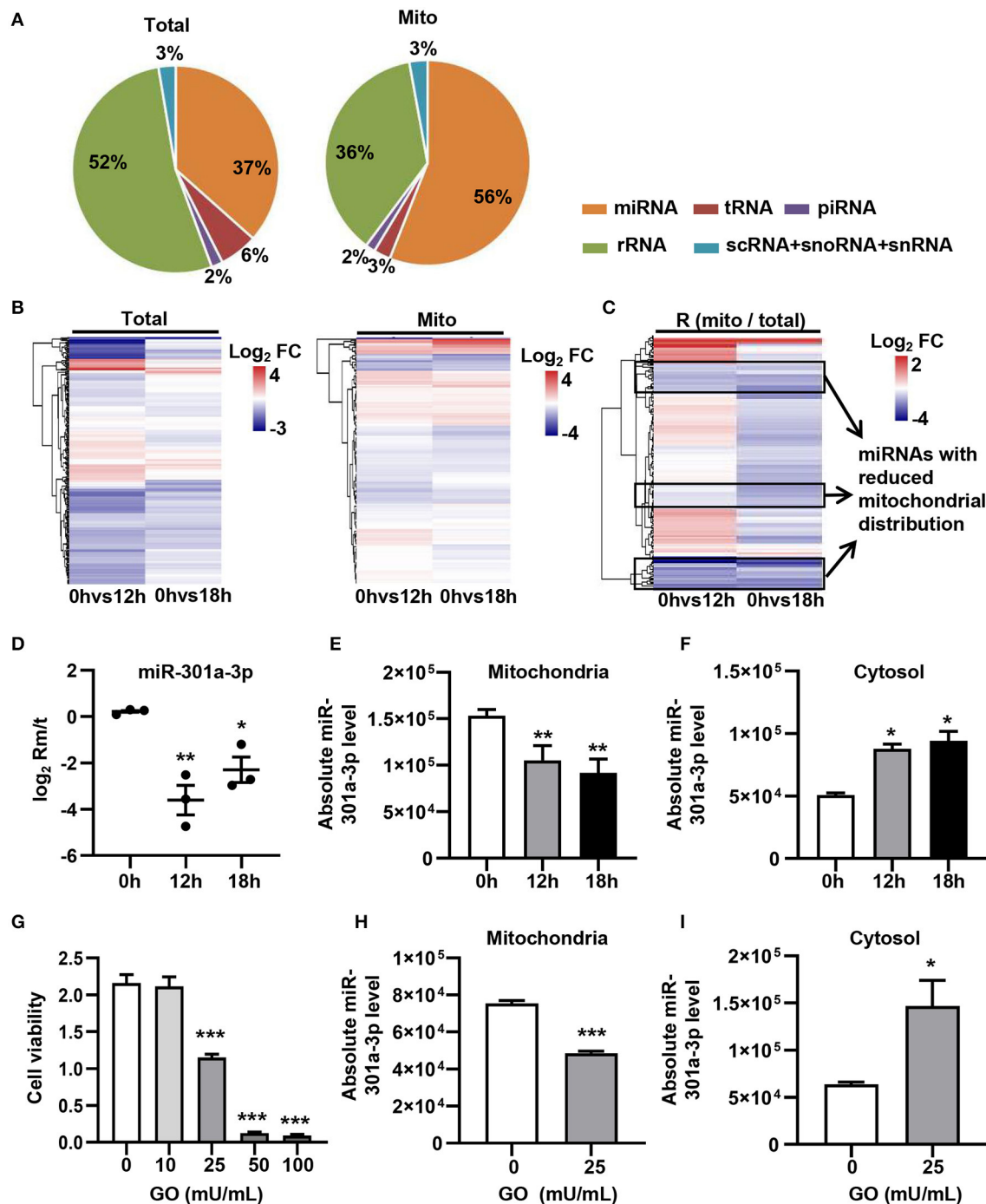
**FIGURE 1 |** Hydrogen peroxide induces apoptosis and inhibits the proliferation of HUVECs. **(A)** The apoptosis of HUVECs treated with 100  $\mu$ M  $H_2O_2$  for 0, 12, and 18 h was detected by Annexin V/ propidium iodide staining. **(B)** The proliferation of HUVECs treated with 100  $\mu$ M  $H_2O_2$  for 0, 12, and 18 h was detected by EdU-iFluor488. One-way analysis of variance (ANOVA) was used for statistical analysis. Data are presented as mean  $\pm$  SEM. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

MSI2 has important roles in maintaining stem cell populations and regulating cancer initiation, progression, and metastasis (Kharas and Lengner, 2017). As an evolutionarily highly conserved RNA binding protein (RBP), MSI2 binds to many mRNAs involved in numerous oncogenic processes and regulates their stability and protein translation (Kudinov et al., 2017). Recent studies have linked certain RBPs to miRNA sub-cellular distribution. Human antigen R (HuR) was reported to regulate the extracellular distribution of multivesicular body (MVB)-associated miRNAs, whereas SYNCRIP and hnRNP A2B1 modulate miRNAs sorting in exosomes (Villarroya-Beltri et al., 2013; Mukherjee et al., 2016; Santangelo et al., 2016). However, the role of RBPs in mitochondrial distribution of miRNAs was not previously reported in the literature. In this study, we found that MSI2 silencing led to a significant decrease in miR-301a-3p levels in mitochondria of HUVECs, whereas MSI2 overexpression promoted the mitochondrial distribution of miR-301a-3p, which extended the physiological function of MSI2 to regulate the mitochondrial distribution of miRNA. A study found that a conserved three-nucleotide core motif UAG defined the RNA binding specificity of MSI2 (Zearfoss et al., 2014). Interestingly, this UAG motif is found in the sequence of miR-301a-3p, implying that MSI2 might bind to miR-301a-3p by recognizing this motif. Moreover, MSI2 may also bind to other miRNAs containing UAG in their sequences. In fact, we found that among the miRNAs that were redistributed between the mitochondria and cytosol induced by  $H_2O_2$ , some miRNAs apart from miR-301a-3p also contained UAG motif in their sequences, such as miR-103b and miR-10b-5p. The results of

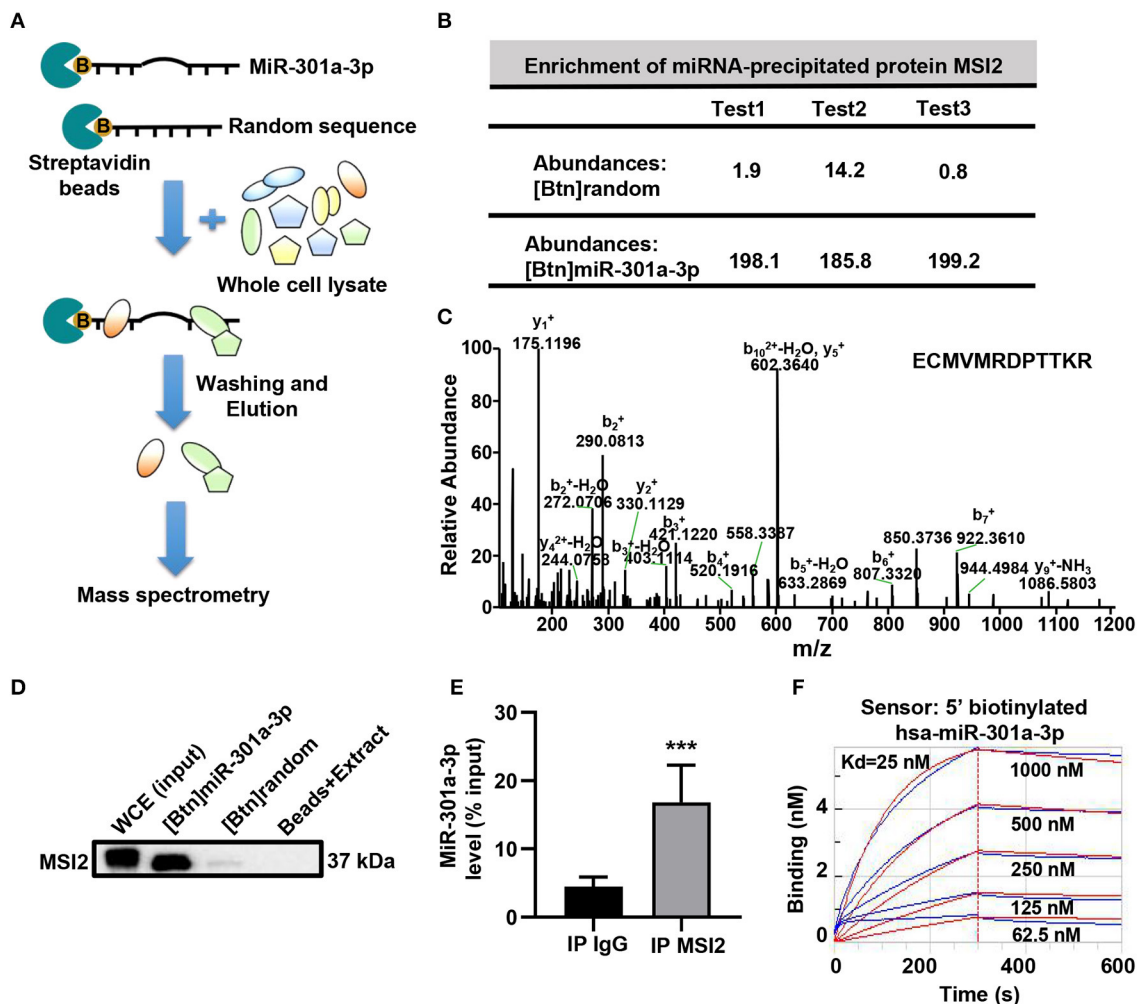
RNA immunoprecipitation with MSI2 antibody in HUVECs showed the high enrichment of miR-103b and miR-10b-5p in the immunoprecipitation products (**Supplementary Figure 2**). However, the molecular mechanisms underlying the binding of MSI2 to these miRNAs should be further studied.

Our findings demonstrated that  $H_2O_2$  or GO significantly reduced the expression of MSI2. However, the mechanism by which ROS downregulates MSI2 is unclear. In fact, the current understanding of the upstream signaling pathways that regulate MSI2 expression is still very limited. It has been reported that Krüppel-like factor 4 (KLF4) transcriptionally inhibits MSI2 expression by directly binding to the MSI2 promoter in multiple pancreatic ductal adenocarcinoma (PDAC) cell lines (Guo et al., 2017). However, we found that  $H_2O_2$  did not affect KLF4 expression in HUVECs, indicating that KLF4 might not be involved in the regulation of MSI2 by ROS (**Supplementary Figure 3**). Recently, ubiquitin carboxyl-terminal hydrolase 10 (USP10) was found to positively regulate MSI2 by post-transcriptional deubiquitination (Ouyang et al., 2019). Whether USP10 regulates the ubiquitination of MSI2 in our *in vitro* model should be further studied.

To the best of our knowledge, this study is the first to demonstrate that MSI2 interacts with miR-301a-3p and contributes to its distribution in mitochondria. Nevertheless, the detailed mechanism underlying the involvement of MSI2 in the mitochondrial distribution of miRNAs including miR-301a-3p is still unclear. However, the mechanisms by which many nucleus encoded RNAs apart from miRNAs are transported into the mitochondria have been studied. For example, 5S rRNA



**FIGURE 2 |** ROS redistributes miR-301a-3p between the cytosol and mitochondria. **(A)** The average frequency distribution of various non-coding RNA species identified in the whole cells (Total) and mitochondria (Mito) of HUVECs by small RNA sequencing. **(B)** Heatmap of the  $\log_2$  fold changes in miRNA levels identified in mitochondria (Mito) and whole cells (Total) between  $H_2O_2$ -treated HUVECs at 0, 12, or 18 h. **(C)** Heatmap of the  $\log_2$  fold changes in the relative mitochondrial distribution (R (mito / total)) of miRNAs between HUVECs treated with  $H_2O_2$  for 0, 12, or 18 h. **(D)** qRT-PCR analysis of the relative mitochondrial distribution ( $\log_2$  Rm/t) of miR-301a-3p in HUVECs treated with  $H_2O_2$  for 0, 12, and 18 h. **(E,F)** Absolute quantification of miR-301a-3p levels in mitochondria **(E)** and cytosol **(F)** of the  $H_2O_2$ -treated HUVECs for 0, 12, and 18 h. **(G)** The viability of HUVECs treated with glucose oxidase (GO) at different concentrations for 24 h estimated by MTT assay. **(H,I)** Absolute quantification of miR-301a-3p levels in mitochondria **(H)** and cytosol **(I)** of HUVECs treated with 25  $\mu$ M GO for 24 h. One-way analysis of variance (ANOVA) was used for **(D–G)**. Two-tailed Student's *t*-test was used for statistical analysis in H and I. Data are presented as mean  $\pm$  SEM. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.

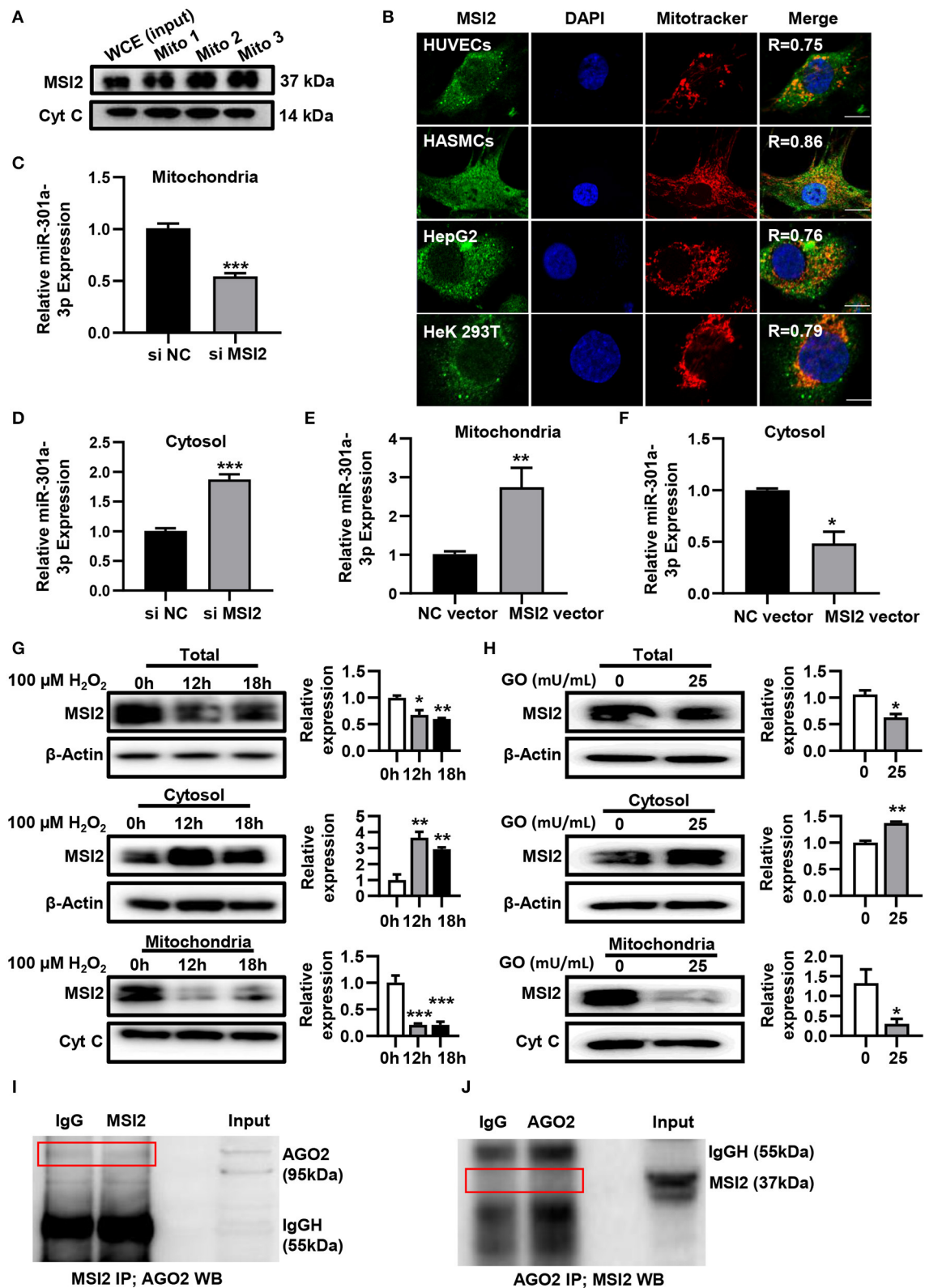


**FIGURE 3 |** MSI2 binds to miR-301a-3p. **(A)** Schematic representation of biotinylated miRNA pull-down assay followed by label-free quantitative proteomic analysis. 5'biotinylated miR-301a-3p or random sequence (control group) was incubated with streptavidin beads and whole cell extracts of HUVECs. After washing and elution, the pull-down proteins were analyzed by MS/MS. **(B)** Relative abundances of MSI2 protein identified in three replicate experiments of label-free quantitative proteomic analysis for the product of biotinylated miR-301a-3p pull-down [(Bt) miR-301a-3p]. Biotinylated random small sequence [(Bt) random] was used as a negative control. **(C)** Representative MS/MS spectrum of a parent ion of MSI2 protein in the label-free quantitative proteomic analysis. The peptide sequence identified by this spectrum was shown in the upper right side. **(D)** Western blotting analysis of MSI2 protein in the product of the biotinylated miR-301a-3p pull-down assay. WCE: whole cell extracts; Beads+Extract: the mixture of streptavidin-agarose beads and whole cell extracts of HUVECs. **(E)** miR-301a-3p level in the immunoprecipitation products obtained by incubating anti-MSI2 antibody with the HUVECs lysates. The result was reported as percentage of the input sample (% input). The normal IgG was used as control (IP IgG). **(F)** The representative sensorgram of bio-layer interferometry (BLI) analysis for the binding kinetics of MSI2 and 5' biotinylated miR-301a-3p. The blue curves represented the measured responses for each tested concentration of MSI2 protein. The overlapped red curves showed the global fitting results of the binding data. Wilcoxon rank-sum test was used for statistical analysis in **(E)**. Median with interquartile range was shown for **(E)**. \*\*\* $P < 0.001$ .

was reported to interact with the precursor of mitochondrial ribosomal protein L18 (MRPL18) in the cytosol, causing a conformational change in 5S rRNA that makes it recognized by Rhodanese and translocated into mitochondria (Smirnov et al., 2010, 2011). Obviously, it is conceivable that other proteins apart from MSI2 should also participate in miR-301a-3p mitochondrial import. In the pull-down materials using biotinylated miR-301a-3p, except for MSI2, we also found another interesting RBP, alpha-enolase. Enolase has been reported to be associated with the mitochondrial transport of yeast tRNA tRK1. The yeast tRNA tRK1 could bind to

the cytosolic enolase (ENO2P) and the precursor of the mitochondrial lysyl-tRNA synthetase (preMSK or pre-LysRS) to form a complex on the mitochondrial membrane surface. Then, this complex was internalized into the mitochondrial matrix via the TOM/TIM protein transport system (Gowher et al., 2013). In addition, MSI2 has been demonstrated to interact with the translocase of inner mitochondrial membrane 10 homolog B (Huttlín et al., 2017), implying that MSI2-miRNA complex might be translocated into mitochondria via the mitochondrial transport system for proteins. Obviously, whether or not these above-mentioned proteins are involved in the mitochondrial





**FIGURE 4 |** MSI2 exists in mitochondria and facilitates the distribution of miR-301a-3p in mitochondria. **(A)** Western blotting analysis of mitochondrial fractions of HUVECs for MSI2 and Cytochrome C. Mito: mitochondria; WCE: whole cell extracts. **(B)** Fluorescence colocalization of MSI2 and mitochondria in different types of cells. *(Continued)*



**FIGURE 4 |** cells. MSI2 was labeled with its Alexa Fluor 488-conjugated antibody (green). Mitochondria and nuclei were stained, respectively, with MitoTracker (red) and DAPI (blue). Yellow areas in the merged images represented the colocalization of MSI2 and mitochondria. R value represented the Mander's overlap coefficient between MSI2 and mitochondria calculated by Image-Pro Plus 6.0 software. Bar=10  $\mu$ m. **(C,D)** The relative levels of miR-301a-3p in mitochondria **(C)** and cytosol **(D)** of HUVECs transfected with MSI2 siRNA compared to those in the control group. The 12S rRNA (mitochondria) and GAPDH (cytosol) were used as internal standards. **(E,F)** The relative levels of miR-301a-3p in mitochondria **(E)** and cytosol **(F)** of HUVECs transfected with MSI2 expression vector compared to those in the control group. **(G)** Western blotting analysis of MSI2 levels in the whole, cytosolic and mitochondrial fractions of HUVECs treated with H<sub>2</sub>O<sub>2</sub> for 0, 12, and 18 h. **(H)** Western blotting analysis of MSI2 levels in the whole, cytosolic and mitochondrial fractions of HUVECs treated with GO for 24 h. **(I,J)** Immunoprecipitation analysis of the binding of Ago2 and MSI2 in HUVECs mitochondria. The immune complexes were formed by incubating mitochondrial lysates with anti-MSI2 (MSI2 IP) and then immunoblotted with anti-Ago2 antibody **(I)**, or by incubating mitochondrial lysates with anti-Ago2 (Ago2 IP) and then immunoblotted with anti-MSI2 antibody **(J)**. Mitochondrial lysates were used as input sample and normal IgG was used as the negative control (IgG). Two-tailed Student's *t*-test was used for statistical analysis in **(C–F,H)**. One-way analysis of variance (ANOVA) was used for **(G)**. Data are presented as mean  $\pm$  SEM. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.

transport of miR-301a-3p is an interesting issue that deserves to be investigated.

A growing body of evidence has suggested that the mitochondrial miRNA dynamic distribution is closely associated with the pathologies of many diseases (Borrallho et al., 2014; Song et al., 2019). For example, miR-378 could redistribute into the interfibrillar mitochondria and regulate mitochondria encoded protein ATP6 in the diabetic heart (Jagannathan et al., 2015). Our study demonstrated that miR-301a-3p redistributed between the mitochondria and cytosol during the process of ROS-induced endothelial injury. It is well-known that endothelial injury plays an important role in the development of many cardiovascular diseases (Lerman and Zeiher, 2005). Moreover, miR-301a-3p has also been reported to induce HUVECs apoptosis and increase endothelial barrier permeability (Liu et al., 2020). Therefore, it might be assumed that the mitochondria-localized miR-301a-3p could rapidly re-enter into the cytosol and regulate its target genes, thus affecting endothelial function under oxidative stress conditions. In line with this assumption, the stored miRNAs in processing bodies have been evidenced to be delocalized and mediated translational repression when amino acid starvation occurs (Bhattacharyya et al., 2006). Nevertheless, more *in vivo* and *in vitro* studies are needed to investigate the pathological meaning of MSI2-mediated redistribution of miR-301a-3p in cardiovascular diseases.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## REFERENCES

- Anderson, S., Bankier, A. T., Barrell, B. G., Bruijn, M. H., Coulson, A. R., Drouin, J., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457–465. doi: 10.1038/290457a0
- Barrey, E., Saint-Auret, G., Bonnamy, B., Damas, D., Boyer, O., and Gidrol, X. (2011). Pre-microRNA and mature microRNA in human mitochondria. *PLoS ONE* 6:e20220. doi: 10.1371/journal.pone.0020220
- Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I., and Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125, 1111–1124. doi: 10.1016/j.cell.2006.04.031

## AUTHOR CONTRIBUTIONS

QZ and ZW conceived the original idea and participated in experimental design. QG, XW, SZ, JG, XY, and XL performed the experiments. QG, QZ, and SZ analyzed the data. QG and QZ wrote the manuscript. ZW, LC, and XZ revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China [NO. 81670247 and 82070820]; the Natural Science Outstanding Youth Foundation of Shandong Province [NO. JQ201519]; Major Science and Technology Innovation Project of Shandong Province [NO. 2018CXGC1218]; Jinan Clinical Medical Science and Technology Innovation Program [NO. 201805055]; Taishan Scholar program of Shandong Province [TS201712092]; and Natural Science Foundation of Shandong Province [ZR2019MH025].

## ACKNOWLEDGMENTS

The authors would like to thank Xinjie Zhang for her help in the experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.609828/full#supplementary-material>

- Borrallho, P. M., Rodrigues, C. M. P., and Steer, C. J. (2014). Mitochondrial MicroRNAs and their potential role in cell function. *Curr. Pathobiol. Rep.* 2, 123–132. doi: 10.1007/s40139-014-0047-x
- Choudhury, N. R., Alves, F. L., Andrés-Aguayo, L., Graf, T., Cáceres, J. F., Rappsilber, J., et al. (2013). Tissue-specific control of brain-enriched miR-7 biogenesis. *Genes Dev.* 27, 24–38. doi: 10.1101/gad.199190.112
- Clayton, D. A., and Shadel, G. S. (2014a). Purification of mitochondria by sucrose step density gradient centrifugation. *Cold Spring Harb. Protoc.* 10:pdb-prot080028. doi: 10.1101/pdb-prot080028
- Clayton, D. A., and Shadel, G. S. (2014b). Isolation of mitochondria from tissue culture cells. *Cold Spring Harb. Protoc.* 10:pdb-prot080002. doi: 10.1101/pdb-prot080002

- Gebert, L. F. R., and MacRae, I. J. (2019). Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* 20, 21–37. doi: 10.1038/s41580-018-0045-7
- Gowher, A., Smirnov, A., Tarassov, I., and Entelis, N. (2013). Induced tRNA import into human mitochondria: implication of a host aminoacyl-tRNA-synthetase. *PLoS ONE* 8:e66228. doi: 10.1371/journal.pone.0066228
- Guo, K., Cui, J., Quan, M., Xie, D., Jia, Z., Wei, D., et al. (2017). The Novel KLF4/MSI2 signaling pathway regulates growth and metastasis of pancreatic cancer. *Clin. Cancer Res.* 23, 687–696. doi: 10.1158/1078-0432.CCR-16-1064
- Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* 545, 505–509. doi: 10.1038/nature22366
- Jagannathan, R., Thapa, D., Nichols, C. E., Shepherd, D. L., Stricker, J. C., Croston, T. L., et al. (2015). Translational regulation of the mitochondrial genome following redistribution of mitochondrial microRNA in the diabetic heart. *Circ. Cardiovasc. Genet.* 8:785–802. doi: 10.1161/CIRCGENETICS.115.001067
- Kharas, M. G., and Lengner, C. J. (2017). Stem cells, cancer, and MUSASHI in blood and guts. *Trends Cancer* 3, 347–356. doi: 10.1016/j.trecan.2017.03.007
- Kim, K. M., Noh, J. H., Abdelmohsen, K., and Gorospe, M. (2017). Mitochondrial noncoding RNA transport. *BMB Rep.* 50, 164–174. doi: 10.5483/BMBRep.2017.50.4.013
- Kudinov, A. E., Karanicolas, J., Golemis, E. A., and Bumber, Y. (2017). Musashi RNA-binding proteins as cancer drivers and novel therapeutic targets. *Clin. Cancer Res.* 23, 2143–2153. doi: 10.1158/1078-0432.CCR-16-2728
- Latronico, M. V., and Condorelli, G. (2012). The might of microRNA in mitochondria. *Circ. Res.* 110, 1540–1542. doi: 10.1161/CIRCRESAHA.112.271312
- Lerman, A., and Zeiher, A. M. (2005). Endothelial function: cardiac events. *Circulation* 111, 363–368. doi: 10.1161/01.CIR.0000153339.27064.14
- Leung, A. K. L. (2015). The whereabouts of microRNA actions: cytoplasm and beyond. *Trends Cell Biol.* 25, 601–610. doi: 10.1016/j.tcb.2015.07.005
- Liu, L. Y., Yin, H. N., Hao, X. X., Song, H. F., Chai, J. K., Duan, H. J., et al. (2020). Down-Regulation of miR-301a-3p reduces burn-induced vascular endothelial apoptosis by potentiating hMSC-secreted IGF-1 and PI3K/Akt/FOXO3a pathway. *iScience* 23:101383. doi: 10.1016/j.isci.2020.101383
- Macgregor-Das, A. M., and Das, S. (2018). A microRNA's journey to the center of the mitochondria. *Am. J. Physiol. Heart Circ. Physiol.* 315, H206–H215. doi: 10.1152/ajpheart.00714.2017
- Mueller, S., Millonig, G., and Waite, G. N. (2009). The GOX/CAT system: a novel enzymatic method to independently control hydrogen peroxide and hypoxia in cell culture. *Adv. Med. Sci.* 54, 121–135. doi: 10.2478/v10039-009-0042-3
- Mukherjee, K., Ghoshal, B., Ghosh, S., Chakrabarty, Y., Shwetha, S., and Das, S. (2016). Reversible HuR-microRNA binding controls extracellular export of miR-122 and augments stress response. *PLoS ONE* 11, 1184–1203. doi: 10.15252/embr.201541930
- Ouyang, S., Liu, T., Liu, X., Zhu, F., Liu, X., et al. (2019). USP10 regulates Musashi-2 stability via deubiquitination and promotes tumour proliferation in colon cancer. *FEBS Lett.* 593, 406–413. doi: 10.1002/1873-3468.13323
- Santangelo, L., Giurato, G., Cicchini, C., Montaldo, C., Mancone, C., Tarallo, R., et al. (2016). The RNA-binding protein SYNCRIP is a component of the hepatocyte exosomal machinery controlling microRNA sorting. *Cell Rep.* 17, 799–808. doi: 10.1016/j.celrep.2016.09.031
- Shepherd, D. L., Hathaway, Q. A., Pinti, M. V., Nichols, C. E., Durr, A. J., Sreekumar, S., et al. (2017). Exploring the mitochondrial microRNA import pathway through polynucleotide phosphorylase (PNPase). *J. Mol. Cell. Cardiol.* 110, 15–25. doi: 10.1016/j.yjmcc.2017.06.012
- Smirnov, A., Comte, C., Mager-Heckel, A. M., Addis, V., Krashenninnikov, I. A., Martin, R. P., et al. (2010). Mitochondrial enzyme rhodanese is essential for 5S ribosomal RNA import into human mitochondria. *J. Biol. Chem.* 285, 30792–30803. doi: 10.1074/jbc.M110.151183
- Smirnov, A., Entelis, N., Martin, R. P., and Tarassov, I. (2011). Biological significance of 5S rRNA import into human mitochondria: role of ribosomal protein MRP-L18. *Genes Dev.* 25, 1289–1305. doi: 10.1101/gad.624711
- Song, R., Hu, X. Q., and Zhang, L. (2019). Mitochondrial MiRNA in cardiovascular function and disease. *Cells* 8:1475. doi: 10.3390/cells8121475
- Sripada, L., Tomar, D., Prajapati, P., Singh, R., Singh, A. K., and Singh, R. (2012b). Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. *PLoS ONE* 7:e44873. doi: 10.1371/journal.pone.0044873
- Sripada, L., Tomar, D., and Singh, R. (2012a). Mitochondria: one of the destinations of miRNAs. *Mitochondrion* 12, 593–599. doi: 10.1016/j.mito.2012.10.009
- Trabucchi, M., and Mategot, R. (2019). Subcellular heterogeneity of the microRNA machinery. *Trends Genet.* 36:70. doi: 10.1016/j.tig.2019.07.008
- Villarroya-Beltri, C., Gutiérrez-Vázquez, C., Sánchez-Cabo, F., Pérez-Hernández, D., Vázquez, J., Martín-Cofreces, N., et al. (2013). Sumoylated hnRNP A2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat. Commun.* 4:2980. doi: 10.1038/ncomms3980
- Wang, W., Visavadiya, N. P., Pandya, J. D., Nelson, P. T., Sullivan, P. G., and Springer, J. E. (2015). Mitochondria-associated microRNAs in rat hippocampus following traumatic brain injury. *Exp. Neurol.* 265, 84–93. doi: 10.1016/j.expneurol.2014.12.018
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322
- Zearfoss, N. R., Deveau, L. M., Clingman, C. C., Schmidt, E., Johnson, E. S., and Massi, F. (2014). A conserved three-nucleotide core motif defines Musashi RNA binding specificity. *J. Biol. Chem.* 289, 35530–35541. doi: 10.1074/jbc.M114.597112

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Guo, Gao, Wang, Yin, Zhang, Li, Chi, Zhou, Wang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Elucidating the Interactions Between Heparin/Heparan Sulfate and SARS-CoV-2-Related Proteins—An Important Strategy for Developing Novel Therapeutics for the COVID-19 Pandemic

Mingjia Yu <sup>1†</sup>, Tianji Zhang <sup>2†</sup>, Wei Zhang <sup>2</sup>, Qianyun Sun <sup>3</sup>, Hongmei Li <sup>2\*</sup> and Jin-ping Li <sup>1,4\*</sup>

<sup>1</sup> Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing, China, <sup>2</sup> Division of Chemistry and Analytical Science, National Institute of Metrology, Beijing, China, <sup>3</sup> Division of Chemistry, Shandong Institute of Metrology, Jinan, China, <sup>4</sup> Department of Medical Biochemistry and Microbiology, University of Uppsala, Uppsala, Sweden

## OPEN ACCESS

### Edited by:

Lianli Chi,

Shandong University, China

### Reviewed by:

Chao Cai,

Ocean University of China, China

Xudong Qu,

Shanghai Jiao Tong University, China

### \*Correspondence:

Jin-ping Li

jin-ping.li@imbim.uu.se

Hongmei Li

lih@nim.ac.cn

<sup>†</sup> These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 12 November 2020

**Accepted:** 16 December 2020

**Published:** 25 January 2021

### Citation:

Yu M, Zhang T, Zhang W, Sun Q, Li H  
and Li J-p (2021) Elucidating the  
Interactions Between  
Heparin/Heparan Sulfate and  
SARS-CoV-2-Related Proteins—An  
Important Strategy for Developing  
Novel Therapeutics for the COVID-19  
Pandemic.  
Front. Mol. Biosci. 7:628551.  
doi: 10.3389/fmolb.2020.628551

Owing to the high mortality and the spread rate, the infectious disease caused by SARS-CoV-2 has become a major threat to public health and social economy, leading to over 70 million infections and 1.6 million deaths to date. Since there are currently no effective therapeutic or widely available vaccines, it is of urgent need to look for new strategies for the treatment of SARS-CoV-2 infection diseases. Binding of a viral protein onto cell surface heparan sulfate (HS) is generally the first step in a cascade of interaction that is required for viral entry and the initiation of infection. Meanwhile, interactions of selectins and cytokines (e.g., IL-6 and TNF- $\alpha$ ) with HS expressed on endothelial cells are crucial in controlling the recruitment of immune cells during inflammation. Thus, structurally defined heparin/HS and their mimetics might serve as potential drugs by competing with cell surface HS for the prevention of viral adhesion and modulation of inflammatory reaction. In this review, we will elaborate coronavirus invasion mechanisms and summarize the latest advances in HS-protein interactions, especially proteins relevant to the process of coronavirus infection and subsequent inflammation. Experimental and computational techniques involved will be emphasized.

**Keywords:** heparin, heparan sulfate, COVID-19, coronavirus, interaction

## INTRODUCTION

Currently, the whole world is facing the deadly coronavirus disease 2019 (COVID-19) outbreak caused by the coronavirus (CoV) SARS-CoV-2, which has been far beyond the outbreaks caused by the other two major coronaviruses (SARS and MERS) in the past 20 years (Drosten et al., 2003; Zaki et al., 2012). So far, there are no specific therapeutic and effective drugs available. Vaccines, although achieving success worldwide, are still far from being widely accessible. In this regard, multidimensional antiviral strategies are strongly needed in preventing the spread of COVID-19 and treating infected individuals.

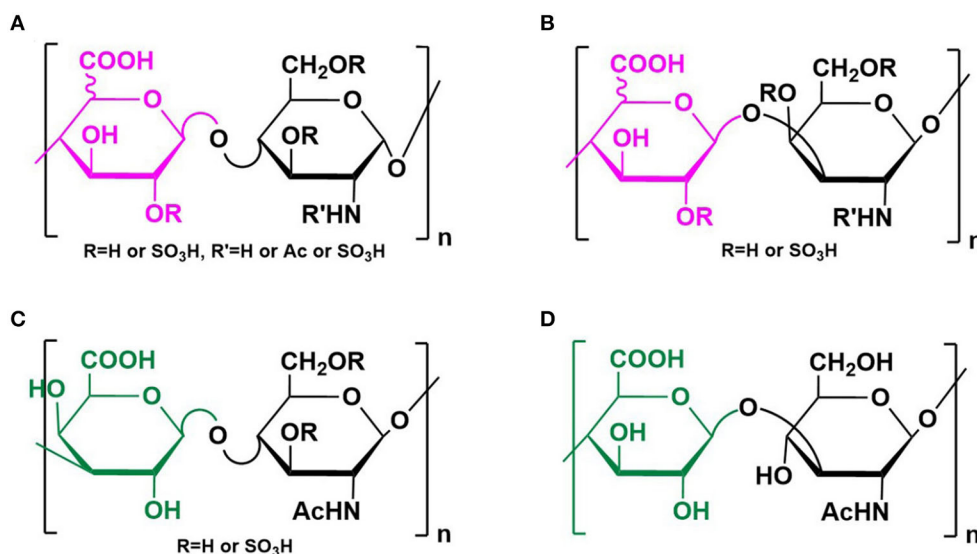
Glycosaminoglycans (GAGs) are a group of anionic polysaccharides composed of repeating disaccharide building blocks, including heparin/heparan sulfate (HS) (-4GlcA $\beta$ /IdoA $\alpha$ 1-4GlcNAc $\alpha$ 1-, x = Ac, SO<sub>3</sub>H or H), chondroitin/dermatan sulfate (-4GlcA $\beta$ /IdoA $\alpha$ 1-3GalNAc $\beta$ 1-), keratan sulfate (-3GalA $\beta$ 1-4GlcNAc $\beta$ 1-), and hyaluronic acid (HA) (-4GlcA $\beta$ 1-3GlcNAc $\beta$ 1-) (Figure 1). Sulfation at various positions of the sugar residues could occur except for the HA, making their structures heterogeneous and extremely difficult to characterize (Tianji Zhang et al., 2019). Among them, heparin and HS exhibit the most diverse biological activities, most of which are mediated by their interactions with proteins (Li and Kusche-Gullberg, 2016). Recent work identified HS on the cell surface as a co-receptor for the SARS-CoV-2 spike protein (S protein) (Clausen et al., 2020), making the HS-S protein interaction an extremely appealing target for manipulating SARS-CoV-2 infection. Designing competing HS mimetics requires the elucidation of the mode of interaction, particularly sequence specificities of the HS.

In this review, we will summarize the classification and invasion mechanisms of the major CoVs and elaborate possible antiviral strategies based on the interactions between heparin/HS and proteins. Relevant technologies involved in elucidating heparin/HS-protein interactions are crucial for developing (sequence) specific antiviral molecules and will thus be underlined.

## CLASSIFICATION AND INVASION MECHANISMS OF THE CORONAVIRUS

CoVs are highly diverse, enveloped, and positive-sense single-stranded (up to 30,000 bp) RNA viruses (Coutard et al.,

2020) belonging to the *Nidovirales* order in the subfamily of *Othocoronavirinae* (Wang et al., 2020). Infection by the viruses can cause severe diseases affecting upper respiratory, gastrointestinal, and central nervous systems in humans and other animals (Gallagher and Buchmeier, 2001). Based on systematic analysis of viral nucleic acid sequence, CoVs can be classified into four genera: alpha, beta, gamma, and delta according to the 10th Report on Virus Taxonomy from the International Committee on Taxonomy of Viruses (ICTV) (Fehr and Perlman, 2015). Among them, alpha- and beta-coronaviruses can infect mammals, gamma-coronaviruses can infect avian species, while delta-coronaviruses can infect both (Li, 2016). Currently, there are seven representative strains of human coronaviruses (HCoVs) including four low-pathogenic coronaviruses [HCoV-229E, HCoV-NL63 (alpha-coronaviruses), HCoV-OC43, and HCoV-HKU1 (beta-coronaviruses)], which cause mild respiratory diseases in humans (Su et al., 2016), and three high pathogenic strains including HCoVs {severe acute respiratory syndrome coronavirus (SARS-CoV) (Drosten et al., 2003), Middle East respiratory syndrome coronavirus (MERS-CoV) (Elfiky et al., 2017), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (beta-coronaviruses) (Hui et al., 2020)}. The three highly pathogenic strains have caused deadly pneumonia in humans since the beginning of the twenty-first century. Unfortunately, so far, there are still no specific therapeutics approved against these human-infecting coronaviruses, mainly due to lacking sufficient knowledge in the pathological process of viral infection. Thus, in-depth understanding of the infection mechanisms will facilitate the development of effective interventions against these highly pathogenic coronaviruses and are of high urgency for the control and treatment of COVID-19.

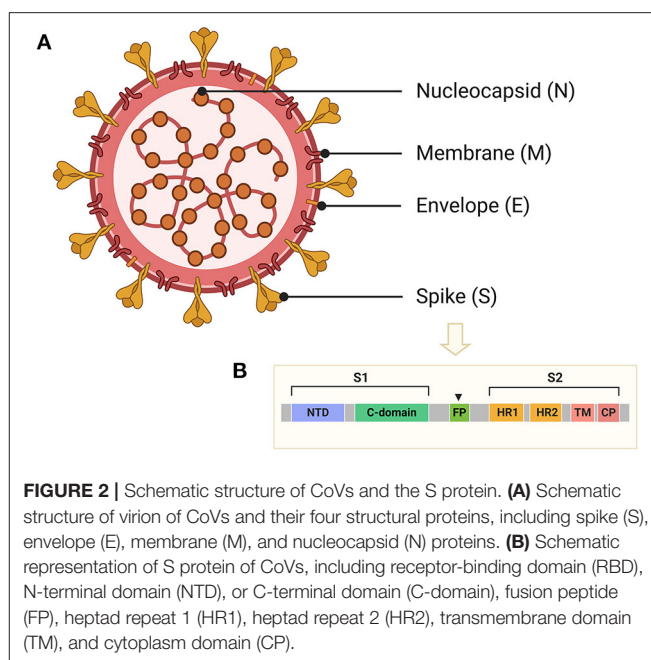


**FIGURE 1 |** Structure of GAGs. Structures of heparin/heparan sulfate (A), chondroitin/dermatan sulfate (B), keratan sulfate (C), and hyaluronic acid (D) were illustrated. Reprinted from ref Tianji Zhang et al. (2019) with permission.



CoVs share similar genome identities. Two-thirds of the genome at the 5'-terminus contain two large overlapping open reading frames (ORFs), ORF 1a and ORF 1b, which encode polyproteins 1a (pp1a) and pp1b/1ab, respectively. The polyproteins can be further cleaved into 15–16 non-structural proteins (nsp2–nsp16 or nsp1–nsp16). One-third of the genome at the 3'-terminus encodes four common structural proteins in the order of Spike (S) that characterizes all coronaviruses, Envelope (E), Membrane (M), and Nucleocapsid (N) (Wang et al., 2020) (**Figure 2A**). The S protein is a trimeric class I fusion protein that protrudes from the virion surface and mediates receptor recognition, membrane fusion, virus entry, and antibody neutralization (Gallagher and Buchmeier, 2001). Considering its significant functions during viral infection (Liu et al., 2004), the S protein serves as a main target for the development of antiviral drugs (Du et al., 2017), antibodies (He et al., 2006), entry inhibitors (Lu et al., 2014), and vaccines (Du et al., 2009). Each monomer of the trimeric S protein is ~180 kDa containing two subunits—a receptor-binding subunit (S1) and a membrane-fusion subunit (S2), which are linked through a fusion peptide. The S1 subunit contains two independent domains—the N-terminal domain (NTD) and the C-terminal domain (C-domain) (Ou et al., 2020) (**Figure 2B**), either of which can serve as the receptor-binding domain (RBD) depending on the virus strains (Kubo et al., 1994; Ou et al., 2017). The S2 subunit consists of four main domains—the heptad repeat 1 (HR1) domain, heptad repeat 2 (HR2) domain, transmembrane domain (TM), and cytoplasm domain (CP) (Xia et al., 2020). During viral infection, a two-step sequential protease cleavage process triggers the activation of S proteins (Belouzard et al., 2009; Millet and Whittaker, 2014), which is modulated by host range and cell tropism. The first cleavage occurs between the S1 and S2 subunits, leading to the release of the S1 subunit and its transition to the post-fusion conformation (Su et al., 2016). Then, as the RBD of the S1 subunit binds to a host cell receptor [CoVs can recognize both angiotensin-converting enzyme 2 (ACE2) and dipeptidyl peptidase 4 (DPP4, also known as CD26)] (Kuhn et al., 2004; Raj et al., 2013), another cleavage site on S2 is exposed and cleaved by host proteases at the S2' site located upstream of the fusion peptide (FP). The HR1 and HR2 domains of S2 form a six-helix bundle (6-HB) fusion core in order to bring viral and cellular membranes into close proximity for subsequent fusion and infection (**Figure 3**) (Bosch et al., 2004). The host proteases such as furin (Millet and Whittaker, 2014), cathepsins (Bertram et al., 2013), human airway trypsin-like protease (Berman et al., 2000; Bertram et al., 2011), and transmembrane protease serine protease-2 (TMPRSS-2) (Gierer et al., 2013) are widely expressed in many important organs, which is a critical reason for the systematic infection, serious pathogenicity, and high mortality of the CoVs. Therefore, the RBD and 6-HB fusion core of CoVs and the proteases on infected cells have become potential targets for the development of virus attachment/fusion inhibitors, neutralizing antibodies, and vaccines.

Invasion of CoVs occurs in two steps, initial binding to the receptor on the cell surface and fusion of S protein with the host cell membrane to deliver their nucleocapsid to the



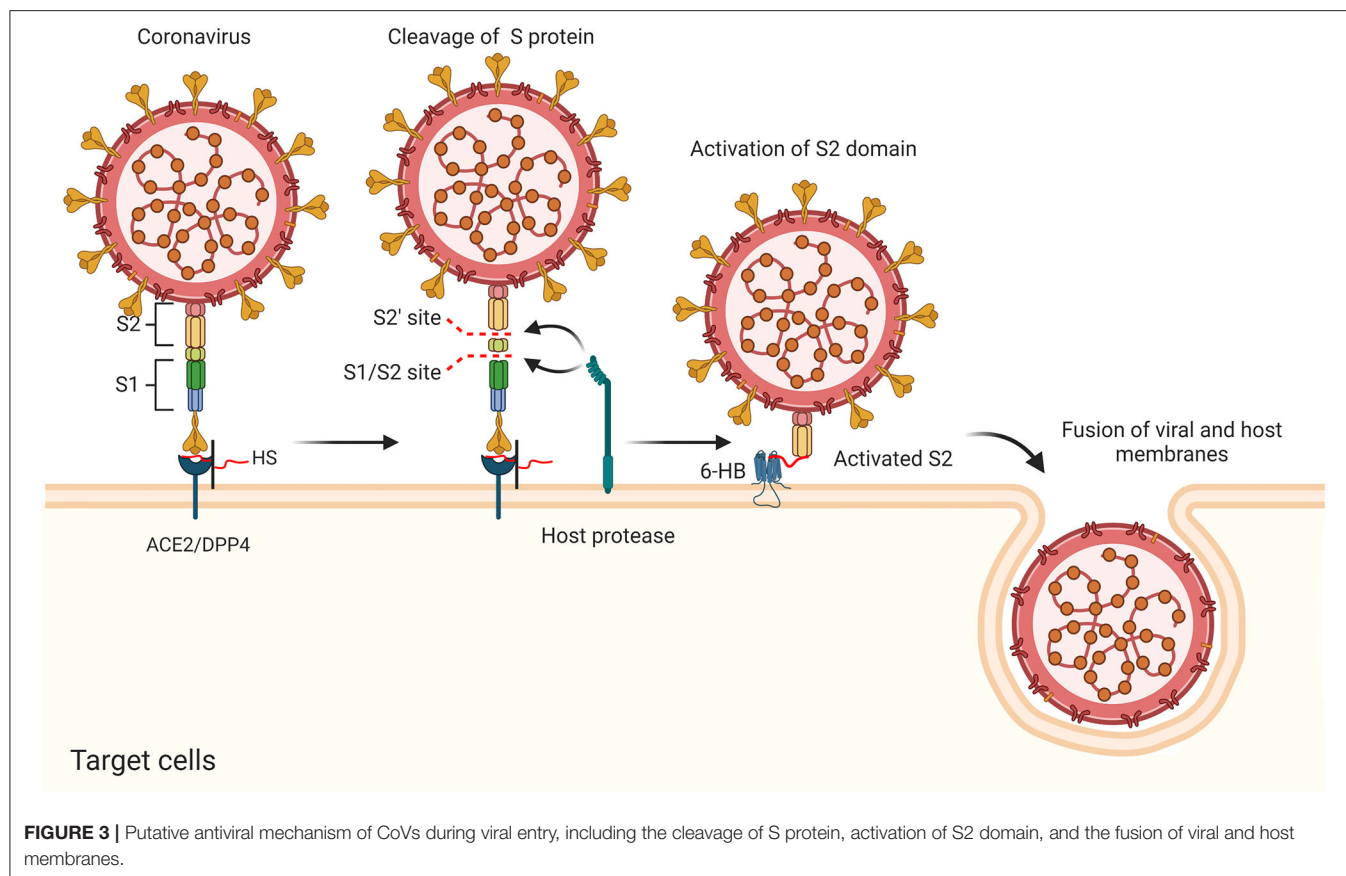
target cell. It has been known that CoVs often initially interact with cell surface molecules to promote their binding to specific receptors. HS proteoglycans (HSPGs) are abundantly present in almost all mammalian cells and serve as a co-receptor for a number of viruses (Gomes and Dietrich, 1982). HSPGs could initially bind to the surface proteins of CoVs, promote subsequent recognition of a secondary Receptor (ACE2/DPP4), and facilitate the attachment and entry of virus by increasing their local concentrations. Studies also suggested that different compositions in HS could impact the tropism of viruses (Wickramasinghe et al., 2011) and HS co-receptors on host cell surface leads to conformational changes of the CoVs' S proteins (Lang et al., 2011; Milewska et al., 2014; Mycroft-West et al., 2020a), possibly through the formation of a ternary complex (Clausen et al., 2020) (**Figure 3**). These findings suggested that the HS–S protein interactions might serve as a potential target to attenuate virus infection.

## HEPARIN/HS AND THEIR INTERACTIONS WITH PROTEIN

### Structures of Heparin/HS

Heparin is a significant anti-coagulant that has been used in clinic over decades. The heparin polysaccharide chains are linear and polyanionic, with repeating disaccharide units of  $\alpha$ -L-iduronic acid (IdoA) or  $\beta$ -D-glucuronic acid (GlcA) residue linked to glucosamine (GlcN) residue by a 1-4 glycosidic bond. The sugar units are sulfated at N-, 6- and 3-O on the GlcN residues as well as 2-O on the hexuronic acid by site-specific sulfotransferases. The 3-O-sulfation is rare but critical for heparin to form a specific pentasaccharide domain that specifically bind to anti-thrombin with high affinity, which is essential for its anti-coagulant activity (Lindahl et al., 1980). In addition to the anti-coagulant





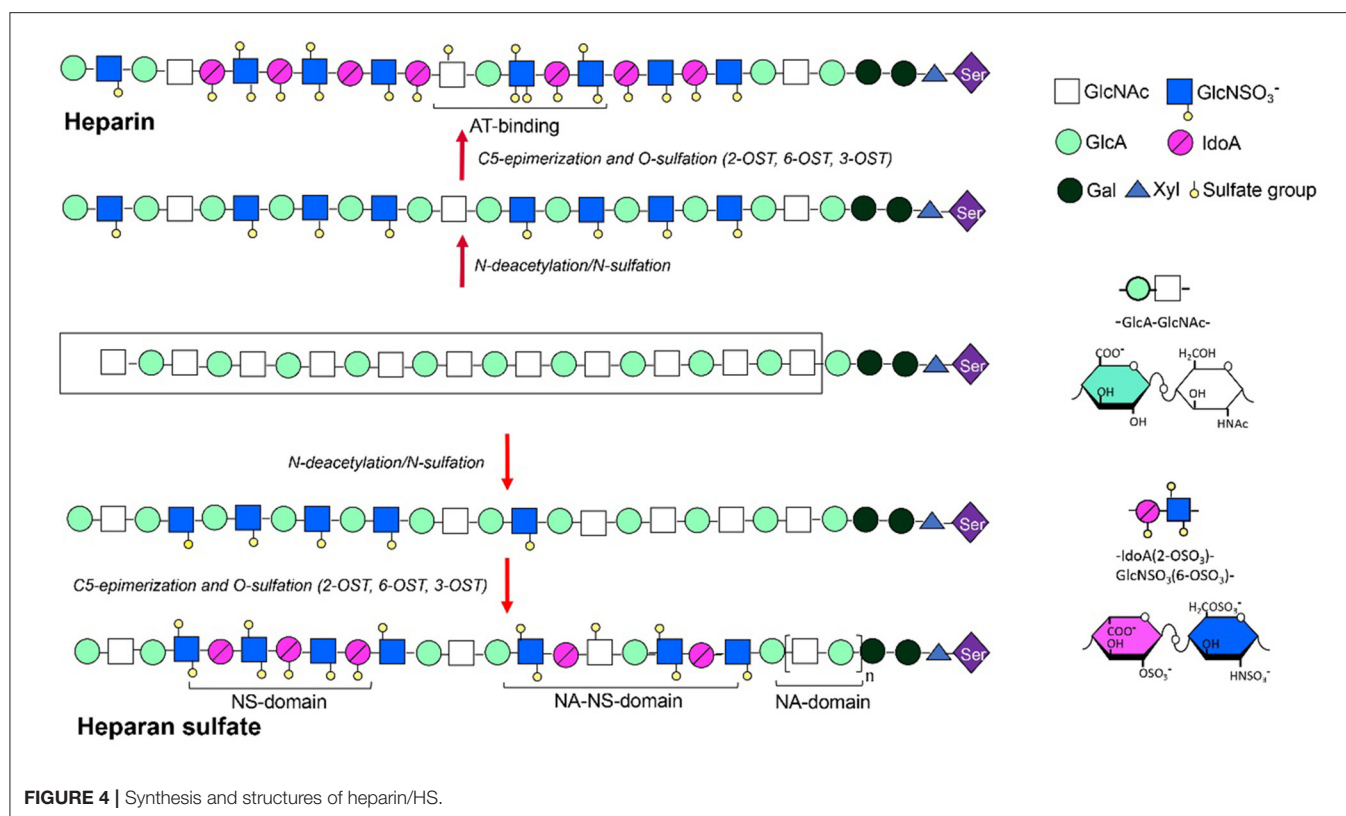
activity, heparin and its derivatives have also been studied for their anti-inflammatory, antiviral, anti-angiogenesis, anti-neoplastic, and anti-metastatic effects (Hao et al., 2019). HS shares high structural similarity with heparin (Linhardt and Toida, 2004), but generally with lower level of sulfation and epimerization, therefore displaying distinct domain structures (Figure 4). The functionalities of heparin and HS are mediated by their interaction with various proteins including proteases, protease inhibitors, chemokines, cytokines, growth factors, and their respective receptors (Xu and Esko, 2014; Seffer et al., 2020), with variable specificities.

## An Overview of Heparin/HS and Protein Interactions

Due to the highly anionic nature of heparin/HS, the interactions between heparin/HS and proteins are primarily through the interaction between negatively charged sulfate and carboxyl groups on heparin/HS and positively charged lysine and arginine residues on the proteins. The role of electrostatics in heparin/HS-protein interactions was elucidated in several studies (Olson et al., 1991; Thompson et al., 1994; Friedrich et al., 2001). Meanwhile, nonionic interactions such as hydrogen bonding and van der Waals packing also contribute to the free energy for the binding reactions (Thompson et al., 1994).

Some heparin/HS binding proteins can be identified by amino acid sequences known as Cardin-Weintraub motifs

corresponding to “XBBXB” and “XBBBXXB”, where X is a hydrophobic residue and B is a basic residue, such as arginine and lysine, responsible for interacting with the sulfate groups present on heparin/HS (Cardin and Weintraub, 1989; Hileman et al., 1998). On the other hand, well-characterized heparin/HS-protein interactions revealed specific requirement of the carbohydrate sequence. The most prominent example is the binding of antithrombin with the unique pentasaccharide sequence, -GlcNS/Ac6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S- in heparin, where the 3-O-sulfation is critical (Richard et al., 2009). Unlike the extremely rigorous sequence requirement as of antithrombin, or the purely non-specific interaction as in the case of heparin and protamine (Hubbard and Jennings, 1985), majority of the heparin/HS-protein interactions are selective. For instance, the interaction between HS and FGF2, a member of the fibroblast growth factor family, prefers the disaccharide unit of IdoA2S and GlcNS on heparin/HS (Turnbull et al., 1992; Jemth et al., 2002). More evidence is emerging, indicating that binding of HS and proteins is somewhat between purely specific and generally non-specific (Forsten-Williams et al., 2008; Nugent et al., 2013). Non-specific bindings solely depend on the high negative charge density of the carbohydrate chain and positively charged residues of the proteins, while modifications or domains on heparin/HS determine specificity levels of the interactions (Xu and Esko, 2014).



## Experimental Technologies Involved in Studying Heparin/HS-Protein Interactions

The major challenge for elucidating heparin/HS-protein interactions is to decipher the carbohydrate sequence that is commonly of high heterogeneity. Multidimensional technologies that facilitate understanding sequence specificities have been comprehensively summarized in a recent review (Yang and Chi, 2017), including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and mass spectrometry (MS).

Heparin/HS oligosaccharide microarrays are valuable tools that can be used to probe the interactions between structurally defined oligosaccharides and proteins with relatively small amounts of samples. The bottleneck of the microarray assay is to synthesize oligosaccharide libraries of intensive diversity. Zong et al. prepared a tetrasaccharide library consisting of 47 unique structures, which is one of the most comprehensive HS microarrays covering a large portion of possible structural variabilities (Zong et al., 2017). In a recent study, chemoenzymatic strategies have been successfully applied to construct microarrays composed of tetrasaccharide to 18-mer containing various N-, 6-O-, 2-O-, and 3-O-sulfation modifications (Horton et al., 2020). Cell-based microarrays have also been developed, aiming at demonstrating the functionality of specific heparin/HS saccharides in real cell signaling (Puvirajesinghe et al., 2012; Sterner et al., 2013).

The surface plasmon resonance (SPR) sensor (Thompson et al., 1994) is one of the most convenient tools for detecting heparin/HS-protein interactions through changes of the refractive index signals. One of the major advantages

of SPR is the capacity of probing biomolecular interactions at the thermodynamic level, offering real-time and label-free measurement of reaction rate constants ( $k_{on}$ ,  $k_{off}$ ) and resultant equilibrium constants ( $K_A$ ,  $K_D$ ) (Homola, 2008). In a recent study, interactions between heparin/HS and various cytokines were characterized by coupling surface plasmon resonance imaging for thermodynamic analysis method and Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF-MS) for structural determination (Przybylski et al., 2020). A self-assembled monolayer of short polyethylene oxide chains was used for grafting cytokines. Captured carbohydrates were carried out directly on the biochip surface using MALDI-TOF-MS, while MS identification was enhanced by on-chip digestion of the cytokine-bound GAGs by heparinase treatment.

## Computational Techniques

Despite the advances in the experimental techniques, there are limitations (e.g., failure in the acquisition of co-crystal structures) in obtaining the information regarding molecular interactions. Thus, computational techniques are indispensable tools for comprehensively understanding the heparin/HS-protein interactions. On the basis of theoretical models, the computational techniques are especially helpful in designing novel drug, performing wide scale analysis against large database (e.g., the PDB database), and for understanding the interaction dynamics. The interactions between protein and heparin/HS can be weak or strong, transient or stable,

non-permanent or permanent, which can be evaluated by basic parameters of binding affinities such as the equilibrium dissociation constant,  $K_D$  (Ma et al., 2018); Gibbs free energy of binding,  $\Delta G$  (Steinbrecher and Labahn, 2010); inhibition constant,  $K_i$  (Pekkarinen et al., 2007); half maximal inhibitory concentration,  $IC_{50}$  (Sebaugh, 2011); and electrostatic potential energy (Bitencourt-Ferreira et al., 2019). Several techniques are available for a wide application in assessment of protein-heparin/HS interaction.

### Homology Modeling

The homology modeling method uses 3D structures deposited in the PDB database to predict protein structures of sequential similarities. Homology modeling can give spatial structures with the highest accuracy (Werner et al., 2012) and thus has been widely applied for rational analysis of interactions between small organic molecule (ligand) and target protein during the docking and virtual screening for drug discovery (Cheng et al., 2012). Homology modeling can be built by four methods, including rigid body assembly [by tools like SWISS-MODEL (Arnold et al., 2006)], segment matching [by tools like SEGMOD/ENCAD (Levitt, 1992)], spatial restraint [by tools like MODELER (Sali and Blundell, 1993)], and artificial evolution [by tools like NEST (Petrey et al., 2003)].

### Molecular Docking

Molecular docking is a computational procedure extensively used in novel drug discovery, through which a small molecule (ligand) is docked into a macromolecule (target protein) at the binding sites to predict the binding conformations and affinity. The conformation of a ligand binding with the receptor depends on its state variables (including position:  $x$ -,  $y$ -, and  $z$ -translations; orientation: euler angles, axis angle, and quaternion; and conformation: torsion angles for each rotatable bond), which decides the extent of the multidimensional search space within the protein–ligand interaction. All docking methods require a scoring function to rank various candidate protein–ligand binding modes and a search method to explore their state variables. Scoring functions are computational approximations to predict protein–ligand binding affinity based on empiricism, force field, and knowledge, while search methods are classified into local and global ones by the extent of search space. Local search methods [such as solis and wets (Solis and Wets, 1981) and the pattern search (Lewis and Torczon, 2002)] tend to find the nearest or local minimum energy to the current conformation, whereas global methods [such as Monte Carlo (MC) simulated annealing (SA) (Kirkpatrick et al., 1983) and the genetic algorithm (GA) (Goldberg, 1988)] search for the best or global minimum energy within the defined search space. Hybrid global-local search methods have been shown to perform even better due to the higher efficiency in finding lower energies among the different candidate protein–ligand binding modes (Morris and Lim-Wilby, 1999). Molecular docking can be performed by various docking programs such as AutoDock (Goodsell et al., 1996), AutoDock Vina (Trott and Olson, 2010), FlexX (Rarey et al., 1996), GOLD (Jones et al., 1997), and Molegro Virtual Docker (MVD) (De Azevedo, 2010).

### Electrostatic Potential Energy

Since electrostatic interactions are part of scoring functions that widely influence the binding affinity of protein–ligand complexes, electrostatic potential energies are calculated in computational models to compare protein–ligand binding affinities. The electrostatic force is conservative as it only depends on the initial and final positions and most protein–ligand complexes show only partial charges. Partial Equalization of Orbital Electronegativity (PEOE) is the most widely used method (Gasteiger and Marsili, 1980) provided by the software AutoDockTools4 (Morris et al., 2009) that estimates partial charges of target protein and ligands in order to calculate electrostatic potential energy.

### Molecular Dynamics

Molecular dynamics (MD) is a computational simulation technique that can obtain not only multiple conformations of target proteins and ligands, but also a wealth of energetic status about the interactions in a time-dependent manner. MD simulations combined with binding free energy calculations can improve the accuracy of binding prediction and are thus suitable for studying the motions of target protein on ligand binding (Radkiewicz and Brooks, 2000; Salsbury et al., 2001). The Newtonian equation of motions is applied for each atom in the MD simulations for approximations (Schlick, 2010), which requires the information of initial coordinates (obtained from experimental structures, models, or combination of the two), potential (obtained from different force fields along with the coordinates) (MacKerell et al., 1998), and algorithms. Given the diverse complexity of the protein–ligand structures, different force field models [such as CHARMM (Miller et al., 2008), AMBER (Guvench and MacKerell, 2008), and GROMACS (Van Der Spoel et al., 2005)] are flexibly used during simulation, which are associated with modeling suites of CHARMM (Brooks et al., 2009), AMBER (Case et al., 2005), GROMACS (Hess et al., 2008), and NAMD (Phillips et al., 2005). Owing to the advances in computers and algorithms, the complexes of biomacromolecules can be simulated in nanoseconds with whole atoms, generating numerous conformations. The characterization of each conformation is accomplished by sophisticated methods that could be divided up to four types, including gross measures of protein and simulation stability, clustering analysis, quasi-harmonic and principal component analysis, and correlation function analysis. The gross measures of protein and simulation stability is the most widely used approach for checking simulation integrity and estimating equilibration timescale of the simulation. Parameters such as root-mean-square deviation (RMSD), structural clustering, free energy of binding and native contacts, and average temperature and pressure are generally calculated for obtaining their fluctuations.

Docking an HS fragment of proper size ( $\geq 4$  monosaccharide units) to a protein is challenging due to the flexibility brought by the glycosidic rings, linkages, and the high density of negative charges. Sapay et al. proposed a two-step method based on molecular docking and MD simulation to explore the binding modes of HS to cellular growth factors (FGF2 and CXCL12 $\alpha$ ) (Sapay et al., 2011). The method provided dynamical modeling



of the protein–ligand complex by building the docking models of HS fragment on protein surface and refining the contacts between HS fragment and the protein.

### Computational Study of SARS-CoV-2 Infection

The powerful tools of computational technology has made significant contribution to the studies on viruses including SARS-CoV-2. The method of homology modeling has made initially important contribution. Based on the rich genomic information and bioinformatics analysis of the proteins encoded by the novel coronavirus genes, Wu *et al.* built 19 structures of SARS-CoV-2 by homology modeling through the Fold and Function Assignment System server, including viral papain like protease (PLpro), main protease (3CLpro, also named 3-chymotrypsin-like protease), RNA-dependent RNA polymerase (RdRp), helicase, and S protein (Wu *et al.*, 2020). As the most efficient way to find anti-SARS-CoV-2 drugs is to screen those that are commonly used in clinic, small-molecule compounds from several resources including the U.S Food and Drug Administration (FDA)-approved drug database (ZINC drug database, ZDD), traditional Chinese medicine/natural products database, and the antiviral drugs database were docked into these computational models by ICM 3.7.3 modeling software (MolSoft LLC) to virtually screen potential druggable targets. Successfully predicted targets and potential drug compounds can be further tested *in vitro* and *in vivo* for treating SARS-CoV-2 infections.

The calculations of electrostatic potential energy were performed to estimate protein-heparin/HS binding affinities combined with the docking technique. Clausen *et al.* calculated the electrostatic potential map of both SARS-CoV-1 RBD (PDB ID: 3BGF) (Pak *et al.*, 2009) and SARS-CoV-2 RBD (PDB ID: 6M17) (Yan *et al.*, 2020) by the Molecular Operating Environment (MOE) software (Clausen *et al.*, 2020). Combined with docking studies of oligosaccharide fragments derived from heparin with RBD, it revealed an extended electropositive surface of RBD composed of positively charged residues including R346, R355, K444, R466, and possibly R509 that could coordinate the electronegative oligosaccharides through hydrogen bonds and hydrophobic interactions. This study demonstrated that the SARS-CoV-2 S protein may mediate an enhanced interaction with HS analogs and heparinoid derivatives compared to SARS-CoV-1 by the evolution of Lys444 and Glu354 (SARS-CoV-1) to Thr444 and Asn354 (SARS-CoV-2), respectively.

MD simulations have also been proven as a convenient method to describe the motions and binding affinities of ligand into target proteins. Han *et al.* designed and simulated several potential peptide inhibitors (including virus-binding domain  $\alpha$ -helices extracted from the protease domain of ACE2) against the SARS-CoV-2 coronavirus (Han and Kral, 2020). Classical MD simulations were performed by the modeling suites NAMD (Phillips *et al.*, 2005) and CHARMM36 protein force field (MacKerell *et al.*, 1998), which screened the most suitable peptide inhibitor with good binding affinity yet low RMSD for critical amino acids, indicative of relatively high binding energies. The novel designed peptide inhibitors have provided insights for researchers to develop therapeutic antiviral inhibitors by offering the  $\alpha$ 1 helix of ACE2 a sulfated ligand. Other molecules of similar

structures, the heparin/HS for instance, could also attach to positively charged residues at the bottom of the RBD.

The timescale of the MD simulations is also a determinant for the convergence of structural clustering, free energy of binding, and native contacts between the GAGs and target proteins. Bojarski *et al.* analyzed the structure of fibroblast growth factor 1 (FGF1) complexed with heparin [PDB ID: 2AXM (DiGabriele *et al.*, 1998)] through microsecond-scale simulations by the force field of AMBER16 (Bojarski *et al.*, 2019). The analysis revealed a conformational selection mechanism of GAGs binding and determined the structural specificity in the FGF1–heparin complex. Their findings could potentially contribute to the development of novel biomaterial resembling GAGs in the field of regenerative medicine.

## UTILIZING HEPARIN/HS–PROTEIN INTERACTIONS TO EXPLORE NOVEL STRATEGIES FOR TREATMENT OF SARS-COV-2 INFECTIONS

### Interactions Between Heparin/HS and the S Protein

Based on the knowledge of virus–heparin/HS interaction, it is assumed that exogenous added heparin/HS may interfere with viral infection. Several excellent studies have focused on the interactions between heparin/HS and the SARS-CoV-2 S protein, especially structure specificity of the carbohydrate chains. Nevertheless, it is worth noting that results from different research groups exhibited inconsistency to some extent, even if the similar analytical methods were conducted. This could be possibly attributed to experimental parameter setting, and the complexity and heterogeneity of the heparin/HS structures.

Using the SPR technique, Mycroft-West *et al.* first reported the SARS-CoV-2 S1 RBD binding to unfractionated heparin (Mycroft-West *et al.*, 2020a). Through circular dichroism (Martino *et al.*, 2020) spectroscopy, the authors further indicated that the RBD underwent conformational change in the presence of heparin, including helix and beta-sheet content alterations. The changes demonstrated that the RBD interacted with heparin in aqueous solution of physiological significance, whereby the major changes induced by heparin were those associated with antiparallel and helix content. In a subsequent study (Mycroft-West *et al.*, 2020b), the authors found that the addition of heparin to Vero cells between 6.25 and 200  $\mu\text{g ml}^{-1}$  inhibited invasion of SARS-CoV-2 by 44–80%. Additionally, SPR data revealed that 2-O, 6-O, and completely desulfated heparin had no inhibitory activity on heparin–RBD binding, proving the significance of 2-O and 6-O-sulfation on heparin/HS–spike interactions. On the other hand, persulfated, N-desulfated/N-re-acetylated, and strikingly 2-O/6-O doubly desulfated heparin possessed inhibitory activity. The authors attributed this phenomenon to a preference of RBD for a particular spatial arrangement of charged groups.

In another study, Kim *et al.* found that both monomeric and trimeric SARS-CoV-2 S proteins bound to immobilized heparin ( $K_D = 40$  and 73 pM, respectively) more tightly than the SARS-CoV and MERS-CoV S proteins (500 and 1 nM, respectively)

(Kim et al., 2020). Heparin-derived oligosaccharides (dp 4 to 18), N-desulfated, 2-O-desulfated, and 6-O-desulfated heparin failed to compete with immobilized heparin for binding to the S protein, suggesting that chain length and all the sulfate groups within heparin were critical in the interaction. On the other hand, in responses to heparin, tri-sulfated non-anti-coagulant HS, and non-anti-coagulant low-molecular-weight heparin (LMWH), the binding of SARS-CoV-2 S protein to the surface-immobilized heparin decreased in a concentration-dependent fashion. The  $IC_{50}$  were determined to be 0.056, 0.12, and 26.4  $\mu$ M, respectively. Additionally, unbiased computational ligand docking predicted putative heparin/HS-binding motifs on the S protein: 453–459 (YRLFRKS), 681–686 (PRRARS), and 810–816 (SKPSKRS), among which the 681–686 (PRRARS) site between the S1 and S2 subunits was a novel insertion not present in the SARS and MERS S proteins.

Liu et al. performed microarray binding experiments using an extensive HS oligosaccharide library (Liu et al., 2020). Their data suggested that the SARS-CoV-2 S protein can bind HS in a length- and sequence-dependent manner, while hexa- and octasaccharides composed of IdoA2S-GlcNS6S repeating units were identified as optimal ligands. Notably, 3-O-sulfation on the GlcN residue was proven not essential for efficient binding. In support of the microarray data, SPR experiments showed that the SARS-CoV-2 S protein bound with higher affinity to heparin ( $K_D = 55$  nM) compared to the RBD ( $K_D = 1$   $\mu$ M) alone. The previously determined octasaccharide composed of IdoA2S-GlcNS6S repeating subunits could inhibit heparin–S protein interaction with an  $IC_{50}$  of 38 nM. Their data supported a model in which the RBD of the SARS-CoV-2 S protein conferred sequence specificity and agreed with Kim et al. (2020) where an additional HS binding site in the S1/S2 proteolytic cleavage site could enhance the avidity of binding.

To obtain insights into heparin/HS–S protein binding and virus infection in a safer circumstance, Tandon et al. pseudotyped SARS-CoV-2 S protein on a third-generation lentiviral (pLV) vector for testing the impact of various sulfated polysaccharides on transduction efficiency in mammalian cells (Tandon et al., 2020). The pLV vector pseudotyped the S protein efficiently and produced high titers on HEK293T cells. Both unfractionated heparin (UFH) and enoxaparin (a low-molecular-weight heparin drug) exhibited high apparent inhibitory activity. However, in contrast with previous SPR results, the authors found that selective desulfation at the 6-O-position of the GlcN residue did not significantly reduce inhibitory activity of either UFH or enoxaparin. Concentration–response curves showed that pLV-S particles were efficiently neutralized by a range of concentrations of UFH, enoxaparin, 6-O-desulfated UFH, and 6-O-desulfated enoxaparin with an  $IC_{50}$  of 5.99  $\mu$ g/L, 1.08 mg/L, 1.77  $\mu$ g/L, and 5.86 mg/L, respectively. This study also enabled SPR analysis using pseudotyped lentiviral virions instead of isolated S protein, which was of more biological relevance. In the binding competition experiments, soluble heparin, non-anti-coagulant heparin, and a non-anti-coagulant low-molecular-weight heparin (NACH) showed  $IC_{50}$  values of 125 nM, 500 nM, and 25  $\mu$ M, respectively.

Tiwari et al. used a model of cellular cell-to-cell fusion assay to show that the SARS-CoV-2 S protein-mediated cell-to-cell

fusion could arise even in the absence of ACE2 (Tiwari et al., 2020). Further, they demonstrated that the S protein differentially recognized the 3-O-sulfated HS structures generated by the two different isoforms, 3OST-3B and 3OST-5, and that the S2 subunit was critical for cell-to-cell fusion mediated by the S protein–3-O-sulfated HSPG pathway. SPGG, a synthetic, small, and highly sulfated non-sugar compound, was capable of serving as an effective inhibitor of cell-to-cell fusion.

In an elegant study, Clausen et al. demonstrated the dependence of HS on SARS-CoV-2 infection (Clausen et al., 2020). Molecular modeling identified the putative binding surface for oligosaccharides that resided in the RBD of the S protein and were adjacent to, but separate from the ACE2 binding site. Interactions between RBD/ectodomain and HS were proved by affinity-based approaches. A ternary complex of heparin, ACE2, and the S protein was demonstrated by binding of S protein to immobilized heparin-BSA and titrating with biotinylated ACE2, in which case the binding of ACE2 increased in proportion to the amount of S protein bound to the heparin-BSA. Through flow cytometry, the authors proved that HS was essential for the spike ectodomains binding to ACE2 and several human cell lines, while heparin lyases treatment dramatically reduced binding. Similarly, targeting heparin/HS synthesis enzymes including NDST1, HS6ST1, HS6ST2, and B4GALT7 (required for GAG assembly) significantly reduced binding. Consistent with Liu et al. (2020), the authors claimed that the interaction between heparin and the S protein was independent of the anti-coagulant activity. Furthermore, infection of pseudotyped vesicular stomatitis virus (VSV) expressing the full-length S protein and SARS-CoV-2 virus was proven to be dependent on cellular HS. Interestingly, Hep3B cells with inactivated HS6ST1/2 responded differently to VSV and SARS-CoV-2.

Relevant to the previously established dependency of HS on SARS-CoV-2 infection, Martino et al. showed that commensal host bacterial communities capable of modifying HS changed with host age and sex in adult COVID-19 patients. The prevalence of those bacteria and the expression of key microbial glycosidases, which were capable of blocking SARS-CoV-2 S protein binding to human lung adenocarcinoma cells *in vitro*, was lower in bronchoalveolar lavage fluid (BALF) compared to healthy controls (Martino et al., 2020). Zhang et al. performed a drug repurposing screen and identified Mitoxantrone, an FDA-approved cancer treatment drug that also directly targets HSPGs and inhibits pseudo-coronavirus infection (Zhang et al., 2020a). Several other drugs, Sunitinib BNTX and Latrunculin, which disrupt actin dynamics on the cell surface, were also proven to inhibit SARS-CoV-2 cell entry. The fact that structurally unrelated actin inhibitors all blocked coronavirus entry strongly suggested that the endocytosis of coronavirus required the actin cytoskeleton in addition to ACE2 and HS.

## Interactions Between Heparin/HS and Cytokines

Severity of SARS, MERS, and COVID-2019 are associated with the presence of lymphopenia and inflammatory cytokine storm (de Wit et al., 2016; Tan et al., 2020; Zhou et al., 2020). The process of inflammation storm is divided into three steps accompanied



by a series of inflammatory responses and recruitment of leukocytes on the infected areas. (1) After initial invasion of virus, macrophages and mast cells immediately release macrophage inflammatory protein 1- $\alpha$  (TNF- $\alpha$ ) and interleukin-1 (IL-1) at the site of pathogen adhesiveness in order to activate leukocyte extravasation. (2) Selectins (E-, L-, and P-selectins) on leukocytes interact with endothelial surface-associated HS, which allows leukocyte tethering and rolling along vessel wall (Wang et al., 2002). (3) An array of inflammatory chemokines and cytokines are activated by HS presented on the endothelial surface, which triggers integrin adhesion molecules binding onto leukocytes and subsequent leukocyte extravasation out of the blood vessel (Butcher, 1991; Norgard-Sumnicht and Varki, 1995; Tanaka et al., 1996; Luo et al., 2001). Rich evidence has shown that various inflammatory chemokines and cytokines including macrophage inflammatory protein (MIP-1 $\alpha$ ), RANTES, granulocyte colony-stimulating factor, interferon- $\gamma$ -inducible protein 10, monocyte chemoattractant protein 1, tumor necrosis factor- $\alpha$  (Kuschert et al., 1999; Huang et al., 2020), as well as IL-2 (Najjam et al., 1998), IL-7 (Borghesi et al., 1999), IL-8 (Spillmann et al., 1998), and IL-10 (Salek-Ardakani et al., 2000) selectively bind to distinct domains of the heterogeneous HS with various affinities and sequence specificities. Gao et al. reported that periodate-oxidized, borohydride-reduced heparin (RO-heparin) could inhibit thioglycollate-induced peritoneal inflammation by preventing neutrophil recruitment dependent on the release of L- and P-selectin (Gao et al., 2005). This is an indication that RO-heparin could attenuate L- and P-selectin-mediated acute inflammation.

### Current/Potential Clinical Applications of Heparin/HS in COVID-19

With the evidence described above, it is most likely that disrupting HS-protein interactions by exogenous and competitive heparin/HS mimetics could interfere with virus infection and/or suppress the inflammatory responses. In fact, COVID-19 patients commonly suffer from hyper-coagulopathy and are routinely treated with heparin/LMWH. Significant differences in 28-days mortality were observed in the subgroup of patients with a concentration of D-dimer ( $>3 \mu\text{g/ml}$ ) higher than sixfold of the normal upper limit, or who had a sepsis-induced coagulopathy (SIC) score  $\geq 4$  (40.0 vs 64.2%,  $P = 0.029$ ) (Shi et al., 2020a). Recent studies have shown that hospitalized patients with severe COVID-19 treated with LMWH or fondaparinux (an ultra-low-molecular-weight heparin) had better prognosis in relation to mortality (Lin et al., 2020; Russo et al., 2020; Tang et al., 2020). It needs to be noted that side effects like heparin-induced thrombocytopenia (HIT) have also been reported in heparin-treated COVID-19 patients (Daviet et al., 2020; Lozano and Franco, 2020). Correct dosage and real-time monitoring of the anti-Xa activity are crucial in heparin/LMWH treatment of COVID-2019 (Duranteau et al., 2018).

Apart from its anti-coagulant effects, a retrospective cohort study found that IL-6 levels were significantly reduced while the percentage of lymphocytes was remarkably increased in the

hospitalized COVID-19 treated with LMWH in comparison to the non-LMWH-treated group (Shi et al., 2020b), demonstrating the anti-inflammatory activity of the drug. In addition to the systemic administration, local application of heparin/LMWH through intranasal or inhalation route have also been reported for the treatment of lung diseases and inhalation injury (Yildiz-Pekoz and Ozsoy, 2017; Zielinski et al., 2019). Considering the antiviral activities of heparin/LMWH, along with data suggesting that the nasal epithelium is a portal for initial infection and transmission, Tandon et al. suggested that intranasal administration of UFH may be an effective and safe prophylactic treatment SARS-CoV-2 transmission. Due to the low bioavailability of intranasally administered heparin (Bendstrup et al., 2002), this approach might avoid dangerous side effects or complications with anti-coagulation treatments while potentially still providing a prophylactic or therapeutic benefit (Tandon et al., 2020).

### CONCLUSIVE REMARKS

Despite the well-established anti-coagulant activity and the observed anti-inflammatory effects, the potential anti-SARS-CoV-2 activity of heparin/HS was only recently proposed. It is still controversial regarding the structure specificities of the heparin/HS chains for its interaction with the S protein; however, *in vitro* experiments and some clinical data have provided promising evidence of heparin/HS (including their mimetics) and heparin/HS-interacting molecules as anti-SARS-CoV-2 drugs. Further elucidation of the heparin/HS-S protein interaction will facilitate the construction of structurally defined oligosaccharide sequences that can be prepared through several methods reported (Roy et al., 2014; Hansen et al., 2015; Baytas and Linhardt, 2020; Zhang et al., 2020b). Non-anti-coagulant heparin, the anti-coagulant activity of which is selectively eliminated, may also be an option to be explored for the treatment of COVID-19 patients without the risk of bleeding complications (Cassinelli et al., 2020; Lindahl and Li, 2020).

### AUTHOR CONTRIBUTIONS

MY and TZ contributed equally to this work by writing the first draft. J-pL contributed to editing the manuscript. All authors contributed to the article and approved the submitted version.

### FUNDING

This work was supported by grants from the National Natural Science Foundation of China (31961133004), the Swedish Research Council (2018-02503, 2018-06016), and Beijing Advanced Innovation Center for Soft Matter Science and Engineering.

### ACKNOWLEDGMENTS

Figures 2, 3 were created with tools obtained from BioRender.com.

## REFERENCES

- Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201. doi: 10.1093/bioinformatics/bti770
- Baytas, S. N., and Linhardt, R. J. (2020). Advances in the preparation and synthesis of heparin and related products. *Drug Discov Today* 25, 2095–2109. doi: 10.1016/j.drudis.2020.09.011
- Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5871–5876. doi: 10.1073/pnas.0809524106
- Bendstrup, K. E., Gram, J., and Jensen, J. I. (2002). Effect of inhaled heparin on lung function and coagulation in healthy volunteers. *Eur. Respir. J.* 19, 606–610. doi: 10.1183/09031936.02.00105202
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucl. Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Bertram, S., Dijkman, R., Habjan, M., Heurich, A., Gierer, S., Glowacka, I., et al. (2013). TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *J. Virol.* 87, 6150–6160. doi: 10.1128/JVI.03372-12
- Bertram, S., Glowacka, I., Muller, M. A., Lavender, H., Gnirss, K., Nehlmeier, I., et al. (2011). Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J. Virol.* 85, 13363–13372. doi: 10.1128/JVI.05300-11
- Bitencourt-Ferreira, G., Veit-Acosta, M., and de Azevedo, W. F. (2019). Electrostatic energy in protein-ligand complexes. *Methods Mol. Biol.* 2053, 67–77. doi: 10.1007/978-1-4939-9752-7\_5
- Bojarski, K. K., Sieradzan, A. K., and Samsonov, S. A. (2019). Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers* 110:e23252. doi: 10.1002/bip.23252
- Borghesi, L. A., Yamashita, Y., and Kincade, P. W. (1999). Heparan sulfate proteoglycans mediate interleukin-7-dependent B lymphopoiesis. *Blood* 93, 140–148. doi: 10.1182/blood.V93.1.140
- Bosch, B. J., Martina, B. E. E., Van Der Zee, R., Lepault, J., Haijema, B. J., Versluis, C., et al. (2004). Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8455–8460. doi: 10.1073/pnas.0400576101
- Brooks, B. R., Brooks, C. L., Mackerell, A. D. Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., et al. (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc.21287
- Butcher, E. C. (1991). Leukocyte-endothelial cell recognition: three (or more) steps to specificity and diversity. *Cell* 67, 1033–1036. doi: 10.1016/0092-8674(91)90279-8
- Cardin, A. D., and Weintraub, H. J. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* 9, 21–32. doi: 10.1161/01.ATV.9.1.21
- Case, D. A., Cheatham, T. E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M. Jr., Onufriev, A., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi: 10.1002/jcc.20290
- Cassinelli, G., Torri, G., and Naggi, A. (2020). Non-anticoagulant heparins as heparanase inhibitors. *Adv. Exp. Med. Biol.* 1221, 493–522. doi: 10.1007/978-3-030-34521-1\_20
- Cheng, T., Li, Q., Zhou, Z., Wang, Y., and Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.* 14, 133–141. doi: 10.1208/s12248-012-9322-0
- Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* 2020:201616. doi: 10.1101/2020.07.14.201616
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., and Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* 176:104742. doi: 10.1016/j.antiviral.2020.104742
- Daviet, F., Guervilly, C., Baldesi, O., Bernard-Guervilly, F., Pilarczyk, E., Genin, A., et al. (2020). Heparin induced thrombocytopenia in severe COVID-19 patients. *Circulation* 2020:49015. doi: 10.1161/CIRCULATIONAHA.120.049015
- De Azevedo, W. F. (2010). MolDock applied to structure-based virtual screening. *Curr. Drug Targets.* 11, 327–334. doi: 10.2174/138945010790711941
- de Wit, E., van Doremalen, N., Falzarano, D., and Munster, V. J. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* 14, 523–534. doi: 10.1038/nrmicro.2016.81
- DiGabelle, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J., et al. (1998). Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* 393, 812–817. doi: 10.1038/31741
- Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., and Jiang, S. (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* 7, 226–236. doi: 10.1038/nrmicro2090
- Du, L., Yang, Y., Zhou, Y., Lu, L., Li, F., and Jiang, S. (2017). MERS-CoV spike protein: a key target for antivirals. *Expert. Opin. Ther. Targets.* 21, 131–143. doi: 10.1080/14728222.2017.1271415
- Duranteau, J., Taccone, F. S., Verhamme, P., Ageno, W., and Force, E. V. G. T. (2018). European guidelines on perioperative venous thromboembolism prophylaxis: intensive care. *Eur. J. Anaesthesiol.* 35, 142–146. doi: 10.1097/EJA.0000000000000707
- Elfiky, A. A., Mahdy, S. M., and Elshemey, W. M. (2017). Quantitative structure-activity relationship and molecular docking revealed a potency of anti-hepatitis C virus drugs against human corona viruses. *J. Med. Virol.* 89, 1040–1047. doi: 10.1002/jmv.24736
- Fehr, A. R., and Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* 1282, 1–23. doi: 10.1007/978-1-4939-2438-7\_1
- Forsten-Williams, K., Chu, C. L., Fannon, M., Buczek-Thomas, J. A., and Nugent, M. A. (2008). Control of growth factor networks by heparan sulfate proteoglycans. *Ann. Biomed. Eng.* 36, 2134–2148. doi: 10.1007/s10439-008-9575-z
- Friedrich, U., Blom, A. M., Dahlback, B., and Villoutreix, B. O. (2001). Structural and energetic characteristics of the heparin-binding site in antithrombotic protein C. *J. Biol. Chem.* 276, 24122–24128. doi: 10.1074/jbc.M011567200
- Gallagher, T. M., and Buchmeier, M. J. (2001). Coronavirus spike proteins in viral entry and pathogenesis. *Virology* 279, 371–374. doi: 10.1006/viro.2000.0757
- Gao, Y., Li, N., Fei, R., Chen, Z., Zheng, S., and Zeng, X. (2005). P-Selectin-mediated acute inflammation can be blocked by chemically modified heparin, RO-heparin. *Mol. Cells* 19, 350–355. Available online at: <http://www.molcells.org/journal/view.html?year=2005&volume=19&number=3&page=350>
- Gasteiger, J., and Marsili, M. (1980). Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 36, 3219–3228. doi: 10.1016/0040-4020(80)80168-2
- Gierer, S., Bertram, S., Kaup, F., Wrensch, F., Heurich, A., Kramer-Kuhl, A., et al. (2013). The spike protein of the emerging betacoronavirus EMC uses a novel coronavirus receptor for entry, can be activated by TMPRSS2, and is targeted by neutralizing antibodies. *J. Virol.* 87, 5502–5511. doi: 10.1128/JVI.00128-13
- Goldberg, D. (1988). *Genetic Algorithms in Search Optimization and Machine Learning*. New York, NY: Addison-Wesley Publishing Company, Inc.
- Gomes, P. B., and Dietrich, C. P. (1982). Distribution of heparin and other sulfated glycosaminoglycans in vertebrates. *Comp. Biochem. Physiol. B.* 73, 857–863. doi: 10.1016/0305-0491(82)90329-7
- Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996). Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* 9, 1–5. doi: 10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6
- Guvench, O., and MacKerell, A. D. (2008). Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* 443, 63–88. doi: 10.1007/978-1-59745-177-2\_4
- Han, Y., and Kral, P. (2020). Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano.* 14, 5143–5147. doi: 10.1021/acsnano.0c02857
- Hansen, S. U., Miller, G. J., Cliff, M. J., Jayson, G. C., and Gardiner, J. M. (2015). Making the longest sugars: a chemical synthesis of heparin-related [4] n oligosaccharides from 16-mer to 40-mer. *Chem. Sci.* 6, 6158–6164. doi: 10.1039/C5SC02091C

- Hao, C., Xu, H., Yu, L., and Zhang, L. (2019). Heparin: an essential drug for modern medicine. *Prog. Mol. Biol. Transl. Sci.* 163, 1–19. doi: 10.1016/bs.pmbts.2019.02.002
- He, Y., Li, J., Heck, S., Lustigman, S., and Jiang, S. (2006). Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design. *J. Virol.* 80, 5757–5767. doi: 10.1128/JVI.00083-06
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi: 10.1021/ct700301q
- Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998). Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays* 20, 156–167. doi: 10.1002/(SICI)1521-1878(199802)20:2<156::AID-BIES8>3.0.CO;2-R
- Homola, J. (2008). Surface plasmon resonance sensors for detection of chemical and biological species. *Chem. Rev.* 108, 462–493. doi: 10.1021/cr068107d
- Horton, M., Su, G., Yi, L., Wang, Z., Xu, Y., Pagadala, V., et al. (2020). Construction of heparan sulfate microarray for investigating the binding of specific saccharide sequences to proteins. *Glycobiology* 2020:cwaa068. doi: 10.1093/glycob/cwaa068
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/S0140-6736(20)30183-5
- Hubbard, A. R., and Jennings, C. A. (1985). Neutralisation of heparan sulphate and low molecular weight heparin by protamine. *Thromb Haemost.* 53, 86–89. doi: 10.1055/s-0038-1661242
- Hui, D. S., E, I. A., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., et al. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* 91, 264–266. doi: 10.1016/j.ijid.2020.01.009
- Jemth, P., Kreuger, J., Kusche-Gullberg, M., Sturiale, L., Gimenez-Gallego, G., and Lindahl, U. (2002). Biosynthetic oligosaccharide libraries for identification of protein-binding heparan sulfate motifs. Exploring the structural diversity by screening for fibroblast growth factor (FGF)1 and FGF2 binding. *J. Biol. Chem.* 277, 30567–73. doi: 10.1074/jbc.M203404200
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748. doi: 10.1006/jmbi.1996.0897
- Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., et al. (2020). Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. *Antiviral. Res.* 181:104873. doi: 10.1016/j.antiviral.2020.104873
- Kirkpatrick, S., Gelatt, C. D. Jr., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* 220, 671–680. doi: 10.1126/science.220.4598.671
- Kubo, H., Yamada, Y. K., and Taguchi, F. (1994). Localization of neutralizing epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J. Virol.* 68, 5403–5410. doi: 10.1128/JVI.68.9.5403-5410.1994
- Kuhn, J. H., Li, W., Choe, H., and Farzan, M. (2004). Angiotensin-converting enzyme 2: a functional receptor for SARS coronavirus. *Cell Mol. Life Sci.* 61, 2738–2743. doi: 10.1007/s00018-004-4242-5
- Kuschert, G. S., Coulin, F., Power, C. A., Proudfoot, A. E., Hubbard, R. E., Hoogewerf, A. J., et al. (1999). Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses. *Biochemistry* 38, 12959–12968. doi: 10.1021/bi990711d
- Lang, J., Yang, N., Deng, J., Liu, K., Yang, P., Zhang, G., et al. (2011). Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. *PLoS ONE* 6:e23710. doi: 10.1371/journal.pone.0023710
- Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.* 226, 507–533. doi: 10.1016/0022-2836(92)90964-L
- Lewis, R. M., and Torczon, V. (2002). A globally convergent augmented lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM J. Optimizat.* 12, 1075–1089. doi: 10.1137/S1052623498339727
- Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3, 237–261. doi: 10.1146/annurev-virology-110615-042301
- Li, J. P., and Kusche-Gullberg, M. (2016). Heparan sulfate: biosynthesis, structure, and function. *Int. Rev. Cell Mol. Biol.* 325, 215–273. doi: 10.1016/bs.ircmb.2016.02.009
- Lin, L., Lu, L., Cao, W., and Li, T. (2020). Hypothesis for potential pathogenesis of SARS-CoV-2 infection—a review of immune changes in patients with viral pneumonia. *Emerg. Microbes Infect.* 9, 727–732. doi: 10.1080/22221751.2020.1746199
- Lindahl, U., Backstrom, G., Thunberg, L., and Leder, I. G. (1980). Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin. *Proc. Natl. Acad. Sci. U. S. A.* 77, 6551–6555. doi: 10.1073/pnas.77.11.6551
- Lindahl, U., and Li, J. P. (2020). Heparin - an old drug with multiple potential targets in Covid-19 therapy. *J. Thromb. Haemost.* 2020:14898. doi: 10.1111/jth.14898
- Linhardt, R. J., and Toida, T. (2004). Role of glycosaminoglycans in cellular communication. *Acc. Chem. Res.* 37, 431–438. doi: 10.1021/ar030138x
- Liu, L., Chopra, P., Li, X., Wolfert, M. A., Tompkins, S. M., and Boons, G. J. (2020). SARS-CoV-2 spike protein binds heparan sulfate in a length- and sequence-dependent manner. *bioRxiv*. doi: 10.1101/2020.05.10.087288
- Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C. R., et al. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* 363, 938–947. doi: 10.1016/S0140-6736(04)15788-7
- Lozano, R., and Franco, M. E. (2020). Incidence of heparin-induced thrombocytopenia in patients with 2019 coronavirus disease. *Med. Clin.* 2020:23. doi: 10.1016/j.medcle.2020.05.023
- Lu, L., Liu, Q., Zhu, Y., Chan, K. H., Qin, L., Li, Y., et al. (2014). Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat. Commun.* 5:3067. doi: 10.1038/ncomms4067
- Luo, J., Kato, M., Wang, H., Bernfield, M., and Bischoff, J. (2001). Heparan sulfate and chondroitin sulfate proteoglycans inhibit E-selectin binding to endothelial cells. *J. Cell Biochem.* 80, 522–531. doi: 10.1002/1097-4644(20010315)80:4<522::AID-JCB1006>3.0.CO;2-H
- Ma, W., Yang, L., and He, L. (2018). Overview of the detection methods for equilibrium dissociation constant KD of drug-receptor interaction. *J. Pharm. Anal.* 8, 147–152. doi: 10.1016/j.jpha.2018.05.001
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616. doi: 10.1021/jp973084f
- Martino, C., Kellman, B. P., Sandoval, D. R., Clausen, T. M., Marotz, C. A., Song, S. J., et al. (2020). Bacterial modification of the host glycosaminoglycan heparan sulfate modulates SARS-CoV-2 infectivity. *bioRxiv* 2020:238444. doi: 10.1101/2020.08.17.238444
- Milewska, A., Zarebski, M., Nowak, P., Stozek, K., Potempa, J., and Pyrc, K. (2014). Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. *J. Virol.* 88, 13221–13230. doi: 10.1128/JVI.02078-14
- Miller, B. T., Singh, R. P., Klauda, J. B., Hodoscek, M., Brooks, B. R., and Woodcock, H. L. (2008). CHARMMing: a new, flexible web portal for CHARMM. *J. Chem. Inf. Model.* 48, 1920–1929. doi: 10.1021/ci800133b
- Millet, J. K., and Whittaker, G. R. (2014). Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15214–15219. doi: 10.1073/pnas.1407087111
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256
- Morris, G. M., and Lim-Wilby, M. (1999). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 443, 365–382.
- Mycroft-West, C., Su, D., Elli, S., Li, Y., Guimond, S., Miller, G., et al. (2020a). The 2019 coronavirus (SARS-CoV-2) surface protein (Spike) S1 Receptor Binding Domain undergoes conformational change upon heparin binding. *bioRxiv* 2020:971093. doi: 10.1101/2020.02.29.971093
- Mycroft-West, C., Su, D., Pagani, I., Rudd, T., Elli, S., Guimond, S., et al. (2020b). Heparin inhibits cellular invasion by SARS-CoV-2: structural dependence of



- the interaction of the surface protein (spike) S1 receptor binding domain with heparin. *bioRxiv* 2020:66761. doi: 10.1101/2020.04.28.066761
- Najjam, S., Mulloy, B., Theze, J., Gordon, M., Gibbs, R., and Rider, C. C. (1998). Further characterization of the binding of human recombinant interleukin 2 to heparin and identification of putative binding sites. *Glycobiology* 8, 509–516. doi: 10.1093/glycob/8.5.509
- Norgard-Sumnicht, K., and Varki, A. (1995). Endothelial heparan sulfate proteoglycans that bind to L-selectin have glucosamine residues with unsubstituted amino groups. *J. Biol. Chem.* 270, 12012–12024. doi: 10.1074/jbc.270.20.12012
- Nugent, M. A., Zaia, J., and Spencer, J. L. (2013). Heparan sulfate-protein binding specificity. *Biochemistry* 78, 726–735. doi: 10.1134/S0006297913070055
- Olson, S. T., Halvorson, H. R., and Bjork, I. (1991). Quantitative characterization of the thrombin-heparin interaction. Discrimination between specific and nonspecific binding models. *J. Biol. Chem.* 266, 6342–6352.
- Ou, X., Guan, H., Qin, B., Mu, Z., Wojdyla, J. A., Wang, M., et al. (2017). Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat. Commun.* 8:15216. doi: 10.1038/ncomms15216
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 11:1620. doi: 10.1038/s41467-020-15562-9
- Pak, J. E., Sharon, C., Satkunarajah, M., Auperin, T. C., Cameron, C. M., Kelvin, D. J., et al. (2009). Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. *J. Mol. Biol.* 388, 815–823. doi: 10.1016/j.jmb.2009.03.042
- Pekkarinen, A. I., Longstaff, C., and Jones, B. L. (2007). Kinetics of the inhibition of fusarium serine proteinases by barley (*Hordeum vulgare* L.) inhibitors. *J. Agric. Food Chem.* 55, 2736–2742. doi: 10.1021/jf0631777
- Petrey, D., Xiang, Z., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T., et al. (2003). Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* 53(Suppl.6), 430–435. doi: 10.1002/prot.10550
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289
- Przybylski, C., Gonnet, F., Saesen, E., Lortat-Jacob, H., and Daniel, R. (2020). Surface plasmon resonance imaging coupled to on-chip mass spectrometry: a new tool to probe protein-GAG interactions. *Anal. Bioanal. Chem.* 412, 507–519. doi: 10.1007/s00216-019-02267-2
- Puvirajesinghe, T. M., Ahmed, Y. A., Powell, A. K., Fernig, D. G., Guimond, S. E., and Turnbull, J. E. (2012). Array-based functional screening of heparin glycans. *Chem. Biol.* 19, 553–558. doi: 10.1016/j.chembiol.2012.03.011
- Radkiewicz, J. L., and Brooks, C. L. (2000). Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. *J. Am. Chem. Soc.* 122, 225–231. doi: 10.1021/ja9913838
- Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H., Muller, M. A., Dijkman, R., et al. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495, 251–254. doi: 10.1038/nature12005
- Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489. doi: 10.1006/jmbi.1996.0477
- Richard, B., Swanson, R., and Olson, S. T. (2009). The signature 3-O-sulfo group of the anticoagulant heparin sequence is critical for heparin binding to antithrombin but is not required for allosteric activation. *J. Biol. Chem.* 284, 27054–27064. doi: 10.1074/jbc.M109.029892
- Roy, S., El Hadri, A., Richard, S., Denis, F., Holte, K., Duffner, J., et al. (2014). Synthesis and biological evaluation of a unique heparin mimetic hexasaccharide for structure-activity relationship studies. *J. Med. Chem.* 57, 4511–4520. doi: 10.1021/jm4016069
- Russo, V., Cardillo, G., Viggiano, G. V., Mangiacapra, S., Cavalli, A., Fontanella, A., et al. (2020). Fondaparinux use in patients with COVID-19: a preliminary multicenter real-world experience. *J. Cardiovasc. Pharmacol.* 76, 369–371. doi: 10.1097/FJC.0000000000000893
- Salek-Ardakani, S., Arrand, J. R., Shaw, D., and Mackett, M. (2000). Heparin and heparan sulfate bind interleukin-18 and modulate its activity. *Blood* 96, 1879–1888. doi: 10.1182/blood.V96.5.1879
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626
- Salsbury, F. R. Jr., Crowley, M. F., and Brooks, C. L. (2001). Modeling of the metallo-beta-lactamase from *B. fragilis*: structural and dynamic effects of inhibitor binding. *Proteins* 44, 448–459. doi: 10.1002/prot.1110
- Sapay, N., Cabannes, E., Petitou, M., and Imbert, A. (2011). Molecular modeling of the interaction between heparan sulfate and cellular growth factors: bringing pieces together. *Glycobiology* 21, 1181–1193. doi: 10.1093/glycob/cwr052
- Schlick, T. (2010). “Molecular Modeling and Simulation: An Interdisciplinary Guide,” in *Molecular Dynamics: Basics* (New York, NY: Springer), 429–435.
- Sebaugh, J. L. (2011). Guidelines for accurate EC50/IC50 estimation. *Pharm. Stat.* 10, 128–134. doi: 10.1002/pst.426
- Seffer, M. T., Cottam, D., Forni, L. G., and Kielstein, J. T. (2020). Heparin 2.0: a new approach to the infection crisis. *Blood Purif.* 47, 1–7. doi: 10.1159/000508647
- Shi, C., Tingting, W., Li, J.-P., Sullivan, M. A., Wang, C., Wang, H., et al. (2020a). Comprehensive landscape of heparin therapy for COVID-19. *Carbohydrate Polym.* 254:117232. doi: 10.1016/j.carbpol.2020.117232
- Shi, C., Wang, C., Wang, H., Yang, C., Cai, F., Zeng, F., et al. (2020b). The potential of low molecular weight heparin to mitigate cytokine storm in severe COVID-19 patients: a retrospective cohort study. *Clin. Transl. Sci.* 2020:12880. doi: 10.1111/cts.12880
- Solis, F. J., and Wets, R. J. B. (1981). Minimization by random search techniques. *Math. Operat. Res.* 6, 19–30. doi: 10.1287/moor.6.1.19
- Spillmann, D., Witt, D., and Lindahl, U. (1998). Defining the interleukin-8-binding domain of heparan sulfate. *J. Biol. Chem.* 273, 15487–15493. doi: 10.1074/jbc.273.25.15487
- Steinbrecher, T., and Labahn, A. (2010). Towards accurate free energy calculations in ligand protein-binding studies. *Curr. Med. Chem.* 17, 767–785. doi: 10.2174/092986710790514453
- Sterner, E., Meli, L., Kwon, S. J., Dordick, J. S., and Linhardt, R. J. (2013). FGF-FGFR signaling mediated through glycosaminoglycans in microtiter plate and cell-based microarray platforms. *Biochemistry* 52, 9009–9019. doi: 10.1021/bi401284r
- Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., et al. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24, 490–502. doi: 10.1016/j.tim.2016.03.003
- Tan, L., Wang, Q., Zhang, D., Ding, J., Huang, Q., Tang, Y. Q., et al. (2020). Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal. Transduct. Target Ther.* 5:33. doi: 10.1038/s41392-020-0148-4
- Tanaka, Y., Kimata, K., Wake, A., Mine, S., Morimoto, I., Yamakawa, N., et al. (1996). Heparan sulfate proteoglycan on leukemic cells is primarily involved in integrin triggering and its mediated adhesion to endothelial cells. *J. Exp. Med.* 184, 1987–1997. doi: 10.1084/jem.184.5.1987
- Tandon, R., Sharp, J. S., Zhang, F., Pomim, V. H., Ashpole, N. M., Mitra, D., et al. (2020). Effective inhibition of SARS-CoV-2 entry by heparin and enoxaparin derivatives. *bioRxiv* 2020:140236. doi: 10.1101/2020.06.08.140236
- Tang, N., Bai, H., Chen, X., Gong, J., Li, D., and Sun, Z. (2020). Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J. Thromb. Haemost.* 18, 1094–1099. doi: 10.1111/jth.14817
- Thompson, L. D., Pantoliano, M. W., and Springer, B. A. (1994). Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain. *Biochemistry* 33, 3831–3840. doi: 10.1021/bi00179a006
- Tianji Zhang, R. Z., Yongqin, L.v., Meng, W., Hongmei, L., Tianwei, T., and Jin-Ping, L. (2019). Glycosaminoglycans in biological samples – towards identification of novel biomarkers. *TrAC* 2019:115732. doi: 10.1016/j.trac.2019.115732
- Tiwari, V., Tandon, R., Sankaranarayanan, N. V., Beer, J. C., Kohlmeier, E. K., Swanson-Mungerson, M., et al. (2020). Preferential recognition and antagonism of SARS-CoV-2 spike glycoprotein binding to 3- O - sulfated heparan sulfate. *bioRxiv* 2020:331751. doi: 10.1101/2020.10.08.331751
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334
- Turnbull, J. E., Fernig, D. G., Ke, Y., Wilkinson, M. C., and Gallagher, J. T. (1992). Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate. *J. Biol. Chem.* 267, 10337–10341.

- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291
- Wang, L., Brown, J. R., Varki, A., and Esko, J. D. (2002). Heparin's anti-inflammatory effects require glucosamine 6-O-sulfation and are mediated by blockade of L- and P-selectins. *J. Clin. Invest.* 110, 127–136. doi: 10.1172/JCI0214996
- Wang, N., Shang, J., Jiang, S., and Du, L. (2020). Subunit vaccines against emerging pathogenic human coronaviruses. *Front. Microbiol.* 11:298. doi: 10.3389/fmicb.2020.00298
- Werner, T., Morris, M. B., Dastmalchi, S., and Church, W. B. (2012). Structural modelling and dynamics of proteins for insights into drug interactions. *Adv. Drug Deliv. Rev.* 64, 323–343. doi: 10.1016/j.addr.2011.11.011
- Wickramasinghe, I. N., de Vries, R. P., Grone, A., de Haan, C. A., and Verheije, M. H. (2011). Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J. Virol.* 85, 8903–8912. doi: 10.1128/JVI.05112-11
- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., et al. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* 10, 766–788. doi: 10.1016/j.apsb.2020.02.008
- Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., et al. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol. Immunol.* 17, 765–767. doi: 10.1038/s41423-020-0374-2
- Xu, D., and Esko, J. D. (2014). Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* 83, 129–157. doi: 10.1146/annurev-biochem-060713-035314
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444–1448. doi: 10.1126/science.abb2762
- Yang, J., and Chi, L. (2017). Characterization of structural motifs for interactions between glycosaminoglycans and proteins. *Carbohydr. Res.* 452, 54–63. doi: 10.1016/j.carres.2017.10.008
- Yildiz-Pekoz, A., and Ozsoy, Y. (2017). Inhaled heparin: therapeutic efficacy and recent formulations. *J. Aerosol. Med. Pulm. Drug Deliv.* 30, 143–156. doi: 10.1089/jamp.2015.1273
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721
- Zhang, Q., Chen, C. Z., Swaroop, M., Xu, M., Wang, L., Lee, J., et al. (2020a). Targeting heparan sulfate proteoglycan-assisted endocytosis as a COVID-19 therapeutic option. *bioRxiv* 2020:2202549. doi: 10.1101/2020.07.14.202549
- Zhang, X., Lin, L., Huang, H., and Linhardt, R. J. (2020b). Chemoenzymatic synthesis of glycosaminoglycans. *Acc. Chem. Res.* 53, 335–346. doi: 10.1021/acs.accounts.9b00420
- Zhou, W., Liu, Y., Tian, D., Wang, C., Wang, S., Cheng, J., et al. (2020). Potential benefits of precise corticosteroids therapy for severe 2019-nCoV pneumonia. *Signal Transduct. Target Ther.* 5:18. doi: 10.1038/s41392-020-0127-9
- Zielinski, M., Wroblewski, P., and Koziński, J. (2019). Is inhaled heparin a viable therapeutic option in inhalation injury? *Adv. Respir. Med.* 87, 184–188. doi: 10.5603/ARM.2019.0029
- Zong, C., Venot, A., Li, X., Lu, W., Xiao, W., Wilkes, J. L., et al. (2017). Heparan sulfate microarray reveals that heparan sulfate-protein binding exhibits different ligand requirements. *J. Am. Chem. Soc.* 139, 9534–9543. doi: 10.1021/jacs.7b01399

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past collaboration with several of the authors TZ, HL, and J-pL.

Copyright © 2021 Yu, Zhang, Zhang, Sun, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease

Deling Shi, Anran Sheng and Lianli Chi\*

National Glycoengineering Research Center, Shandong University, Qingdao, China

## OPEN ACCESS

### Edited by:

Emil Alexov,  
Clemson University, Clemson, SC,  
United States

### Reviewed by:

Umesh R Desai,  
Virginia Commonwealth University,  
Richmond, VA, United States  
Grzegorz Wegrzyn,  
University of Gdansk, Poland

### \*Correspondence:

Lianli Chi  
lianlich@sdu.edu.cn

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 09 December 2020

**Accepted:** 27 January 2021

**Published:** 09 March 2021

### Citation:

Shi D, Sheng A and Chi L (2021)  
Glycosaminoglycan-Protein  
Interactions and Their Roles in  
Human Disease.  
Front. Mol. Biosci. 8:639666.  
doi: 10.3389/fmolb.2021.639666

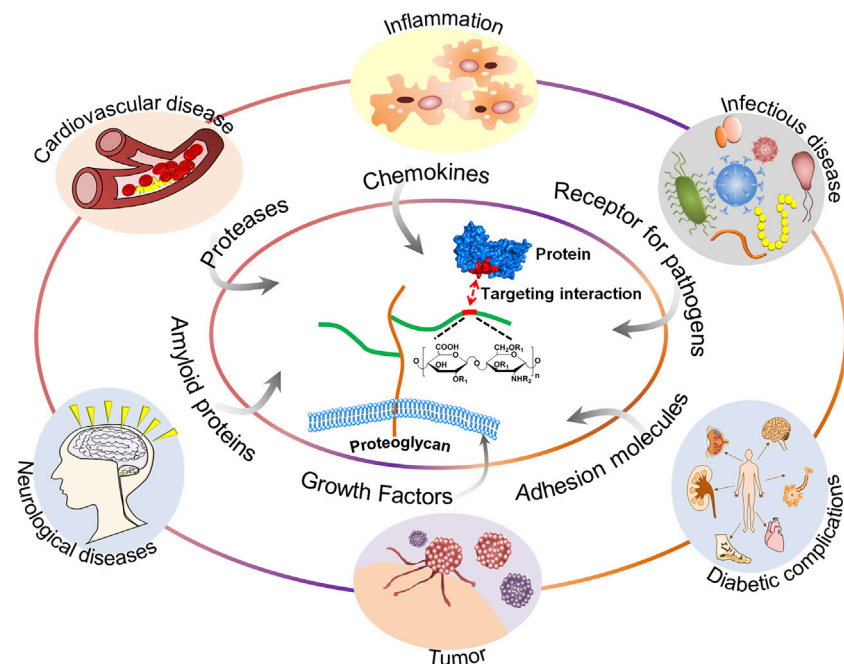
Glycosaminoglycans (GAGs) are a family of linear and negatively charged polysaccharides that exist ubiquitously on the human cell surface as well as in the extracellular matrix. GAGs interact with a wide range of proteins, including proteases, growth factors, cytokines, chemokines and adhesion molecules, enabling them to mediate many physiological processes, such as protein function, cellular adhesion and signaling. GAG-protein interactions participate in and intervene in a variety of human diseases, including cardiovascular disease, infectious disease, neurodegenerative diseases and tumors. The breakthrough in analytical tools and approaches during the last two decades has facilitated a greater understanding of the importance of GAG-protein interactions and their roles in human diseases. This review focuses on aspects of the molecular basis and mechanisms of GAG-protein interactions involved in human disease. The most recent advances in analytical tools, especially mass spectrometry-based GAG sequencing and binding motif characterization methods, are introduced. An update of selected families of GAG binding proteins is presented. Perspectives on development of novel therapeutics targeting specific GAG-protein interactions are also covered in this review.

**Keywords:** glycosaminoglycan, protein, human disease, interaction, molecular recognition

## INTRODUCTION

Recently, COVID-19 disease, caused by severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2), has led to medical and economic disruptions worldwide. Reports have shown that heparan sulfate (HS) is an indispensable cofactor for SARS-CoV-2 infection by interacting with both SARS-CoV-2 spike glycoprotein and angiotensin-converting enzyme 2 (ACE2) in the receptor-binding domain (RBD) (Clausen et al., 2020; Kim et al., 2020). Evidence has shown that heparin and its derivatives may contribute to the fight against SARS-CoV-2 infection and side effects (Liu et al., 2020; Tandon et al., 2020) by targeting the interaction between HS and related proteins. These studies have emphasized the importance of the interactions between glycosaminoglycans (GAGs) and proteins in disease and their roles as novel therapeutic targets, these interactions have been studied for decades but still lag behind the study of protein-protein and protein-nucleic acid interactions due to the structural complexity of GAGs and limitations of analytical tools.

GAGs are a family of linear and negatively charged polysaccharides that are commonly expressed in the interior and surrounding environment of most cell types, with a molecular weight of approximately 10–100 kDa (Kowitsch et al., 2018). Among the naturally occurring polysaccharides, the structure of GAGs is extremely complex due to alterations in residue types, glycosidic bond types, sulfation levels, sulfation positions and chain lengths. According to the type of hexosamine, hexose or hexuronic acid in the disaccharide repeating units and the glycosidic linkage



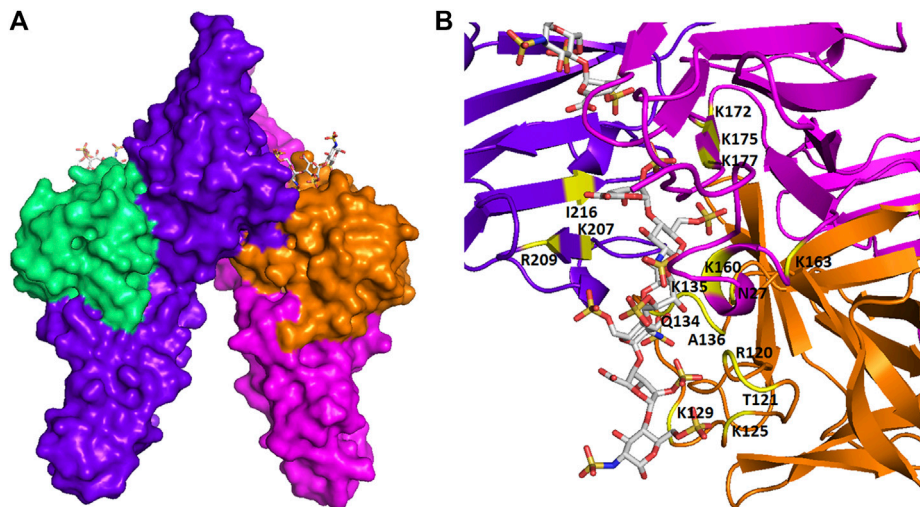
**FIGURE 1** | A schematic representation of the structure of GAGs and their interactions with proteins and functions relevant to specific diseases.

between these units, GAGs are divided into five main types: nonsulfated GAGs, such as hyaluronic acid (HA) (Dymarska et al., 2016), and sulfated GAGs, including heparin and HS (Shriver et al., 2012), chondroitin sulfate (CS) (Purushothaman et al., 2012), dermatan sulfate (DS) (Yamada and Sugahara, 2008), and keratan sulfate (KS) (Pomin, 2015). Heparin (~2.3 sulfate groups per disaccharide) and HS (~0.8 sulfate groups per disaccharide) consist of basic disaccharide repeats (GlcA/IdoA $\beta$ 1-4GlcNAc $\alpha$ 1-4) $_n$ , while the 3- and 6-positions of the glucosamine residue or the carboxyl group of uronic acid may be substituted or not substituted with sulfate groups. Heparin and HS have received the most attention and have been studied extensively due to their high sulfation and diverse biological activities, which are also our first concerns herein. Except for HA, all mammalian GAGs are linked to a core protein to form proteoglycans (PGs). The structure of the protein cores, the composition of the glycosaminoglycan chains, and the distribution of the proteoglycan all affect the biological activity of proteoglycans (Lindahl et al., 2015).

GAGs are of vital importance in the field of glycobiology, especially their multiple roles as signal molecules that regulate protein activity and act as structural components and effectors of cellular activity. GAGs have been demonstrated to modulate numerous biological processes, ranging from embryonic development, regulation of enzymatic activities, extracellular matrix assembly, and ligand binding to receptors to the regulation of cell signaling, through the regulation of distinct proteins, such as growth factors, chemokines, and adhesion molecules (Vallet et al., 2021). These processes are particularly important when related to diseases, including cardiovascular disease (Wight, 2018), cancer (Ma et al., 2020), infectious

diseases (Kamhi et al., 2013), neurodegenerative diseases (Huynh et al., 2019), inflammatory responses (Morla, 2019), and wound healing (Salbach et al., 2012). A schematic representation of the structure of GAGs and their interactions with proteins and functions relevant to specific diseases is shown in **Figure 1**.

The binding between GAGs and proteins are prominently ionic. Non-ionic forces, including hydrogen bonding and hydrophobic interaction, sometimes also play a major role in forming the GAG-protein complexes (Capila and Linhardt, 2002). It has been controversial that the binding between GAGs and proteins are non-specific until recently, as more and more studies have revealed the relatively high selectivity of GAG sequence to specifically bind to certain proteins. The binding posture and specificity were demonstrated in **Figure 2**, using a fibroblast growth factor (FGF)- FGF receptor (FGFR)-heparin complex as an example (Schlessinger et al., 2000). Furthermore, FGF1, and FGF2 signaling through FGFR 1c showed clearly different specificity when screening against a library of chemoenzymatically synthesized HS with defined structures (Schultz et al., 2017). Additional examples on specificity of GAG-protein interactions include a 2-O-sulfate-GlcA containing HS hexasaccharide selectively activating heparin cofactor II (Sankarayanarayanan et al., 2017), a 3-O-sulfated HS being preferentially recognized by SARS-CoV-2 spike glycoprotein (Tiwari et al., 2020), and a 3-O-sulfated HS octasaccharide specifically binding to herpes simplex virus type 1 glycoprotein D (Huang et al., 2017). Besides, high-throughput study using HS microarray revealed that HS-binding proteins, including FGF2 and several chemokines, require clearly different ligands on HS (Zong et al., 2017). A review focused on the topic of



**FIGURE 2 |** The spatial structure of an FGF-FGFR-heparin complex. **(A)** The surface view. **(B)** The view of ribbon structure. The heparin fragments ( $\Delta$ UA-GlcNS6S-IdoA2S-GlcNS6S-IdoA2S-GlcNS6S) that make contacts to two FGF2s (shown in green and orange) and two FGFR1s (shown in purple and red) are represented as balls and sticks. The amino acid residues that participate in the interaction are indicated. The figure was prepared by using PDB code 1FQ9, which was originally reported in the reference Schlessinger et al. (2000).

selectivity of GAG-protein interactions has been recently written by Kjellén and Lindahl (Kjellén and Lindahl, 2018). The selectivity of these interactions is fundamental for designing HS mimetics as promising therapeutics.

There is increasing interest in exploring the essentials of GAG-protein interactions and their roles in human diseases. In particular, novel therapeutics targeting specific GAG-protein interactions have important application value, such as the treatment of coronary pneumonia. As the interaction between GAGs and proteins involves a wide range of physiological processes, the influence of their interaction on specific diseases and their potential therapeutic effects have attracted much attention in an effort to find new methods for treatment or prevention of disease. To synthesize structural analogs, remove or modify structures, or block the interaction with reagents, it is necessary to obtain defined mechanisms and binding sequences. The recent breakthroughs in analytical tools and approaches, especially mass spectrometry (MS)-based GAG sequencing and binding motif characterization methods, have facilitated a greater understanding of the structural basis and mechanisms of GAG-protein interactions, creating an opportunity to utilize the structural diversity of GAGs to discover novel therapeutics. Further understanding of the interaction process and mechanism between GAGs and proteins will contribute to the proper understanding of the occurrence and development of a great number of diseases and the development of new therapeutic approaches.

This review focuses on the interaction between GAGs and proteins and their effect in human disease. In addition, the molecular basis and mechanisms of GAG-protein interactions are introduced. The latest progress in GAG-binding proteins and analytical tools is also discussed. Moreover, perspectives on

development of novel therapeutics targeting specific GAG-protein interactions are presented.

## HUMAN DISEASES RELATED TO GAG-PROTEIN INTERACTIONS

### Cardiovascular Disease

The first specific GAG-protein interaction described was heparin and antithrombin, which has important physiological significance and was used in the production of pharmaceutical heparin products as anticoagulants for treatment of thrombosis, embolism and thrombophlebitis. Heparin and low molecular weight heparin (LMWH) inhibit coagulation factors Xa and IIa by combining with antithrombin III to prevent thrombosis. Since then, the function of GAGs and their interaction with proteins in the vascular system have been studied. Although heparin is successfully used to prevent thrombosis in hospitalized patients, it was reported to present a risk of bleeding at prophylactic doses (Sunseri et al., 2018). This prothrombotic adverse reaction, named heparin-induced thrombocytopenia (HIT), is mediated by immunity and is also caused by an interaction between heparin and protein. Heparin products form multimolecular complexes with antigenic platelet factor 4 (PF4), resulting in the formation of IgG platelet-activating antibodies which are against the heparin/PF4 complex, which triggers an immune response and induces platelet activation and aggregation (Ho and Siordia, 2016). This process leads to platelet reduction and thrombin generation, ultimately resulting in thrombocytopenia. At the same time, the process may also be accompanied by the formation of venous or arterial thrombosis, which then

develops into deep venous thrombosis and pulmonary embolism (Warkentin, 2018).

Early studies have shown that GAGs accumulate in disease-prone areas of the vascular system, such as at branch points, and are often consistent with lipid deposition. Subsequent studies have shown that GAGs are covalently linked to specific core proteins and interact with different ligands within the interstitial space to help regulate vascular structure and function. PGs also interact with a variety of receptors on the surface of vascular cells, partially regulating the phenotype of vascular cells (Wight, 2018). For example, DSPG can promote the formation of atherosclerosis (Edwards et al., 2004), while CSPG may participate in the process of early atherosclerosis intimal thickening (Wight and Merrilees, 2004). HSPG is negatively regulated by atherogenic molecules; thus, the lipoprotein regulation of endothelin may play a key role in the formation of atherosclerosis (Pillarsetti, 2000). Recently, the relationship between cardiovascular disease and heparin-binding protein (HBP) was confirmed by using bioinformatics methods (Cai et al., 2020), which showed that HBPs may act as a novel biomarker linking cardiovascular diseases, such as atherosclerosis, myocarditis, myocardial ischemia, and myocardial infarction (MI). Specific HBPs or signaling pathways can be developed as new therapies for cardiovascular disease.

## Tumors

In the last few decades, PGs have been found to be involved in the functions and mechanisms of cancer cells and play a key role in cancer cell adhesion, migration, invasion, and metastasis. HS proteoglycans (HSPGs) are proteins that are covalently linked with HS. The main HSPGs can be classified into two main categories: cell surface HSPGs (syndecans and glypicans) and basement membrane HSPGs (perlecan, agrin and collagen type XVIII). HSPGs are downregulated or upregulated in different tumors (De Pasquale and Pavone, 2020). GPC1, a cell surface HSPG, was found to be overexpressed in breast cancer (Matsuda et al., 2001), glioma (Saito et al., 2017), and pancreatic cancer (Kleeff et al., 1998) but downregulated in colorectal cancer (Knelson et al., 2014). HS can bind growth factors to regulate angiogenesis, including fibroblast growth factors (FGFs), vascular endothelial growth factors (VEGFs) and platelet-derived growth factors (PDGFs). Perlecan on the tumor cell surface can interact with ligand and adaptor proteins to enhance FGF signaling and tumor angiogenesis (Whitelock and Iozzo, 2005). If the C-terminus of perlecan is lacking, VEGF synthesis would be reduced to suppress tumor angiogenesis (Sharma et al., 1998). The other GAGs also have important functions in tumors. CS-E is not expressed in normal ovaries or cystadenomas but is highly expressed in extracellular matrices (ECMs) of ovarian adenocarcinomas to mediate VEGF binding (Ten Dam et al., 2007). It has been reported that the tumor microenvironment can induce HA production (Tammi et al., 2011). HA is highly expressed in breast cancer (Auvinen et al., 2000), lung cancer (Pirinen et al., 2001) and ovarian cancer (Anttila et al., 2000), while HA expression is low in squamous cell carcinoma and melanoma (Karjalainen et al., 2000; Kosunen et al., 2004). From these studies, the abnormal PG expression levels or structural changes in PGs during tumorigenesis and progression indicate their importance as

potential biomarkers of cancer occurrence and progression and as therapeutic targets.

## Infectious Disease

Given their ubiquity and abundant biological functions, GAGs are the main target of pathogens in the infection process and play an important role in the initial attachment of pathogens to host cells. Studies have shown that GAGs interact with microbial pathogens on the cell surface and ECMs to modulate microbial pathogenesis and host defense. Many pathogenic microorganisms, such as viruses (e.g., human papilloma virus (HPV) (Kines et al., 2009), hepatitis C virus (HCV) (Barth et al., 2003), dengue virus (Dalrymple and Mackow, 2011), bacteria (e.g., *Listeria monocytogenes* (Banerjee et al., 2004) and protozoa (e.g., malaria sporozoites (Clausen et al., 2012) can express proteins that bind to HS, DS, and CS on cell surfaces, thereby facilitating the host cell infection process.

The latest evidence shows that HS, as a cofactor of SARS-CoV-2 infection, transforms the spinous process structure into an open conformation through interaction of the spike glycoprotein in the RBD of SARS-CoV-2 to promote the binding of adjacent ACE2 (Clausen et al., 2020). Previous experiments have shown that HSPGs are essential cell-surface molecules involved in SARS-CoV cell entry by providing binding sites for SARS-CoV invasion at the early stage (Lang et al., 2011). Coronavirus NL63 entry into host cells relies on HS interactions that increase virus density at the cell surface. The entry of coronavirus NL63 into host cells is achieved by using GAGs as adhesion molecules to increase the virus density on the cell surface, which is an example of pathogens using GAGs to survive (Milewska et al., 2014). Other microbial pathogens, such as Middle East respiratory syndrome coronavirus (MERS-CoV) and the Gram-negative bacterium *Pseudomonas aeruginosa* (Park et al., 2001), which can cause respiratory infections, have also been reported to interact with GAGs.

When the skin barrier is damaged, the GAGs at the wound site will change and can bind to pathogens, such as Merkel cell polyoma virus (MCV) (Schowalter et al., 2011), *S. aureus* (Liang et al., 1992), *Candida* (Green et al., 2013) and *Leishmania* (Fatoux-Ardore et al., 2014). Merkel cell polyoma virus (MCV) infection is an example. MCV is a circular double-stranded DNA virus and the causative agent of Merkel cell carcinoma, which is a rare but fatal skin cancer. When MCV first attaches to cells, it mainly binds to HS on the cell surface and, to a lesser extent, binds to CS. After treatment of cells with heparanase and chondroitinase sulfate, MCV infection is significantly affected. In addition, other diseases are related to the interaction of GAGs and pathogenic microorganisms, including enterocolitis (Boyd et al., 1998), diarrhea (Viboud and Bliska, 2005), keratitis (Hayashida et al., 2011), and AIDS (Hayashida et al., 2015).

## Diabetic Complications

Diabetes encompasses a group of lifelong metabolic diseases characterized by chronic hyperglycemia due to multiple causes. According to World Health Organization statistics, diabetes is the disease with the most complications, including diabetic



cardiopathy, diabetic ocular surface, and diabetic foot. One of the most important complications for diabetic patients is diabetic nephropathy. Diabetic nephropathy is a major microvascular complication in long-term diabetic patients. The prolonged hyperglycemia caused by diabetes can lead to glycosylation and non-enzymatic cross-linking between proteins and glucose or its derivatives (Qiu et al., 2020). A series of further complex molecular rearrangements produces irreversible advanced glycation end products (AGEs). AGEs initiate and accelerate the development of renal disease by activating the receptor for advanced glycation end products (RAGE). Through surface plasmon resonance (SPR) analysis, it was found that the affinity of RAGE for low molecular weight heparins (LMWHs) was approximately 6 times higher than that for AGEs. The antagonistic effect of LMWHs on RAGE helps to improve diabetic nephropathy (Myint et al., 2006). A permeability change in the capillary wall of the glomerulus is an early manifestation of diabetic nephropathy, which clinically manifests as abnormal proteinuria. The basement membrane of the glomerulus contains highly negatively charged GAGs represented by HS, which can prevent passage of charged macromolecules. Neutralization of anions in the capillary wall of the glomerulus is related to the loss of charge-dependent glomerular permeability selectivity. The decrease in HS is due to the increase in heparanase-1 gene expression in glomerular epithelial cells induced by glucose in patients with diabetic nephropathy. Heparin or LMWHs can be used as heparinase inhibitors to effectively reverse the abnormal permeation selectivity of the glomerulus and improve diabetic nephropathy (Lewis and Xu, 2008).

## Mucopolysaccharidoses

Mucopolysaccharidoses (MPS) are a group of diseases caused by abnormal accumulation of GAGs. The patients are of genetic defects and produce no or deficient lysosomal enzymes to degrade metabolic GAGs. Based on the deficient enzyme and symptom, MPS are divided into seven different types and more subtypes. Unfortunately, there is no medical treatment can cure these diseases. Most studies are focused on the early diagnostics of MPS. Currently, enzyme replacement therapy and hematopoietic stem cell transplantation are primarily used in clinic to control the progress of MPS and improve the conditions of patients (Zhou et al., 2020).

In MPS patients, GAGs are accumulated in cells, blood and tissues, which consequents to pathological symptoms over time. However, the exact mechanism of biological interactions with accumulated GAGs and proteins remains unclear. Most recent research in this field suggested that abnormally accumulated HS in MPS patients tightly bound to cathepsin V and inhibited its elastolytic activity. HS antagonist was able to restore the activity of cathepsin V (Chazeirat et al., 2021). The new findings encourage exploring novel approaches for treating MPS and associated disorders based on the molecular interaction between GAGs and proteins.

## Other Diseases

GAGs also play a crucial role in inflammation, neurological diseases (e.g., Parkinson's disease, Alzheimer's disease (AD),

and mad cow disease) and other diseases. The important role of GAGs in the inflammatory response has been reported in previous studies. As the structural heterogeneity of HS is usually concentrated in the high-sulfate region, it can participate in almost every stage of leukocyte passage through the vascular wall and can interact with a variety of proteins, such as L-selectin, CXC-chemokine ligand 8 (CXCL8), and histidine-rich glycoprotein (HRG) (Parish, 2006). The interaction of HA with CD44 and tumor necrosis factor-stimulated gene-6 (TSG-6) activates a variety of inflammatory cells (Baranova et al., 2011), and HA also interacts with Toll-like receptor four to promote the release of cytokines by dendritic cells (Taylor and Gallo, 2006). LMWHs can combine with tumor necrosis factor (TNF) and the nuclear transcription factor NF- $\kappa$ B to prevent leukocyte extravasation (Luan et al., 2014). Moreover, some studies have shown that GAGs may be used to treat AD and other age-related dementias. GAGs can interact with basic fibroblast growth factor (FGF-2), VEGF, brain-derived neurotrophic factor (BDNF) and tau growth factors (Huynh et al., 2019). Heparin can inhibit the activity of  $\beta$ -site APP cleaving enzyme (BACE1) to reduce  $\beta$ -amyloid protein content (Cui et al., 2011). Similarly, CS extract from *Sardina pilchardus* can also inhibit BACE1 (Mycroft-West et al., 2020). In addition, GAGs are of great value in the treatment of sinusitis, asthma, chronic obstructive pulmonary disease, cystic fibrosis, and primary ciliary dyskinesia. For example, TSG-6, CD44, and Toll-like receptor 4 (TLR4) can be activated by HA, leading to calcium channel activation and immune activation (Garantziotis et al., 2016). In addition, a reduction in contractile protein content in the diaphragm and some growth factors has been reported to lead to changes in glycosaminoglycan epitopes in patients with chronic obstructive pulmonary disease (Ottenheim et al., 2007).

As summarized herein, nearly all types of major human diseases are related to GAGs more or less. There are still great demands for therapeutics to treat these diseases. Understanding the role of GAGs in these diseases and knowing how to modulate these physiological or pathological processes using artificial GAGs might open an era of discovering new drugs based on GAGs or targeting GAGs.

## GLYCOSAMINOGLYCAN-BINDING PROTEINS

### Serpins

Serpin family protein proteinase inhibitors play a critical role in regulating proteinases in diverse physiologic processes by regulating the activity of serine and cysteine proteinases through a conformational trapping mechanism, providing a finely tuned time- and location-dependent regulation of proteinase activity (Huntington, 2006). In plasma, antithrombin III (AT III) and heparin cofactor II (HC II) are major heparin-dependent protease inhibitors that maintain blood fluidity by interacting with cell surface GAGs. Antithrombin, in cooperation with heparin and HS, causes anticoagulation by preventing activation of blood clotting proteinases at the site of vascular injury. Under normal conditions, antithrombin



inhibits blood clotting proteinases in a repressed reactivity state because the exposed reactive center loop (RCL) of serpin only provides the minimum specificity determinants to identify thrombin, factor Xa and factor IXa. In addition, unfavorable interactions diminish the favorable RCL and exosite interactions with proteinases. The combination of specific heparin or HS with antithrombin can induce allosteric activation, thus reducing adverse interactions and promoting template bridging of the serpin and proteinase (Olson et al., 2010). The defined protein-binding motif and molecular basis for the anticoagulant function of heparin have been reported to involve a specific pentasaccharide sequence that can bind to AT III. At least 16 saccharides of the heparin chain are required, although only the pentasaccharide is necessary (Guerrini et al., 2014). By interacting with AT III, heparin enhances AT III-mediated inhibition of thrombin and factor Xa. Inactivation of these proteases by AT III is greatly accelerated by the binding of heparin, increasing the bimolecular rate constant by a factor of 2000 (Rosenberg and Damus, 1973). Interestingly, heparin also binds to HC II but does not exhibit selectivity. Instead, the sequence of a unique DS hexasaccharide has been elucidated to interact with HC II of high affinity (Maimone and Tollefsen, 1990; Raghuraman et al., 2010). These again demonstrated the selectivity of binding between GAGs and proteins. However, Other serpins that rely on binding to GAGs to enhance their inhibition include heparin cofactor II, protein C inhibitor and protease nexin I (Munoz and Linhardt, 2004; Rein et al., 2011).

## Growth Factors

HSPGs interact with growth factors [e.g., FGFs (Huynh et al., 2019), VEGF (Gitay-Goren et al., 1992), transforming growth factor  $\beta$  (TGF- $\beta$ ) (Lee et al., 2015), and PDGF (Fager et al., 1992)] to promote their biological activities. The proteins in the FGF family may be the most extensively studied heparin-binding proteins and have a high affinity for cell surface HSPGs. FGFs participate in developmental and physiological processes through binding cell surface FGFRs as well as GAGs. These growth factors, such as acidic fibroblast growth factor (FGF-1) and FGF-2, must interact with and be activated by an active ternary complex comprising canonical receptors (FGFRs) and GAGs on endothelial surface PGs. Then, the three components FGF, FGFR, and HS interact simultaneously with signal transduction, thus triggering cell division and further processing (Fannon et al., 2000). In addition, the GAG interaction is necessary to stabilize the FGF-FGFR complex by balancing the surface charges. This interaction also limits the activity of growth factors to a certain extent. In fact, FGF binding is achieved through selected sequences (protein-binding motifs) within the HS backbone, although the minimal binding sequences are still controversial (Pomin, 2016). Heparin-binding epidermal growth factor-like growth factor (HB-EGF) is a member of the EGF family of growth factors and interacts with the EGF receptor to exert mitogenic activity in various cell types. HB-EGF is considered to play a key role in advanced brain functions in the central nervous system (Oyagi and Hara, 2012), as well as in tumor formation and other biological processes (Tsujioka et al., 2011).

## Chemokines

Chemokines are a family of small cytokines that can be classified into four groups, CXC, CC, C, and CX3C, according to their shared structural characteristics and four cysteine residues in conserved locations. Some chemokines can be induced during an immune response to promote cells of the immune system to reach the infection site, while others participate in controlling the migration of cells during normal tissue maintenance or development processes (Mantovani et al., 2006). These proteins interact with G protein-linked transmembrane receptors (called chemokine receptors) to exert their biological effects, including selective recruitment and activation of cells during inflammation, stimulation of leukocyte degranulation, and promotion of angiogenesis or angiostasis (Crijns et al., 2020). Locally produced chemokines bind to their chemokine receptors and induce leukocytes to adhere to endothelial cells, followed by extravasation of the leukocytes and subsequent migration to inflammation sites. To expose to the endothelial layer of blood vessels and form a concentration gradient, chemokines must bind to GAGs in endothelial cells and tissues (Johnson et al., 2005). In addition to PF4, which can lead to HIT, other important members of the chemokine family (e.g., stromal cell derived factor-1a (SDF-1a) and monocyte chemoattractant protein-1 (MCP-1) also bind to heparin, although with varying affinity and specificity. For example, studies have shown that HS is involved in binding and localization of SDF-1a to the cell surface. The sulfated-acetylated-sulfated domain of HS has subsequently been found to be recognized by a number of chemokines, such as IL-8, PF4 and MIP-1a (Gandhi and Mancera, 2008). Increasing evidence has confirmed that the binding and oligomerization of chemokines with GAGs are indispensable factors in the activity of chemokines *in vivo* (Proudfoot et al., 2003). Chemokines have been shown to be selective when interacting with GAGs. For example, for CCL5, the order of interaction strength is heparin, DS, HS, and CS, while mutant CCL5 has a reduced affinity for heparin. Studies have revealed that the main GAG-binding motifs on chemokines usually appear to be BBXB or BBBXXBX, where B and X represent a basic amino acid and any amino acid, respectively (Hileman et al., 1998). In addition, specific chemokine binding epitopes on GAGs have been found, such as the 2-O-sulfate group on the iduronic acid unit, which is necessary for formation of the GAG-dependent chemokine PF4 (Stringer and Gallagher, 1997).

## Receptor for Pathogens

The interaction of GAGs with specific proteins on the surface of a variety of pathogens, including viruses, bacteria, parasites and fungi, enables microorganisms to take the first step in establishing infection. Heparin-binding adhesins associated with intracellular pathogens, including gpB, gpC, and gpD of herpes simplex virus (HSV), gp120 of human immunodeficiency virus (HIV), herpesvirus filamentous hemagglutinin (FHA) of *Bordetella pertussis*, CS surface protein of *Plasmodium falciparum*, and the trypanosome adhesin penetrin, are likely the best studied proteins (Rostand and Esko, 1997). The protein sequences involved in the interaction between HSV and HS are

conserved and functional in other alpha-herpesvirus glycoproteins. CD4 is the main receptor of the HIV-1 envelope glycoprotein gp120. The V3 and C4 domains of gp120 contain positively charged regions that can be aggregated in the oligomeric gp120 to form HS binding sites. Heparin and HS binding to Tat protein is also important in HIV-1 infection. Tat protein is one of the essential proteins for HIV-1 replication and is believed to play a role in triggering cell infection. The smallest heparin fragment involved in Tat binding is a hexasaccharide. Therefore, heparin is a “multi-target” compound that can affect different aspects of HIV infection (Capila and Linhardt, 2002). Dengue virus causes several human diseases, such as dengue fever, and infection is initiated by an interaction between the dengue E protein and protein, lipids, or carbohydrate host receptor(s). E protein, which is the major antigen, is involved in viral attachment and other biological processes. The structures and antibody binding sites of dengue virus E protein have been elucidated, and the results showed that specific carbohydrate residues with sulfation are common structures shared by CS-E and heparin and could be essential determinants for controlling dengue virus entry mediated by the E protein (Kato et al., 2010).

## Other Proteins

In addition to the above proteins, other proteins can also interact with GAGs, such as adhesion molecules, lipid or membrane-binding proteins, amyloid proteins and proteases. Cell adhesion molecules (CAMs) are a group of molecules that mediate contact and binding between cells or between cells and the extracellular matrix and can be divided into four main groups: the integrin family, the immunoglobulin superfamily, selectins (P, E, L) and cadherins. The interaction of GAGs with adhesion proteins involves a variety of physiological and pathological processes. For example, heparin tetrasaccharides specifically block the interactions of L- and P-selectins with antigen sialyl Lewis X-containing ligands, which show anti-inflammatory activity *in vivo* and prevent the adhesion of colon cancer cells to L- and P-selectin (Norgard-Sumnicht et al., 1993). Annexins belong to a homologous protein family that is closely related to the cell membrane, indicating that they are involved in various processes. Calcium-dependent lectin activity (Kojima et al., 1996) and/or binding to specific glycoproteins and binding of annexins IV, V, and VI to GAGs (including heparin, HS, or CS) have been reported. This interaction is not only based on the affinity of annexin to polyanions but also has structural specificity. The interaction between sucrose octathiosulfate and annexin V was found to be weaker than that of heparin-derived octasaccharide and annexin V combined with heparin and HS but not CS, which confirmed the specificity of the annexin V-heparin interaction (Ishitsuka et al., 1998). Apolipoprotein E (ApoE) is an important protein that can regulate lipid transport in human plasma and in the brain. The interaction between ApoE and cell-surface HSPGs is important for the liver to absorb lipoprotein residues. HSPGs on the cell surface can locate ApoE-enriched remnant lipoproteins to receptors through rapid correlation and separation (Futamura et al., 2005), facilitating lipoprotein uptake. The

increased risk of AD associated with ApoE4 (Arg112, Arg158) appears to be associated with changes in amyloid- $\beta$  (A $\beta$ ) homeostasis (O’Callaghan et al., 2014). The interaction between ApoE and low-density lipoprotein receptor (the LDLR family) and HSPG is also important for cell signaling events (Tai et al., 2016). The binding of heparin to neutrophil elastase, a serine protease, is involved in inflammation and pulmonary diseases, and targeting their binding site has led to discovering promising synthetic mimetics to treat cystic fibrosis (Morla et al., 2019).

Both specific and nonspecific interactions in protein/glycosaminoglycan associations reconcile the two opposing views that emphasize either the dominance of structural complementarity, similar to that encountered in protein/protein interactions, or electrostatic forces. An enormous structural heterogeneity makes the search for specific protein “recognition elements” an extremely challenging undertaking. At the same time, the polyanionic nature of GAGs highlights the role of charge density as an important determinant of affinity to a range of proteins. To date, a large number of GAG-binding proteins have been identified. New cases of GAGs interacting with proteins are being discovered, and the update of selected families of GAG binding proteins is summarized in **Table 1**. Due to the structural heterogeneity of GAGs, the negatively charged GAGs tend to attract proteins in a nonspecific manner, and due to the specificity of different protein binding sequences, it is reasonable to believe that there are still numerous unknown GAG-protein interactions waiting to be discovered.

In summary, GAGs interplay with a wide range of important proteins. These proteins belong to different families and play various roles in physiological or pathological processes. Selectivity is the key when studying the binding between GAGs and proteins. Because the ionic force between negative charges of GAGs and positive charges on proteins is the basis of their interaction, abnormally highly charged GAGs, such as oversulfated CS or oversulfated HS, usually bind to basic proteins with high affinity but little specificity, which will cause uncontrollable side effects if being used as drugs. Elucidating and designing defined GAG sequence that specifically interacts to certain protein will be the only plausible way to develop promising new GAG therapeutics.

## ANALYTICAL TOOLS AND APPROACHES FOR CHARACTERIZATION OF GAG-PROTEIN INTERACTIONS

Obviously, safe and effective therapeutic intervention for diseases associated with GAGs depends on the selection of appropriate structures with the desired characteristics and a lack of harmful effects. For example, when using heparin or related compounds to treat COVID-19, the candidate drugs must have the ability to hinder the ACE2/S-Protein interaction with few deleterious effects (e.g., the HIT caused by binding to PF4). This work can be greatly facilitated by analytical tools that provide detailed information on the interactions between candidate drugs and their therapeutic targets.

**TABLE 1 |** An update of selected families of GAG binding proteins is summarized.

Heparin-binding protein	Related diseases	Physiological/Pathological role	Characteristics of GAG binding	References
Spike glycoprotein	COVID-19	HS is a necessary co-factor for SARS-CoV-2 infection by interacting with both SARS-CoV-2 spike glycoprotein and ACE2 in the RBD.	HS transforms spinous process structure into open conformation through the interaction of receptor binding domain of spike glycoprotein of SARS-CoV-2, so as to promote the binding of ACE2	Clausen et al. (2020)
Tau, $\alpha$ -synuclein, and A $\beta$	Neurodegenerative diseases	Tau and $\alpha$ -synuclein aggregates bind HSPGs on the cell surface to mediate uptake and intracellular seeding	Tau aggregates require a precise gag structure with definite GAG fractions at the N- and 6-O- positions be substituted with sulfate groups, while the binding of $\alpha$ -synuclein to a A $\beta$ is not so strict	Stopschinski et al. (2018)
HB-EGF	Cervical cancer	The expression of HB-EGF in tumor tissue was higher than that in stroma. Cervical cancer cells are the main source of HB-EGF.	HB-EGF is an important EGFR ligand in cervical cancer	Schrevel et al. (2017)
Transmembrane protein 184A (TMEM184A)	Angiogenesis	TMEM184A regulates angiogenesis by limiting endothelial cells proliferation and regulating extracellular growth	TMEM184A was identified as a heparin receptor in vascular cells. Heparin specifically binds to TMEM184A to induce anti-proliferative signaling	Farwell et al. (2017)
CXCL8	Inflammation	The binding of CXCL8 to GAGs on endothelial cell surfaces regulate neutrophil recruitment	Syndecan-4 (SDC4) was the potential proteoglycan co-receptor of CXCL8. CXCL8 binds to cell-surface HSPGs and leads to intracellular signal transduction in inflammatory tissue endothelium	Derler et al. (2017)
<i>Borrelia</i> glycosaminoglycan binding protein (Bgp)	Lyme disease caused by <i>Borrelia burgdorferi</i>	A variety of Bgp present in <i>B. burgdorferi</i> provide functional redundancy during infection, which highlights the importance of GAGs as co-receptors for spirochetes adhering to host cells	The binding efficiency of Bgp to heparin was higher than that of chondroitin sulfate C	Schlachter et al. (2018)
FGF-2	Ischemic heart repair	FGF-2 promotes angiogenesis after MI. HSPG enhances cell adhesion, promotes the biological activity of FGF-2 in angiogenesis, and protects FGF-2 from enzymatic hydrolysis	The specific binding of HSPG to FGF-2 protein 6 times stronger than that of FGF-2 and heparin	Shi et al. (2019)
Receptor protein tyrosine phosphatase (RPTP $\sigma$ )	Neural development and regeneration	RPTP $\sigma$ has important functions in modulating neural development and regeneration	Both HS and CS bind to a series of lys residues located in the first Ig domain of RPTP $\sigma$ . RPTP $\sigma$ was aggregated by GAGs rich in 4,6-O-disulfated disaccharides	Katagiri et al. (2018)
C-type lectin 14a (CLEC14A)	Angiogenesis	CLEC14A is up-regulated during tumor angiogenesis and regulates endothelial cell migration and adhesion <i>in vitro</i> and angiogenesis <i>in vivo</i>	C-type lectin domain of CLEC14A binds 1:1 to heparin with nanomolar affinity. CLEC14A prefers highly charged polysaccharides	Sandoval et al. (2020)
Keratinocyte-derived chemokine (KC or mCXCL1) and macrophage inflammatory protein 2 (MIP2 or mCXCL2)	Inflammation	KC and MIP2 play important roles in transporting neutrophils to infected and injured sites	Different combinations of residues from the N-loop, 40s turn, $\beta_3$ -strand, and C-terminal helix form a binding surface within a monomer and both conserved residues. The binding interaction is mediated by both conserved residues and residues specific to chemokines	Sepuru et al. (2018)
Pre-S region of HBV envelope proteins	Hepatitis B	The human hepatic cell-binding site (i.e., the sodium taurocholate co-transporting polypeptide (NTCP)-binding site, with myristoylated pre-S1 (2–47)) and the low pH-dependent fusogenic domain (pre-S1 (9–24)) are required for targeting and endosomal escape, respectively	A novel heparin-binding site (pre-S1 (30–42)) in the N-terminal half of the pre-S1 region may interact with cell-surface HSPG. The amino acid residues Asp-31, Trp-32, and Asp-33 are essential for heparin-binding activity	Liu et al. (2018)

Currently, numerous analytical approaches have been developed and applied to reveal the molecular mechanism and binding sequence of GAG-protein complexes (Yang and Chi,

2017). Affinity approaches, such as affinity chromatography, surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC), are used to measure the binding strength

**TABLE 2 |** The recent advances in mass spectrometry based analytical tools is summarized.

Method	Principle	Applications	References
Mass spectrometry combined with gas-phase ion manipulation technique	Intact macromolecules or macromolecular complexes are directly ionized from non-denaturing solvent, and key noncovalent interactions that hold the complexes together can be preserved for MS analysis in the gas phase	Characterizing biomacromolecular structure and interactions under physiological conditions. For example, obtaining meaningful information about the complex formed by ACE2 and S protein, and the role of heparin in disrupting ACE2/RBD binding	Yang et al. (2020)
SEC-MS	When the SEC-MS system is applied to heparin, a series of oligomers with different sulfation levels can be generated	Enzymatic lysis was used to product the proteinbound chains, then mass spectrometry detection (SEC/MS) can detect the tight association with the protein, including the characterization, oligomer length and the number of incorporated sulfate and acetyl groups	Niu et al. (2020)
Cross-linking MS	Cross-linking with MS approach has been recently recognized as a powerful tool to study protein-protein interaction. It can also study GAG-protein interactions by "locking" binding motifs together through covalently cross-linking carboxyl groups of GAGs to amine side chains of protein	The cross-linking technique locks down the binding motifs of GAGs and proteins through covalent reactions. For example, the carboxyl groups of GAGs can be activated by EDC and sulfo-NHS, then form a zero-length linkage with the amine side chains of proteins. After digestion by protease, LC/MS/MS analysis showed that the binding motif was oligosaccharide peptide conjugate	Yugandhar et al. (2020)
Limited Proteolysis in the absence of denaturation, heparin-Affinity chromatography, and high-resolution LC-MS/MS proteomics (LPHAMS)	By using suboptimal conditions for proteolysis, limited cleavage occurs at exposed hinges or loops, resulting in the release of intact protein domains. Liberated domains by chromatography on heparin-affinity resin would identify potential HSBPs and enrich HS-binding domains	Identification and characterization of membrane-anchored and extracellular proteins that bind HS. Application of LPHAMS has led to the identification of large number of HSBPs. In many cases, this method reveals subdomains that promote HS binding	Sandoval et al. (2020)

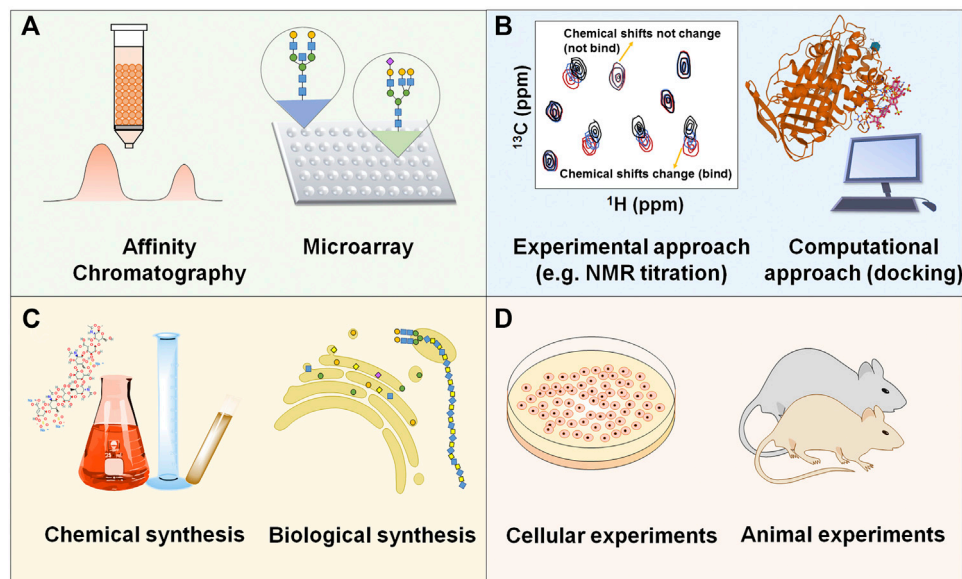
between GAGs and proteins. NMR spectroscopy and X-ray crystallography can present extremely valuable information about GAG-protein interactions, providing structural and conformational data that are useful in identifying the precise contact points between interacting molecules. The microarray platform and molecular docking are powerful tools to screen protein interactions against large GAG structure libraries, and intricate dynamic details of molecular-level events can be visualized with a relatively small time and cost investment. However, the structural heterogeneity of GAGs and the extensive glycosylation of the proteins involved still make discovery of the specificity of the binding sequence challenging.

MS techniques have several unique advantages in the characterization of GAG-protein complexes due to their superior sensitivity, tolerance of lower sample purity and ability to characterize amino acid/sugar residues and modification. Ion mobility spectroscopy (IMS) (Gray et al., 2016), hydroxyl radical footprinting (HRF) (Li et al., 2015) and cross-linking MS (Yang et al., 2012) have been used to study the interactions between GAGs and proteins. Recently, native mass spectrometry has been used as a tool to support mechanistic study of drug/therapeutic target interactions (Tong and Wang, 2018). Using gas-phase ion manipulation (limited charge reduction) and molecular modeling to supplement native MS has allowed obtainment of meaningful information about the complex formed by ACE2 and S protein and the role of heparin in destroying ACE2/RBD binding (Yang et al., 2020). A top-down approach was used to maintain the chemical diversity of heparin by allowing complex long chains to interact with the target

protein. After enzymolysis, the protein-binding heparin chains were analyzed using size exclusion chromatography with online mass spectrometry detection (SEC/MS) (Niu et al., 2020), which revealed the oligomers that were not cleaved by lysis due to their binding to the protein and enabled characterization of chain length and sulfate and acetyl groups. Some of the latest mass spectrometry techniques and their applications in GAG-protein interactions are shown in **Table 2**.

Another emerging field is developing computational tools to facilitate the study of GAG-protein interactions. Unlike proteins, GAGs are highly charged and highly flexible at the aspect of confirmation. Therefore, it is difficult to obtain high quality cocrystals of GAG-protein complexes. Computational approaches provided an alternative way to predict the binding patterns and residues contributed the binding. A systematic study has been carried out by computationally characterizing all known GAG-protein bindings from the Protein Data Bank, which proved the feasibility of the computational methodology (Bojarski et al., 2019). Furthermore, a GAG-Dock methodology has been developed to evaluate the binding between various GAG ligands and receptors that are essential in axonal growth, and their plausible structures were provided (Griffith et al., 2017). Our group has also applied the molecular docking approach to explain the pharmacokinetic behavior of heparin in diabetic patients by simulating the binding of heparin and glycosylated human serum albumin (Qiu et al., 2020). Computational study of GAG-protein binding is also useful in developing potential therapeutics, such as sulfated small molecules mimicking the function of GAGs (Nagarajan et al.,





**FIGURE 3 |** Schematic strategy of developing novel therapeutics based on the specific interaction between GAGs and proteins **(A)** Discovering GAG-protein binding. **(B)** Elucidating the molecular mechanism between the binding. **(C)** Synthesizing specific GAG oligosaccharides or analogs. **(D)** Evaluating the efficacy and toxicity *in vitro* and *in vivo*.

2017). Indeed, the computational methodology has become extremely useful and easily accessible to non-computational researchers (Sankaranarayanan et al., 2018).

## THERAPEUTICS TARGETING SPECIFIC GAG-PROTEIN INTERACTIONS

The eventual goal of studying interaction between GAGs and proteins is to develop novel therapeutics from this promising but inadequately explored field. The schematic strategy is shown in **Figure 3**. Actually, some therapeutics targeting specific GAG-protein interactions, including GAG oligosaccharides and synthetic analogs, removal or modification of GAGs by enzymes, exogenous heparin/HS or synthetic GAG mimetics as competitive inhibitors, cationic proteins and polymers as HS antagonists, and small molecule antagonists of heparin and HS, are currently being developed or have been applied to treat related diseases.

The application of heparin oligosaccharides and synthetic analogs is an important aspect of the clinical treatment of many diseases. HP and LMWH have long been used as anticoagulants (Hirsh et al., 2001). The synthetic pentasaccharide Arixtra (Fondaparinux) binds to AT and has better efficacy at low doses (Walenga et al., 2002). Additionally, some therapeutic applications of heparin and its derivatives beyond anticoagulation have been explored in patients with bronchial asthma, chronic obstructive pulmonary disease (COPD) and cancer. The anti-inflammatory, antioxidant, and lysogenic effects of heparin administered via the inhalation pathway may alter the progression of COPD and asthma (Shute et al., 2018).

Several different strategies to target the interactions of HS and proteins have been explored, including HS removal or modification by enzymes (Rek et al., 2009). Some heparin enzymes (such as bacterial heparinases and mammalian endosulfatases) have been shown to be potential inhibitors of HS-protein interactions. Heparinase therapy has also been used to inhibit tumor growth/metastasis and amyloid-related diseases. Cells treated with heparinase can resist the attachment or entry of several HS-binding pathogens, including viruses, bacteria, and parasites (Weiss et al., 2017).

Another way to inhibit HS-protein interactions is to use exogenous heparin/HS or synthetic GAG analogs as competitive inhibitors. Exogenous addition of heparin and HS chains can inhibit infection of host cells with HS-binding pathogens, such as HSV, HPV, hepatitis B and various bacteria. Additionally, cancer cell growth and metastasis can be blocked by HS and heparin. LMWHs and HS mimetics (Lee et al., 1999), such as rhamnan sulfate, have shown anticancer and antiviral activity, which was promising when tested *in vitro*.

Cationic proteins and foldamers have been used as antagonists of HS-protein interactions. These molecules depend on electrostatic interactions between their positively charged functional groups and the high anion sulfate and carboxylic acid groups of heparin and HS. Lactoferrin (Lonnerdal and Iyer, 1995) has been tested to neutralize heparin and antagonize certain HS-protein interactions. Protamine (Taylor and Folkman, 1982) has been demonstrated to be a potent antagonist of the GAG-protein interaction and has been used clinically to reverse anticoagulants.

Certain small molecule drugs have been developed as HS-protein antagonists due to their specific characteristics and advantages. For example, a dispirotriperazine derivative (DSTP 27) (Schmidtke et al., 2003) was found to bind cell surface HS and

inhibit attachment and absorption of some viruses and to block HS-dependent viral attachment of an HPV virus in the long term.

## CONCLUSION AND MARKS

GAGs are involved in a large number of biological processes and play an important role in growth and development, maintaining homeostasis and resisting disease. GAGs affect cell adhesion, migration, signal transduction and other biological activities through interactions with proteins, such as growth factors and adhesion factors, thereby affecting numerous physiological activities. Due to the diversity of the types and functions of the proteins that interact with GAGs, GAGs exert a variety of biological functions. The occurrence and development of many diseases, from the invasion of pathogens to the occurrence and development of tumors, are related to GAGs. Elucidation of the specific sequence and mechanism by which GAGs interact with proteins is essential for finding novel therapeutics targeting specific GAG-protein interactions.

## INSIGHTS AND FUTURE

The research, development and market of carbohydrate-based drugs, especially GAG-based drugs, are far behind the protein-based drugs. Except for heparin drugs as anticoagulants, few GAGs have been widely used in clinic, although GAGs exhibit a

wide range of bioactivities. However, the situation is changing now. With the advances of analytical tools and synthetic/biosynthetic approaches, identifying specific sequence and obtaining sufficient structure uniform GAG oligosaccharides become feasible. In the next five to ten years, we can expect quite a few GAG or GAG mimetics proceed to clinical trials. It will boost the GAG study and lead to new solutions for diseases that are difficult to be cured by current small molecule or protein drugs.

## AUTHOR CONTRIBUTIONS

All the authors contributed to this work and have approved it for publication.

## FUNDING

This study was supported by the National Natural Science Foundation of China (21877072) and the Natural Science Foundation of Shandong Province (ZR2019MB014).

## ACKNOWLEDGMENTS

The authors would like to thank Xiaojun Sun for proofreading the manuscript.

## REFERENCES

- Anttila, M. A., Tammi, R. H., Tammi, M. I., Syrjänen, K. J., Saarikoski, S. V., and Kosma, V. M. (2000). High levels of stromal hyaluronan predict poor disease outcome in epithelial ovarian cancer. *Cancer Res.* 60 (1), 150–155.
- Auvinen, P., Tammi, R., Parkkinen, J., Tammi, M., Agren, U., Johansson, R., et al. (2000). Hyaluronan in peritumoral stroma and malignant cells associates with breast cancer spreading and predicts survival. *Am. J. Pathol.* 156 (2), 529–536. doi:10.1016/S0002-9440(10)64757-8
- Banerjee, M., Copp, J., Vuga, D., Marino, M., Chapman, T., van der Geer, P., et al. (2004). GW domains of the *Listeria monocytogenes* invasion protein InlB are required for potentiation of met activation. *Mol. Microbiol.* 52 (1), 257–271. doi:10.1111/j.1365-2958.2003.03968.x
- Baranova, N. S., Nileback, E., Haller, F. M., Briggs, D. C., Svedhem, S., Day, A. J., et al. (2011). The inflammation-associated protein TSG-6 cross-links hyaluronan via hyaluronan-induced TSG-6 oligomers. *J. Biol. Chem.* 286 (29), 25675–25686. doi:10.1074/jbc.M111.247395
- Barth, H., Schafer, C., Adah, M. I., Zhang, F., Linhardt, R. J., Toyoda, H., et al. (2003). Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. *J. Biol. Chem.* 278 (42), 41003–41012. doi:10.1074/jbc.M302267200
- Bojarski, K. K., Becher, J., Riemer, T., Lemmnitzer, K., Moller, S., Schiller, J., et al. (2019). Synthesis and in silico characterization of artificially phosphorylated glycosaminoglycans. *J. Mol. Struct.* 1197, 401–416. doi:10.1016/j.molstruc.2019.07.064
- Boyd, A. P., Sory, M. P., Iriarte, M., and Cornelis, G. R. (1998). Heparin interferes with translocation of Yop proteins into HeLa cells and binds to LcrG, a regulatory component of the *Yersinia* Yop apparatus. *Mol. Microbiol.* 27 (2), 425–436. doi:10.1046/j.1365-2958.1998.00691.x
- Cai, Y., Zhang, X., Shen, J., Jiang, B., Hu, D., and Zhao, M. (2020). Heparin-binding protein: a novel biomarker linking four different cardiovascular diseases. *Cardiol. Res. Pract.* 2020, 9575373. doi:10.1155/2020/9575373
- Capila, I., and Linhardt, R. J. (2002). Heparin-protein interactions. *Angew. Chem. Int. Ed. Engl.* 41 (3), 391–412. doi:10.1002/1521-3773(20020201)41:3<390::AID-ANIE390>3.0.CO;2-B
- Chazeirat, T., Denamur, S., Bojarski, K. K., Andrault, P. M., Sizaret, D., Zhang, F., et al. (2021). The abnormal accumulation of heparan sulfate in patients with mucopolysaccharidosis prevents the elastolytic activity of cathepsin V. *Carbohydr. Polym.* 253, 117261. doi:10.1016/j.carbpol.2020.117261
- Clausen, T. M., Christoffersen, S., Dahlback, M., Langkilde, A. E., Jensen, K. E., Resende, M., et al. (2012). Structural and functional insight into how the *Plasmodium falciparum* VAR2CSA protein mediates binding to chondroitin sulfate A in placental malaria. *J. Biol. Chem.* 287 (28), 23332–23345. doi:10.1074/jbc.M112.348839
- Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* 183 (4), 1043–1057. doi:10.1016/j.cell.2020.09.033
- Crijns, H., Vanheule, V., and Proost, P. (2020). Targeting chemokine-glycosaminoglycan interactions to inhibit inflammation. *Front. Immunol.* 11, 483. doi:10.3389/fimmu.2020.00483
- Cui, H., Hung, A. C., Klaver, D. W., Suzuki, T., Freeman, C., and Narkowicz, C. (2011). Effects of heparin and enoxaparin on APP processing and Abeta production in primary cortical neurons from Tg2576 mice. *PLoS One* 6 (7), e23007. doi:10.1371/journal.pone.0023007
- Dalrymple, N., and Mackow, E. R. (2011). Productive dengue virus infection of human endothelial cells is directed by heparan sulfate-containing proteoglycan receptors. *J. Virol.* 85 (18), 9478–9485. doi:10.1128/JVI.05008-11
- De Pasquale, V., and Pavone, L. M. (2020). Heparan sulfate proteoglycan signaling in tumor microenvironment. *Int. J. Mol. Sci.* 21 (18), 6588. doi:10.3390/ijms21186588
- Derler, R., Gesslbauer, B., Weber, C., Strutzmann, E., Miller, I., and Kungl, A. (2017). Glycosaminoglycan-mediated downstream signaling of CXCL8 binding to endothelial cells. *Int. J. Mol. Sci.* 18 (12), 2605. doi:10.3390/ijms18122605

- Dymarska, M., Juros, W., Janeczko, T., and Kostrzewa-Suslow, E. (2016). Hyaluronic acid. Structure, properties and uses. *Przem. Chem.* 95 (4), 814–816. doi:10.15199/62.2016.4.19
- Edwards, I. J., Wagner, J. D., Vogl-Willis, C. A., Litwak, K. N., and Cefalu, W. T. (2004). Arterial heparan sulfate is negatively associated with hyperglycemia and atherosclerosis in diabetic monkeys. *Cardiovasc. Diabetol.* 3, 6. doi:10.1186/1475-2840-3-6
- Fager, G., Camejo, G., Olsson, U., Ostergren-Lunden, G., and Bondjers, G. (1992). Heparin-like glycosaminoglycans influence growth and phenotype of human arterial smooth muscle cells *in vitro*. II. The platelet-derived growth factor A-chain contains a sequence that specifically binds heparin. *Vitro Cell. Dev. Biol.* 28A (3 Pt 1), 176–180. doi:10.1007/BF02631088
- Fannon, M., Forsten, K. E., and Nugent, M. A. (2000). Potentiation and inhibition of bFGF binding by heparin: a model for regulation of cellular response. *Biochemistry* 39 (6), 1434–1445. doi:10.1021/bi991895z
- Farwell, S. L. N., Reylander, K. G., Iovine, M. K., and Lowe-Krentz, L. J. (2017). Novel heparin receptor transmembrane protein 184a regulates angiogenesis in the adult zebrafish caudal fin. *Front. Physiol.* 8, 671. doi:10.3389/fphys.2017.00671
- Fatoux-Ardore, M., Peysselon, F., Weiss, A., Bastien, P., Pralong, F., and Ricard-Blum, S. (2014). Large-scale investigation of Leishmania interaction networks with host extracellular matrix by surface plasmon resonance imaging. *Infect. Immun.* 82 (2), 594–606. doi:10.1128/IAI.01146-13
- Futamura, M., Dhanasekaran, P., Handa, T., Phillips, M. C., Lund-Katz, S., and Saito, H. (2005). Two-step mechanism of binding of apolipoprotein E to heparin: implications for the kinetics of apolipoprotein E-heparan sulfate proteoglycan complex formation on cell surfaces. *J. Biol. Chem.* 280 (7), 5414–5422. doi:10.1074/jbc.M411719200
- Gandhi, N. S., and Mancera, R. L. (2008). The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des.* 72 (6), 455–482. doi:10.1111/j.1747-0285.2008.00741.x
- Garantziotis, S., Brezina, M., Castelnovo, P., and Drago, L. (2016). The role of hyaluronan in the pathobiology and treatment of respiratory disease. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 310 (9), L785–L795. doi:10.1152/ajplung.00168.2015
- Gitay-Goren, H., Soker, S., Vlodavsky, I., and Neufeld, G. (1992). The binding of vascular endothelial growth factor to its receptors is dependent on cell surface-associated heparin-like molecules. *J. Biol. Chem.* 267 (9), 6093–6098. doi:10.1016/s0021-9258(18)42666-x
- Gray, C. J., Thomas, B., Upton, R., Migas, L. G., Eysers, C. E., Barran, P. E., et al. (2016). Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis. *Biochim. Biophys. Acta* 1860 (8), 1688–1709. doi:10.1016/j.bbagen.2016.02.003
- Green, J. V., Orsborn, K. I., Zhang, M., Tan, Q. K., Greis, K. D., Porollo, A., et al. (2013). Heparin-binding motifs and biofilm formation by *Candida albicans*. *J. Infect. Dis.* 208 (10), 1695–1704. doi:10.1093/infdis/jit391
- Griffith, A. R., Rogers, C. J., Miller, G. M., Abrol, R., Hsieh-Wilson, L. C., and Goddard, W. A., 3rd (2017). Predicting glycosaminoglycan surface protein interactions and implications for studying axonal growth. *Proc. Natl. Acad. Sci. U.S.A.* 114 (52), 13697–13702. doi:10.1073/pnas.1715093115
- Guerrini, M., Mourier, P. A., Torri, G., and Viskov, C. (2014). Antithrombin-binding oligosaccharides: structural diversities in a unique function? *Glycoconj. J.* 31 (6–7), 409–416. doi:10.1007/s10719-014-9543-9
- Hayashida, A., Amano, S., Gallo, R. L., Linhardt, R. J., Liu, J., and Park, P. W. (2015). 2-O-Sulfated domains in syndecan-1 heparan sulfate inhibit neutrophil cathelicidin and promote *Staphylococcus aureus* corneal infection. *J. Biol. Chem.* 290 (26), 16157–16167. doi:10.1074/jbc.M115.660852
- Hayashida, A., Amano, S., and Park, P. W. (2011). Syndecan-1 promotes *Staphylococcus aureus* corneal infection by counteracting neutrophil-mediated host defense. *J. Biol. Chem.* 286 (5), 3288–3297. doi:10.1074/jbc.M110.185165
- Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998). Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays* 20 (2), 156–167. doi:10.1002/(SICI)1521-1878(199802)20:2<156::AID-BIES8>3.0.CO;2-R
- Hirsh, J., Warkentin, T. E., Shaughnessy, S. G., Anand, S. S., Halperin, J. L., Raschke, R., et al. (2001). Heparin and low-molecular-weight heparin: mechanisms of action, pharmacokinetics, dosing, monitoring, efficacy, and safety. *Chest* 119 (Suppl. 1), 64S–94S. doi:10.1378/chest.119.1\_suppl.64s
- Ho, P. J., and Siordia, J. A. (2016). Dabigatran approaching the realm of heparin-induced thrombocytopenia. *Blood Res.* 51 (2), 77–87. doi:10.5045/br.2016.51.277
- Huang, T. Y., Irene, D., Zulueta, M. M., Tai, T. J., Lain, S. H., Cheng, C. P., et al. (2017). Structure of the complex between a heparan sulfate octasaccharide and mycobacterial heparin-binding hemagglutinin. *Angew. Chem. Int. Ed. Engl.* 56 (15), 4192–4196. doi:10.1002/anie.201612518
- Huntington, J. A. (2006). Shape-shifting serpins—advantages of a mobile mechanism. *Trends Biochem. Sci.* 31 (8), 427–435. doi:10.1016/j.tibs.2006.06.005
- Huynh, M. B., Ouidja, M. O., Chantepie, S., Carpentier, G., Maiza, A., Zhang, G., et al. (2019). Glycosaminoglycans from Alzheimer's disease hippocampus have altered capacities to bind and regulate growth factors activities and to bind tau. *PLoS One* 14 (1), e0209573. doi:10.1371/journal.pone.0209573
- Ishitsuka, R., Kojima, K., Utsumi, H., Ogawa, H., and Matsumoto, I. (1998). Glycosaminoglycan binding properties of annexin IV, V, and VI. *J. Biol. Chem.* 273 (16), 9935–9941. doi:10.1074/jbc.273.16.9935
- Johnson, Z., Proudfoot, A. E., and Handel, T. M. (2005). Interaction of chemokines and glycosaminoglycans: a new twist in the regulation of chemokine function with opportunities for therapeutic intervention. *Cytokine Growth Factor Rev.* 16 (6), 625–636. doi:10.1016/j.cytogr.2005.04.006
- Kamhi, E., Joo, E. J., Dordick, J. S., and Linhardt, R. J. (2013). Glycosaminoglycans in infectious disease. *Biol. Rev.* 88 (4), 928–943. doi:10.1111/brv.12034
- Karjalainen, J. M., Tammi, R. H., Tammi, M. I., Eskelinen, M. J., Agren, U. M., Parkkinen, J. J., et al. (2000). Reduced level of CD44 and hyaluronan associated with unfavorable prognosis in clinical stage I cutaneous melanoma. *Am. J. Pathol.* 157 (3), 957–965. doi:10.1016/S0002-9440(10)64608-1
- Katagiri, Y., Morgan, A. A., Yu, P., Bangayan, N. J., Junka, R., and Geller, H. M. (2018). Identification of novel binding sites for heparin in receptor protein-tyrosine phosphatase (RPTPsigma): implications for proteoglycan signaling. *J. Biol. Chem.* 293 (29), 11639–11647. doi:10.1074/jbc.RA118.003081
- Kato, D., Era, S., Watanabe, I., Arihara, M., Sugiura, N., Kimata, K., et al. (2010). Antiviral activity of chondroitin sulphate E targeting dengue virus envelope protein. *Antiviral Res.* 88 (2), 236–243. doi:10.1016/j.antiviral.2010.09.002
- Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., et al. (2020). Glycosaminoglycan binding motif at S1/S2 proteolytic cleavage site on spike glycoprotein may facilitate novel coronavirus (SARS-CoV-2) host cell entry. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.04.14.041459
- Kines, R. C., Thompson, C. D., Lowy, D. R., Schiller, J. T., and Day, P. M. (2009). The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proc. Natl. Acad. Sci. U.S.A.* 106 (48), 20458–20463. doi:10.1073/pnas.0908502106
- Kjellen, L., and Lindahl, U. (2018). Specificity of glycosaminoglycan-protein interactions. *Curr. Opin. Struct. Biol.* 50, 101–108. doi:10.1016/j.sbi.2017.12.011
- Kleeff, J., Ishiwata, T., Kumbasar, A., Friess, H., Buchler, M. W., Lander, A. D., et al. (1998). The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *J. Clin. Invest.* 102 (9), 1662–1673. doi:10.1172/JCI4105
- Knelson, E. H., Nee, J. C., and Blobel, G. C. (2014). Heparan sulfate signaling in cancer. *Trends Biochem. Sci.* 39 (6), 277–288. doi:10.1016/j.tibs.2014.03.001
- Kojima, K., Yamamoto, K., Irimura, T., Osawa, T., Ogawa, H., and Matsumoto, I. (1996). Characterization of carbohydrate-binding protein p33/41: relation with annexin IV, molecular basis of the doublet forms (p33 and p41), and modulation of the carbohydrate binding activity by phospholipids. *J. Biol. Chem.* 271 (13), 7679–7685. doi:10.1074/jbc.271.13.7679
- Kosunen, A., Ropponen, K., Kellokoski, J., Pukkila, M., Virtaniemi, J., Valtonen, H., et al. (2004). Reduced expression of hyaluronan is a strong indicator of poor survival in oral squamous cell carcinoma. *Oral Oncol.* 40 (3), 257–263. doi:10.1016/j.oraloncology.2003.08.004
- Kowitsch, A., Zhou, G., and Groth, T. (2018). Medical application of glycosaminoglycans: a review. *J. Tissue Eng. Regen. Med.* 12 (1), e23–e41. doi:10.1002/term.2398
- Lang, J., Yang, N., Deng, J., Liu, K., Yang, P., Zhang, G., et al. (2011). Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. *PLoS One* 6 (8), e23710. doi:10.1371/journal.pone.0023710
- Lee, J. B., Hayashi, K., Hayashi, T., Sankawa, U., and Maeda, M. (1999). Antiviral activities against HSV-1, HCMV, and HIV-1 of rhamnan sulfate from

- Monostroma latissimum*. *Planta Med.* 65 (5), 439–441. doi:10.1055/s-2006-960804
- Lee, J., Wee, S., Gunaratne, J., Chua, R. J., Smith, R. A., Ling, L., et al. (2015). Structural determinants of heparin-transforming growth factor-beta1 interactions and their effects on signaling. *Glycobiology* 25 (12), 1491–1504. doi:10.1093/glycob/cwv064
- Lewis, E. J., and Xu, X. (2008). Abnormal glomerular permeability characteristics in diabetic nephropathy: implications for the therapeutic use of low-molecular weight heparin. *Diabetes Care* 31 (Suppl. 2), S202–S207. doi:10.2337/dc08-s251
- Li, Z., Moniz, H., Wang, S., Ramiah, A., Zhang, F., Moremen, K. W., et al. (2015). High structural resolution hydroxyl radical protein footprinting reveals an extended Robo1-heparin binding interface. *J. Biol. Chem.* 290 (17), 10729–10740. doi:10.1074/jbc.M115.648410
- Liang, O. D., Ascencio, F., Fransson, L. A., and Wadstrom, T. (1992). Binding of heparan sulfate to *Staphylococcus aureus*. *Infect. Immun.* 60 (3), 899–906. doi:10.1128/IAI.60.3.899-906.1992
- Lindahl, U., Couchman, J., Kimata, K., Esko, J. D., Varki, R. A., Cummings, R. D., et al. (2015). “Proteoglycans and sulfated glycosaminoglycans,” in *Essentials of glycobiology*. Editor: S. DarvillSchnaar (New York, NY: Cold Spring Harbor), 207–221.
- Liu, J., Li, J., Arnold, K., Pawlinski, R., and Key, N. S. (2020). Using heparin molecules to manage COVID-2019. *Res. Pract. Thromb. Haemost.* 4 (4), 518–523. doi:10.1002/rth2.12353
- Liu, Q., Somya, M., Iijima, M., Tatematsu, K., and Kuroda, S. (2018). A hepatitis B virus-derived human hepatic cell-specific heparin-binding peptide: identification and application to a drug delivery system. *Biomater. Sci.* 7 (1), 322–335. doi:10.1039/c8bm01134f
- Lonnerdal, B., and Iyer, S. (1995). Lactoferrin: molecular structure and biological function. *Annu. Rev. Nutr.* 15, 93–110. doi:10.1146/annurev.nu.15.070195.000521
- Luan, Z. G., Naranpure, M., and Ma, X. C. (2014). Treatment of low molecular weight heparin inhibits systemic inflammation and prevents endotoxin-induced acute lung injury in rats. *Inflammation* 37 (3), 924–932. doi:10.1007/s10753-014-9812-6
- Ma, S. N., Mao, Z. X., Wu, Y., Liang, M. X., Wang, D. D., Chen, X., et al. (2020). The anti-cancer properties of heparin and its derivatives: a review and prospect. *Cell Adh. Migr.* 14 (1), 118–128. doi:10.1080/1936918.2020.1767489
- Maimone, M. M., and Tollefsen, D. M. (1990). Structure of a dermatan sulfate hexasaccharide that binds to heparin cofactor II with high affinity. *J. Biol. Chem.* 265 (30), 18263–18271. doi:10.1016/S0021-9258(17)44747-8
- Mantovani, A., Bonecchi, R., and Locati, M. (2006). Tuning inflammation and immunity by chemokine sequestration: decoys and more. *Nat. Rev. Immunol.* 6 (12), 907–918. doi:10.1038/nri1964
- Matsuda, K., Maruyama, H., Guo, F., Kleeff, J., Itakura, J., Matsumoto, Y., et al. (2001). Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer Res.* 61 (14), 5562–5569.
- Milewska, A., Zarebski, M., Nowak, P., Stozek, K., Potempa, J., and Pyrc, K. (2014). Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. *J. Virol.* 88 (22), 13221–13230. doi:10.1128/JVI.02078-14
- Morla, S. (2019). Glycosaminoglycans and glycosaminoglycan mimetics in cancer and inflammation. *Int. J. Mol. Sci.* 20 (8), 1963. doi:10.3390/ijms20081963
- Morla, S., Sankaranarayanan, N. V., Afosah, D. K., Kumar, M., Kummarapurugu, A. B., Voynow, J. A., et al. (2019). On the process of discovering leads that target the heparin-binding site of neutrophil elastase in the sputum of cystic fibrosis patients. *J. Med. Chem.* 62 (11), 5501–5511. doi:10.1021/acs.jmedchem.9b00379
- Munoz, E. M., and Linhardt, R. J. (2004). Heparin-binding domains in vascular biology. *Arterioscler. Thromb. Vasc. Biol.* 24 (9), 1549–1557. doi:10.1161/01.ATV.0000137189.22999.3f
- Mycroft-West, C. J., Devlin, A. J., Cooper, L. C., Procter, P., Miller, G. J., Fernig, D. G., et al. (2020). Inhibition of BACE1, the beta-secretase implicated in Alzheimer's disease, by a chondroitin sulfate extract from *Sardina pilchardus*. *Neural Regen. Res.* 15 (8), 1546–1553. doi:10.4103/1673-5374.274341
- Myint, K. M., Yamamoto, Y., Doi, T., Kato, I., Harashima, A., Yonekura, H., et al. (2006). RAGE control of diabetic nephropathy in a mouse model: effects of RAGE gene disruption and administration of low-molecular weight heparin. *Diabetes* 55 (9), 2510–2522. doi:10.2337/db06-0221
- Nagarajan, B., Sankaranarayanan, N. V., Patel, B. B., and Desai, U. R. (2017). A molecular dynamics-based algorithm for evaluating the glycosaminoglycan mimicking potential of synthetic, homogenous, sulfated small molecules. *PLoS One* 12 (2), e0171619. doi:10.1371/journal.pone.0171619
- Niu, C., Zhao, Y., Bobst, C. E., Savinov, S. N., and Kaltashov, I. A. (2020). Identification of protein recognition elements within heparin chains using enzymatic foot-printing in solution and online SEC/MS. *Anal. Chem.* 92 (11), 7565–7573. doi:10.1021/acs.analchem.0c00115
- Norgard-Sumnicht, K. E., Varki, N. M., and Varki, A. (1993). Calcium-dependent heparin-like ligands for L-selectin in nonlymphoid endothelial cells. *Science* 261 (5120), 480–483. doi:10.1126/science.7687382
- O'Callaghan, P., Noborn, F., Sehlin, D., Li, J. P., Lannfelt, L., Lindahl, U., et al. (2014). Apolipoprotein E increases cell association of amyloid-beta 40 through heparan sulfate and LRP1 dependent pathways. *Amyloid* 21 (2), 76–87. doi:10.3109/13506129.2013.879643
- Olson, S. T., Richard, B., Izaguirre, G., Schedin-Weiss, S., and Gettins, P. G. (2010). Molecular mechanisms of antithrombin-heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors. *Biochimie* 92 (11), 1587–1596. doi:10.1016/j.biochi.2010.05.011
- Ottenheijm, C. A., Jenniskens, G. J., Geraedts, M. C., Hafmans, T., Heunks, L. M., van Kuppevelt, T. H., et al. (2007). Diaphragm dysfunction in chronic obstructive pulmonary disease: a role for heparan sulphate? *Eur. Respir. J.* 30 (1), 80–89. doi:10.1183/09031936.00125106
- Oyagi, A., and Hara, H. (2012). Essential roles of heparin-binding epidermal growth factor-like growth factor in the brain. *CNS Neurosci. Ther.* 18 (10), 803–810. doi:10.1111/j.1755-5949.2012.00371.x
- Parish, C. R. (2006). The role of heparan sulphate in inflammation. *Nat. Rev. Immunol.* 6 (9), 633–643. doi:10.1038/nri1918
- Park, P. W., Pier, G. B., Hinkes, M. T., and Bernfield, M. (2001). Exploitation of syndecan-1 shedding by *Pseudomonas aeruginosa* enhances virulence. *Nature* 411 (6833), 98–102. doi:10.1038/35075100
- Pillarsetti, S. (2000). Lipoprotein modulation of subendothelial heparan sulfate proteoglycans (perlecan) and atherogenicity. *Trends Cardiovasc. Med.* 10 (2), 60–65. doi:10.1016/s1050-1738(00)00048-7
- Pirinen, R., Tammi, R., Tammi, M., Hirvikoski, P., Parkkinen, J. J., Johansson, R., et al. (2001). Prognostic value of hyaluronan expression in non-small-cell lung cancer: increased stromal expression indicates unfavorable outcome in patients with adenocarcinoma. *Int. J. Cancer* 95 (1), 12–17. doi:10.1002/1097-0215(20010120)95:1<12::AID-IJC1002>3.0.CO;2-E
- Pomin, V. H. (2015). Keratan sulfate: an up-to-date review. *Int. J. Biol. Macromol.* 72, 282–289. doi:10.1016/j.ijbiomac.2014.08.029
- Pomin, V. H. (2016). Paradigms in the structural biology of the mitogenic ternary complex FGF:FGFR:heparin. *Biochimie* 127, 214–226. doi:10.1016/j.biochi.2016.05.017
- Proudfoot, A. E., Handel, T. M., Johnson, Z., Lau, E. K., LiWang, P., Clark-Lewis, I., et al. (2003). Glycosaminoglycan binding and oligomerization are essential for the *in vivo* activity of certain chemokines. *Proc. Natl. Acad. Sci. U.S.A.* 100 (4), 1885–1890. doi:10.1073/pnas.0334864100
- Purushothaman, A., Sugahara, K., and Faissner, A. (2012). Chondroitin sulfate “wobble motifs” modulate maintenance and differentiation of neural stem cells and their progeny. *J. Biol. Chem.* 287 (5), 2935–2942. doi:10.1074/jbc.R111.298430
- Qiu, H., Jin, L., Chen, J., Shi, M., Shi, F., Wang, M., et al. (2020). Comprehensive glycomic analysis reveals that human serum albumin glycation specifically affects the pharmacokinetics and efficacy of different anticoagulant drugs in diabetes. *Diabetes* 69 (4), 760–770. doi:10.2337/db19-0738
- Raghuraman, A., Mosier, P. D., and Desai, U. R. (2010). Understanding dermatan sulfate-heparin cofactor II interaction through virtual library screening. *ACS Med. Chem. Lett.* 1 (6), 281–285. doi:10.1021/ml100048y
- Rein, C. M., Desai, U. R., and Church, F. C. (2011). Serpin-glycosaminoglycan interactions. *Methods Enzymol.* 501, 105–137. doi:10.1016/B978-0-12-385950-1.00007-9
- Rek, A., Krenn, E., and Kungl, A. J. (2009). Therapeutically targeting protein-glycan interactions. *Br. J. Pharmacol.* 157 (5), 686–694. doi:10.1111/j.1476-5381.2009.00226.x
- Rosenberg, R. D., and Damas, P. S. (1973). The purification and mechanism of action of human antithrombin-heparin cofactor. *J. Biol. Chem.* 248 (18), 6490–6505. doi:10.1016/S0021-9258(19)43472-8



- Rostand, K. S., and Esko, J. D. (1997). Microbial adherence to and invasion through proteoglycans. *Infect. Immun.* 65 (1), 1–8. doi:10.1128/IAI.65.1.1-8.1997
- Saito, T., Sugiyama, K., Hama, S., Yamasaki, F., Takayasu, T., Nosaka, R., et al. (2017). High expression of glypican-1 predicts dissemination and poor prognosis in glioblastomas. *World Neurosurg.* 105, 282–288. doi:10.1016/j.wneu.2017.05.165
- Salbach, J., Rachner, T. D., Rauner, M., Hempel, U., Anderegg, U., Franz, S., et al. (2012). Regenerative potential of glycosaminoglycans for skin and bone. *J. Mol. Med.* 90 (6), 625–635. doi:10.1007/s00109-011-0843-2
- Sandoval, D. R., Gomez Toledo, A., Painter, C. D., Tota, E. M., Sheikh, M. O., West, A. M. V., et al. (2020). Proteomics-based screening of the endothelial heparan sulfate interactome reveals that C-type lectin 14a (CLEC14A) is a heparin-binding protein. *J. Biol. Chem.* 295 (9), 2804–2821. doi:10.1074/jbc.RA119.011639
- Sankaranarayanan, N. V., Nagarajan, B., and Desai, U. R. (2018). So you think computational approaches to understanding glycosaminoglycan-protein interactions are too dry and too rigid? Think again! *Curr. Opin. Struct. Biol.* 50, 91–100. doi:10.1016/j.sbi.2017.12.004
- Sankaranarayanan, N. V., Strebel, T. R., Boothello, R. S., Sheerin, K., Raghuraman, A., and Sallas, F. (2017). A hexasaccharide containing rare 2-O-Sulfate-Glucuronic acid residues selectively activates heparin cofactor II. *Angew. Chem. Int. Ed. Engl.* 56 (9), 2312–2317. doi:10.1002/anie.201609541
- Schlachter, S., Seshu, J., Lin, T., Norris, S., and Parveena, N. (2018). The Borrelia burgdorferi glycosaminoglycan binding protein bgp in the B31 strain is not essential for infectivity despite facilitating adherence and tissue colonization. *Infect. Immun.* doi:10.1128/IAI.00667-17
- Schlessinger, J., Plotnikov, A. N., Ibrahim, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., et al. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* 6 (3), 743–750. doi:10.1016/s1097-2765(00)00073-3
- Schmidtke, M., Karger, A., Meerbach, A., Egerer, R., Stelzner, A., and Makarov, V. (2003). Binding of a N,N'-bis(tert-butyl) derivative of dispirotrypiperazine to heparan sulfate residues on the cell surface specifically prevents infection of viruses from different families. *Virology* 311 (1), 134–143. doi:10.1016/s0042-6822(03)00166-1
- Schwalter, R. M., Pastrana, D. V., and Buck, C. B. (2011). Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry. *PLoS Pathog.* 7 (7), e1002161. doi:10.1371/journal.ppat.1002161
- Schrevel, M., Osse, E. M., Prins, F. A., Trimbos, J., Fleuren, G. J., Gorter, A., et al. (2017). Autocrine expression of the epidermal growth factor receptor ligand heparin-binding EGF-like growth factor in cervical cancer. *Int. J. Oncol.* 50 (6), 1947–1954. doi:10.3892/ijo.2017.3980
- Schultz, V., Sulflita, M., Liu, X., Zhang, X., Yu, Y., Li, L., et al. (2017). Heparan sulfate domains required for fibroblast growth factor 1 and 2 signaling through fibroblast growth factor receptor 1c. *J. Biol. Chem.* 292 (6), 2495–2509. doi:10.1074/jbc.M116.761585
- Sepuru, K. M., Nagarajan, B., Desai, U. R., and Rajarathnam, K. (2018). Structural basis, stoichiometry, and thermodynamics of binding of the chemokines KC and MIP2 to the glycosaminoglycan heparin. *J. Biol. Chem.* 293 (46), 17817–17828. doi:10.1074/jbc.RA118.004866
- Sharma, B., Handler, M., Eichstetter, I., Whitelock, J. M., Nugent, M. A., and Iozzo, R. V. (1998). Antisense targeting of perlecan blocks tumor growth and angiogenesis in vivo. *J. Clin. Invest.* 102 (8), 1599–1608. doi:10.1172/JCI3793
- Shi, J., Fan, C., Zhuang, Y., Sun, J., Hou, X., Chen, B., et al. (2019). Heparan sulfate proteoglycan promotes fibroblast growth factor-2 function for ischemic heart repair. *Biomater. Sci.* 7 (12), 5438–5450. doi:10.1039/c9bm01336a
- Shriver, Z., Capila, I., Venkataraman, G., and Sasisekharan, R. (2012). Heparin and heparan sulfate: analyzing structure and microheterogeneity. *Handb. Exp. Pharmacol.* 207, 159–176. doi:10.1007/978-3-642-23056-1\_8
- Shute, J. K., Puxeddu, E., and Calzetta, L. (2018). Therapeutic use of heparin and derivatives beyond anticoagulation in patients with bronchial asthma or COPD. *Curr. Opin. Pharmacol.* 40, 39–45. doi:10.1016/j.coph.2018.01.006
- Stopschinski, B. E., Holmes, B. B., Miller, G. M., Manon, V. A., Vaquer-Alicea, J., Prueitt, W. L., et al. (2018). Specific glycosaminoglycan chain length and sulfation patterns are required for cell uptake of tau versus alpha-synuclein and beta-amyloid aggregates. *J. Biol. Chem.* 293 (27), 10826–10840. doi:10.1074/jbc.RA117.000378
- Stringer, S. E., and Gallagher, J. T. (1997). Specific binding of the chemokine platelet factor 4 to heparan sulfate. *J. Biol. Chem.* 272 (33), 20508–20514. doi:10.1074/jbc.272.33.20508
- Sunseri, M., Ahuja, T., Wilcox, T., and Green, D. (2018). Acquired coagulopathy and hemorrhage secondary to subcutaneous heparin prophylaxis. *Case Rep. Hematol.* 2018, 9501863. doi:10.1155/2018/9501863
- Tai, L. M., Thomas, R., Marottoli, F. M., Koster, K. P., Kanekiyo, T., Morris, A. W., et al. (2016). The role of APOE in cerebrovascular dysfunction. *Acta Neuropathol.* 131 (5), 709–723. doi:10.1007/s00401-016-1547-z
- Tammi, R. H., Passi, A. G., Rilla, K., Karousou, E., Vigetti, D., Makkonen, K., et al. (2011). Transcriptional and post-translational regulation of hyaluronan synthesis. *FEBS J.* 278 (9), 1419–1428. doi:10.1111/j.1742-4658.2011.08070.x
- Tandon, R., Sharp, J. S., Zhang, F., Pomin, V. H., Ashpole, N. M., Mitra, D., et al. (2020). Effective inhibition of SARS-CoV-2 entry by heparin and enoxaparin derivatives. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.06.08.140236
- Taylor, K. R., and Gallo, R. L. (2006). Glycosaminoglycans and their proteoglycans: host-associated molecular patterns for initiation and modulation of inflammation. *FASEB J.* 20 (1), 9–22. doi:10.1096/fj.05-4682rev
- Taylor, S., and Folkman, J. (1982). Protamine is an inhibitor of angiogenesis. *Nature* 297 (5864), 307–312. doi:10.1038/297307a0
- ten Dam, G. B., van de Westerloo, E. M., Purushothaman, A., Stan, R. V., Bulten, J., Sweep, F. C., et al. (2007). Antibody GD3G7 selected against embryonic glycosaminoglycans defines chondroitin sulfate-E domains highly up-regulated in ovarian cancer and involved in vascular endothelial growth factor binding. *Am. J. Pathol.* 171 (4), 1324–1333. doi:10.2353/ajpath.2007.070111
- Tiwari, V., Tandon, R., Sankaranarayanan, N. V., Beer, J. C., Kohlmeier, E. K., Swanson-Mungerson, M., et al. (2020). Preferential recognition and antagonism of SARS-CoV-2 spike glycoprotein binding to 3-O-sulfated heparan sulfate. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.10.08.331751
- Tong, W., and Wang, G. (2018). How can native mass spectrometry contribute to characterization of biomacromolecular higher-order structure and interactions? *Methods* 144, 3–13. doi:10.1016/j.ymeth.2018.04.025
- Tsujioka, H., Yotsumoto, F., Hikita, S., Ueda, T., Kuroki, M., and Miyamoto, S. (2011). Targeting the heparin-binding epidermal growth factor-like growth factor in ovarian cancer therapy. *Curr. Opin. Obstet. Gynecol.* 23 (1), 24–30. doi:10.1097/GCO.0b013e3283409c91
- Vallet, S. D., Clerc, O., and Ricard-Blum, S. (2021). Glycosaminoglycan-protein interactions: the first draft of the glycosaminoglycan interactome. *J. Histochem. Cytochem.* 69 (2), 93–104. doi:10.1369/0022155420946403
- Viboud, C. I., and Bliska, J. B. (2005). Yersinia outer proteins: role in modulation of host cell signaling responses and pathogenesis. *Annu. Rev. Microbiol.* 59, 69–89. doi:10.1146/annurev.micro.59.030804.121320
- Walenga, J. M., Jeske, W. P., Samama, M. M., Frapaise, F. X., Bick, R. L., and Fareed, J. (2002). Fondaparinux: a synthetic heparin pentasaccharide as a new antithrombotic agent. *Expert Opin. Investig. Drugs* 11 (3), 397–407. doi:10.1517/13543784.11.3.397
- Warkentin, T. E. (2018). Heparin-induced thrombocytopenia-associated thrombosis: from arterial to venous to venous limb gangrene. *J. Thromb. Haemost.* 16 (11), 2128–2132. doi:10.1111/jth.14264
- Weiss, R. J., Esko, J. D., and Tor, Y. (2017). Targeting heparin and heparan sulfate protein interactions. *Org. Biomol. Chem.* 15 (27), 5656–5668. doi:10.1039/c7ob01058c
- Whitelock, J. M., and Iozzo, R. V. (2005). Heparan sulfate: a complex polymer charged with biological activity. *Chem. Rev.* 105 (7), 2745–2764. doi:10.1021/cr10213m
- Wight, T. N. (2018). A role for proteoglycans in vascular disease. *Matrix Biol.* 71–72, 396–420. doi:10.1016/j.matbio.2018.02.019
- Wight, T. N., and Merrilees, M. J. (2004). Proteoglycans in atherosclerosis and restenosis: key roles for versican. *Circ. Res.* 94 (9), 1158–1167. doi:10.1161/01.RES.0000126921.29919.51
- Yamada, S., and Sugahara, K. (2008). Potential therapeutic application of chondroitin sulfate/dermatan sulfate. *Curr. Drug Discov. Technol.* 5 (4), 289–301. doi:10.2174/157016308786733564
- Yang, B., Wu, Y. J., Zhu, M., Fan, S. B., Lin, J., Zhang, K., et al. (2012). Identification of cross-linked peptides from complex samples. *Nat. Methods* 9 (9), 904–906. doi:10.1038/nmeth.2099

- Yang, J., and Chi, L. (2017). Characterization of structural motifs for interactions between glycosaminoglycans and proteins. *Carbohydr. Res.* 452, 54–63. doi:10.1016/j.carres.2017.10.008
- Yang, Y., Du, Y., and Kaltashov, I. A. (2020). The utility of native MS for understanding the mechanism of action of repurposed therapeutics in COVID-19: heparin as a disruptor of the SARS-CoV-2 interaction with its host cell receptor. *Anal. Chem.* 92 (16), 10930–10934. doi:10.1021/acs.analchem.0c02449
- Yugandhar, K., Wang, T. Y., Wierbowski, S. D., Shayhidin, E. E., and Yu, H. (2020). Structure-based validation can drastically underestimate error rate in proteome-wide cross-linking mass spectrometry studies. *Nat. Methods* 17 (10), 985–988. doi:10.1038/s41592-020-0959-9
- Zhou, J., Lin, J., Leung, W. T., and Wang, L. (2020). A basic understanding of mucopolysaccharidosis: incidence, clinical features, diagnosis, and management. *Intractable Rare Dis. Res.* 9 (1), 1–9. doi:10.5582/irdr.2020.01011
- Zong, C., Venot, A., Li, X., Lu, W., Xiao, W., Wilkes, J. L., et al. (2017). Heparan sulfate microarray reveals that heparan sulfate-protein binding exhibits different ligand requirements. *J. Am. Chem. Soc.* 139 (28), 9534–9543. doi:10.1021/jacs.7b01399
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2021 Shi, Sheng and Chi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# NMR Characterization of the Interactions Between Glycosaminoglycans and Proteins

Changkai Bu and Lan Jin\*

National Glycoengineering Research Center, Shandong Key Laboratory of Carbohydrate Chemistry and Glycobiology, Shandong University, Qingdao, China

## OPEN ACCESS

### Edited by:

Fuming Zhang,  
Rensselaer Polytechnic Institute,  
United States

### Reviewed by:

Xiaojun Sun,  
University of Jinan, China  
Yasuteru Shigeta,  
University of Tsukuba, Japan

### \*Correspondence:

Lan Jin  
lanjin@sdu.edu.cn

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 28 December 2020

**Accepted:** 24 February 2021

**Published:** 16 March 2021

### Citation:

Bu C and Jin L (2021) NMR  
Characterization of the Interactions  
Between Glycosaminoglycans  
and Proteins.  
Front. Mol. Biosci. 8:646808.  
doi: 10.3389/fmolb.2021.646808

Glycosaminoglycans (GAGs) constitute a considerable fraction of the glycoconjugates found on cellular membranes and in the extracellular matrix of virtually all mammalian tissues. The essential role of GAG-protein interactions in the regulation of physiological processes has been recognized for decades. However, the underlying molecular basis of these interactions has only emerged since 1990s. The binding specificity of GAGs is encoded in their primary structures, but ultimately depends on how their functional groups are presented to a protein in the three-dimensional space. This review focuses on the application of NMR spectroscopy on the characterization of the GAG-protein interactions. Examples of interpretation of the complex mechanism and characterization of structural motifs involved in the GAG-protein interactions are given. Selected families of GAG-binding proteins investigated using NMR are also described.

**Keywords:** glycosaminoglycans, proteins, interaction, NMR, conformation

## INTRODUCTION

Glycosaminoglycans (GAGs) are linear acidic heteropolysaccharides that exist in all mammals and are formed by repeating disaccharide units composed of N-acetyl-hexosamine and hexuronic or hexose (Table 1; Vasconcelos and Pomin, 2017). GAGs can have different sulfation patterns with different charge densities and heterogeneous monosaccharide compositions (Uhl et al., 2020). In addition to HA, GAGs are synthesized from the Golgi apparatus in the form of proteoglycans (Sasarman et al., 2016). According to the disaccharide composition and sulfation pattern, GAGs can be divided into several groups, including heparin/heparan sulfate (HS), chondroitin sulfate (CS)/dermatan sulfate (DS), keratan sulfate (KS) and hyaluronic acid (HA) (Pomin and Mulloy, 2018). Heparin/HS is composed of repeating disaccharide units of glucosamine (GlcNAc) and glucuronic acid (GlcA) or iduronic acid (IdoA). The initial substrate is  $[\rightarrow 4)\text{-}\beta\text{-D-GlcA-(1}\rightarrow 4)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow ]_n$ . GlcNAc can be substituted by sulfate groups at the amide, 3 or/and 6 hydroxyl groups, and the persulfation can be written as GlcNS<sub>3</sub>S<sub>6</sub>S. GlcA can be converted into IdoA by C5 epimerase, and both can be modified by 2-O-sulfation (written as IdoA<sub>2</sub>S or GlcA<sub>2</sub>S). CS consists of repeating disaccharide units of glucuronic acid (GlcA) and galactosamine (GalNAc). The initial substrate is  $[\rightarrow 4)\text{-}\beta\text{-D-GlcA-(1}\rightarrow 3)\text{-}\beta\text{-D-GalNAc-(1}\rightarrow ]_n$ . CS can undergo sulfation modification similar to heparin except for N-sulfation. However, due to the difference in glycosidic linkage, 3-O-sulfation in heparin becomes 4-O-sulfation. DS is obtained by converting GlcA in CS by C5-epimerase into IdoA. KS consists of repeating disaccharide units of Gal and GlcNAc, both of which can be 6-O-sulfated (Pomin, 2015). HA is the only GAG that is not modified by sulfation

and is not synthesized as proteoglycans. It is composed of repeating disaccharide units of GlcA and GlcNAc. According to the monosaccharide composition and sulfation pattern, GAG disaccharides can have 408 possible compositions (Soares et al., 2017).

As an important component of the extracellular matrix (ECM), GAGs play important roles in the construction of biological systems and the transduction of biological signals (Theocharis et al., 2016). Signal transduction occurs mainly through the interaction between GAGs and proteins, and these interactions are critical to the biological activity of these proteins. GAGs participate in a variety of physiological processes, including binding, activating and fixing a variety of protein ligands, such as growth factors, cytokines, chemokines, lipoproteins, proteases and their inhibitors, and other ECM components (Dyer et al., 2017; Rider and Mulloy, 2017; Crijns et al., 2020). GAGs are also associated with many pathological processes, including degenerative neurological diseases (Alzheimer's disease), cardiovascular diseases (thrombosis and atherosclerosis) and cancer (Vigetti et al., 2016; Huynh et al., 2019; Morla, 2019). In the invasion of viruses, GAGs also play roles that cannot be ignored (such as in herpes simplex virus and COVID-19) (Liu et al., 2020). The interaction between GAGs and proteins occurs mainly through electrostatic forces. This puts forward requirements for amino acid sequences in proteins and meets some rules, such as the XBBXB and XBBBXXB heparin-binding sequences proposed by Cardin, where B is a basic amino acid and X is any amino acid (Cardin and Weintraub, 1989). However, long-term research has found that the interaction between GAGs and proteins is not simply determined by the primary structure sequence. A large number of studies have proven that hydrogen bonds and van der Waals forces sometimes even play roles far exceeding electrostatic forces in the interaction; a proper tertiary structure of the protein is also required (Rudd et al., 2017). This poses more serious and complex problems for studying the interactions between GAGs and proteins.

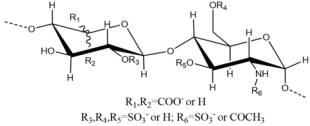
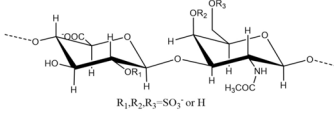
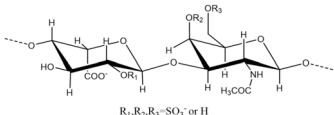
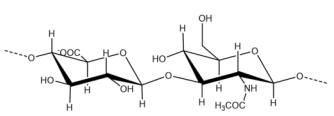
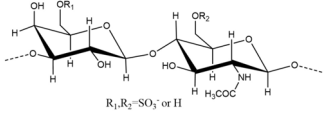
The interactions between GAGs and proteins are closely related to many factors, including saccharide unit composition, degree of sulfation, sulfation pattern, chain length, monosaccharide ring conformation and glycosidic linkage. The research methods used to characterize the interaction between GAGs and proteins mainly include gel electrophoresis (GE) (Nogueira et al., 2019), affinity chromatography (AC) (Sandoval et al., 2020), surface plasmon resonance (SPR) (Przybylski et al., 2020), biological layer interferometry (BLI) (Xiao et al., 2016), isothermal titration (ITC) (Zsila et al., 2018), microarray methods (Pomin and Wang, 2018b), crystal diffraction methods (X-ray) (Dahms et al., 2015), mass spectrometry (MS) (Yang and Chi, 2017), and nuclear magnetic resonance spectroscopy (NMR) (Kato and Peters, 2017). NMR is an insensitive technique compared with other analytical method for the study of interactions between biomolecules. The amount of sample needs to be in milligrams with high purity. In the study of proteins, NMR can characterize a protein with a molecular weight around 20 KD very well. However, proteins need to be isotope labeled by  $^{15}\text{N}$  and/or  $^{13}\text{C}$  when the molecular weight

increases and can be studied up to 100 KD. The cross peaks will become broadening and overlapped severely for larger proteins. Even with the above limitations, NMR is still an irreplaceable technique in the characterization of the biomolecule interactions at the atomic level especially in the case of glycosaminoglycans. Both X-ray diffraction and NMR can provide more precise tertiary structure information, and they do not require sample derivatization and will not cause structural damage to the sample during the experiment. Due to the accuracy and refinement of the data, both types of data can be used for model construction. However, X-ray diffraction studies a crystal in solid state and provide only few conformations of the interaction. While, NMR studies a solution under physiological condition and records dynamic conformations during the whole interaction period. Glycosaminoglycans are very hard to obtain a crystal due to their high flexibilities and exchangeable conformations. The solution NMR can not only show the natural state of the complex, but also detect the change of the complex conformation on the ns-ms time scale (Pomin and Wang, 2018a). Compared with the immobilization study of crystal diffraction, solution NMR can also be used for the dynamic study of interactions under physiological conditions.

Nuclear magnetic resonance is widely used to study the conformation of GAGs alone or in complex with proteins (Pomin, 2014), but the information usually obtained indicates that there are multiple GAGs or complex structures in solution. According to NMR data, GAGs present different folds configurations in solution according to their type and environment (Mulloy, 2006), such as the controversial 3-folds and 4-folds coexisting left-handed helix of HA (Gargiulo et al., 2010), which will directly affect the distribution of acidic groups in space. Generally speaking, the conformational changes of GAGs are mainly caused by two factors, one is the ring conformation of monosaccharides, and the second is the flexibility of the glycosidic linkages (Skidmore et al., 2009). The conformation of the IdoA residue in heparin, HS and DS is different from that of the other three monosaccharides (GlcNAc, GalNAc, and GlcA). IdoA exist in the conformational equilibrium, with two chairs ( $^1\text{C}_4$  and  $^4\text{C}_1$ ) and one shewboat ( $^2\text{S}_0$ ), instead of the fixed conformation  $^4\text{C}_1$  adopted in GalNAc, GlcNAc, or GlcA (Pomin, 2014). This gives these three different types of GAGs more flexible and various protein binding activities. This balance is affected by chain length, the degree of sulfation of adjacent monosaccharides, and its own 2-O sulfation (Haasnoot et al., 2020). When interacting with proteins, the conformational balance of IdoA will be tilted, such as binding to fibroblast growth factor-2 (FGF2), fibroblast growth factor-2 receptor (FGF2R), and eosinophil cationic protein (Ecp) (Hricovini et al., 2002; García-Mayoral et al., 2013). In free state, when the conformational balance ratio is closer to the required binding state, the binding affinity is stronger (Hricovini et al., 2002). Conversely, when the required conformation of the bound state cannot be achieved, the activity may be completely lost. But even if the protein has a clear tendency to a certain conformation of IdoA, there will generally be a conformational balance. The binding of AT III to heparin requires an absolute  $^2\text{S}_0$  conformation, but according to the NMR



**TABLE 1 |** Structures and tissue distribution of glycosaminoglycans.

Glycosaminoglycans	Degree of sulfation per disaccharide unit	Molecular weight range	Tissue distribution
Heparin/Heparan sulfate (HS)  $R_1, R_2, R_3 = \text{COO}^- \text{ or } \text{H}; R_4, R_5, R_6 = \text{SO}_3^- \text{ or } \text{COCH}_3$	Heparin about 1.8~2.4 HS about 0.8~1.8	Heparin about 3~30 kDa HS about 10~100 kDa	Heparin in liver, lungs and skin; HS was widely distributed on the cell surface.
Chondroitin sulfate (CS)  $R_1, R_2, R_3 = \text{SO}_3^- \text{ or } \text{H}; R_4, R_5, R_6 = \text{COO}^- \text{ or } \text{H}$	0.1~1.3	5~50 kDa	cartilage, tendon, aorta, ligament
Dermatan sulfate (DS)  $R_1, R_2, R_3 = \text{SO}_3^- \text{ or } \text{H}; R_4, R_5, R_6 = \text{COO}^- \text{ or } \text{H}$	< 1	15~40 kDa	skin, blood vessels, heart valves
Hyaluronic acid (HA)  $R_1, R_2, R_3 = \text{SO}_3^- \text{ or } \text{H}; R_4, R_5, R_6 = \text{COO}^- \text{ or } \text{H}$	0	4~12000 kDa	synovial fluid, vitreous humour, ECM of loose connective tissue
Keratan sulfate (KS)  $R_1, R_2, R_3 = \text{SO}_3^- \text{ or } \text{H}; R_4, R_5, R_6 = \text{COO}^- \text{ or } \text{H}$	< 1	5~25 kDa	KS I in cornea; KS II in cartilage aggregated; KS III in brain tissue

structure information, there is negligible  $^1\text{C}_4$  conformation in the whole binding process (Guerrini et al., 2006). Even though IdoA brings more variable binding conformational selectivity, recent studies have shown that GlcA has a better effect on the overall conformation of GAGs (Whitmore et al., 2020). In order to adapt to the ECM environment, the angle of the glycosidic linkages is allowed to change to a certain extent. The angle of the glycosidic linkages is affected by temperature, and the increase in temperature will result in a transition to the higher energy state (Hughes et al., 2017). When interacting with proteins, the glycosidic linkages can adopt proper orientations to meet the structural requirements during binding to proteins, and even cause the kinking of the GAGs polymer chain, thereby further enhancing the binding affinity (Hricovini, 2004). Compared with the obvious conformational equilibrium of IdoA, sometimes GAGs have  $\alpha/\beta$  isomeric equilibrium at the reducing end (Silipo et al., 2008) and rapid intramolecular hydrogen bond exchange (Almond et al., 1998). Due to the flexibility of GAGs, there may be multiple interaction modes at the same binding domain in the GAG-protein interaction process (Tjong et al., 2007). In the interaction between GAGs and proteins, the structure of the proteins is normally changed or stabilized. The weak interaction between GAGs and proteins undergoes on the ns-ms time scale, so the conformation of the protein in the system will change over time. Due to the structural heterogeneity and conformational flexibility of GAGs or the dynamic changes of the complex, it is also

very difficult to construct a model of complexes in solution (Almond, 2018).

Solution NMR can provide information about conformational changes and kinetic data during interactions between proteins and GAGs (Pomin and Wang, 2018a). NMR can also reveal the effects of different temperatures, pH values, salt concentrations, and ligand concentrations on the binding activity. There are three main goals in using NMR to study GAG-protein interactions: the first is to detect the amino acids involved in binding from the perspective of proteins, the second is to analyze the saccharide and its groups involved in binding from the perspective of GAGs, and the third is to observe the conformational changes and kinetic information during binding from the perspective of the interaction. To achieve these three goals, three technologies, chemical shift perturbation (CSP), saturation transfer difference (STD), and exchange-transferred nuclear Overhauser effect (trNOE), are initially used (Vignovich and Pomin, 2020), while other technologies, such as saturation transfer double difference (STDD) (Ledwith et al., 2016), paramagnetic relaxation enhancement (PRE) (Orton et al., 2016), pseudocontact shifts (PCS) (Srb et al., 2019), and exchange-transferred rotating-frame Overhauser effect (ROE), have been developed to compensate for the shortcomings of the former. The latest pulse sequences have been developed to provide a more detailed and accurate description of the binding process, such as the gradient spectroscopic observation of water ligands (waterLOGSY) (Huang and Leung, 2019) and

heteronuclear in-phase single quantum coherence experiment (HISQC) (Sepuru et al., 2018a). In addition, solid-state NMR has also been applied to study interactions involving ligands with low solubility (Malmos et al., 2016; Stewart et al., 2016). These techniques are based on four types of data: nuclear Overhauser effect (NOE), scalar coupling ( $J$ ), residual dipole coupling (RDC) and chemical shift anisotropy (CSA). The purpose of this paper is to introduce some important findings of the application of NMR to the study of the interactions between GAGs and proteins (Table 2) and the review is classified according to the type of GAGs.

## HEPARIN/HEPARAN SULFATE

Heparin is the most negatively charged polymer found in nature, and it is also the most studied in the GAG family (Conrad, 1997). One way to distinguish between heparin and HS is based on whether the mature body is still connected to the core protein. HS will be secreted out of the cell in the form of glycoproteins, most of which are fixed on the cell membrane to mediate many intercellular signaling pathways. Heparin is cleaved by  $\beta$ -endoglucuronidase and is combined with alkaline protease in the form of oligosaccharide chains to be stored in secretory granules (Oduah et al., 2016). The binding of heparin to protein mostly relies on its own high electronegativity and the positively charged domains in the protein. Hydrogen bonds and van der Waals forces also play important roles in the binding process. Moreover, the binding of heparin and protein is sometimes ion-dependent. For example, the binding of Langerin and heparin is mainly  $\text{Ca}^{2+}$ -dependent, although there are additional non- $\text{Ca}^{2+}$ -dependent binding sites (Muñoz-García et al., 2015; Hanske et al., 2017; José García-Jiménez et al., 2019). HS can be divided into a high-sulfation domain (NS domain) and a low-sulfation domain (NA domain). Heparin essentially contains all possible sulfation modification structures of the NS domain due to the degree of high sulfation. Most of the biological functions of HS are concentrated in the NS domain, although the NA domain is more flexible and more suitable for bending. Due to the early large-scale clinical application of heparin, it was relatively easy to obtain. Early research mainly used heparin as a substitute for HS to carry out functional and structural studies. In approximately the past thirty years, the study of the interaction between heparin and various proteins has become a hot spot, and the gradual maturity of chemical enzyme synthesis has given this field new vitality. Heparin can induce the oligomerization or heteromerization of proteins, which can prevent proteins from being hydrolyzed by protein-degrading enzymes and increase or decrease the possibility of their binding to receptors.

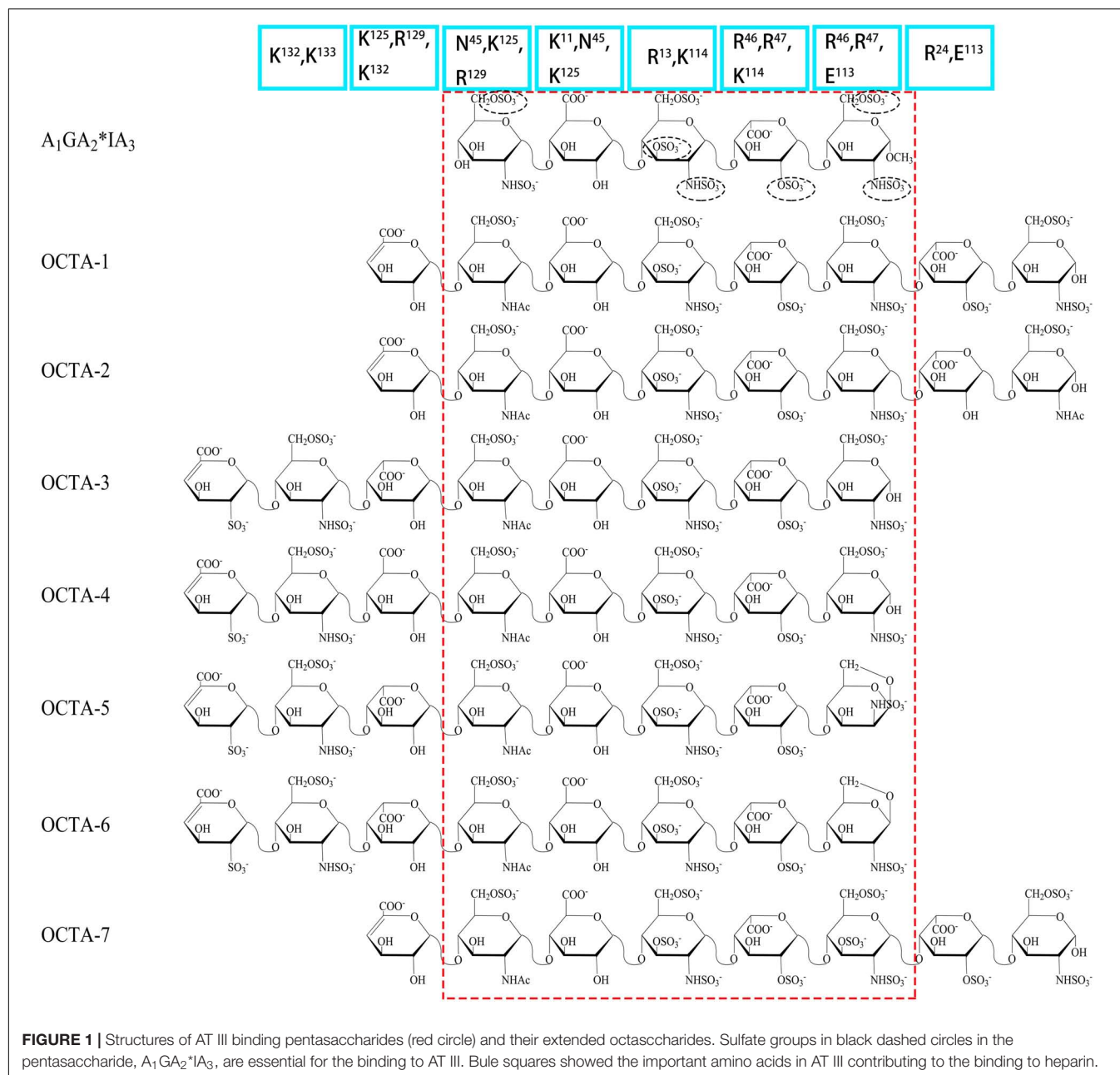
Antithrombin III (AT III) is an absolutely conserved serine protease with two different glycosylation forms ( $\alpha$ ,  $\beta$ ), consisting of three  $\beta$ -sheets (A-C) and nine  $\alpha$ -helices (A-I) (Rezaie and Giri, 2020). Heparin is a cofactor of the antithrombin-mediated coagulation cascade, and the interaction between them directly affects the activities of factors IXa, Xa and IIa (Gray et al., 2012). Choay, J used chemical enzymatic synthesis of various heparin-related oligosaccharides to determine that the minimum specific

sequence required for binding to AT III was the pentasaccharide  $\text{A}_1\text{GA}_2^*\text{IA}_3$  (Figure 1), which is also the only specific recognition sequence for heparin and protein binding found thus far (Thunberg et al., 1982; Choay et al., 1983). Although the specific pentasaccharide can meet the requirement of binding to AT III, it can only inhibit the activity of Xa. Inhibiting thrombin activity requires a heparin chain containing more than 16 saccharides, which can form a ternary complex with antithrombin and thrombin (Lane et al., 1984). The interaction between heparin and AT III was described as a three-state, two-step kinetic process (Figure 2; Olson et al., 1981), which assumed that AT III was in a balance of 'native unactivated,' 'intermediate-activated' and 'fully activated' states under physiological conditions (Roth et al., 2015). First,  $\text{A}_1\text{GA}_2^*$  was driven by  $\text{K}^{125}$  and  $\text{K}^{114}$  to combine with the C-terminus of helix D in "native unactivated" AT III, and the reducing end faced the N-terminus (Desai et al., 1998). Then, accompanied by conformational changes in AT III (helix D extension, reactive center loop exposure, and closure of sheet A) and heparin (IdoA from equilibrium conformation between  $^1\text{C}_4$  and  $^2\text{S}_0$  to complete  $^2\text{S}_0$ ), each unit in the pentasaccharide was further combined with AT III (van Boeckel et al., 1994). The combined complex can interact with the target protease or enzymatically decompose, and heparin is dissociated accordingly. In the electrostatic binding of heparin and AT III, several sulfate groups of heparin-specific pentasaccharide (N- $\text{SO}_3$  for  $\text{A}_2^*$  and  $\text{A}_3$ , 6-O- $\text{SO}_3$  for  $\text{A}_1$ , and 3-O- $\text{SO}_3$  for  $\text{A}_2^*$ ) and carboxyl groups were irreplaceable (Olson et al., 2002).

Further research using NMR focused on the specific role of each monosaccharide in the binding of heparin to AT III and the effect of extended pentasaccharide on the binding. The ratio of the  $^2\text{S}_0$  conformation in IdoA in the  $\text{A}_1\text{GA}_2^*\text{IA}_3$  sequence was 20% higher than that in the general heparin sequence (Ferro et al., 1987). In the three different chemically synthesized heparin pentasaccharides, the pentasaccharide had anticoagulant activity only when IdoA was in  $^2\text{S}_0$  (Das et al., 2001). Therefore, the proportion of  $^2\text{S}_0$  of IdoA in the heparin pentasaccharide sequence was one of the factors affecting the binding rate, which was affected by the degree of sulfation of glucosamine on both sides and its own 2-O- $\text{SO}_3$  (Haasnoot et al., 2020). Although the absence of 2-O- $\text{SO}_3$  in IdoA had no significant effect on the binding conformation, it resulted in a decrease in the proportion of free state  $^2\text{S}_0$  and a two-fold decrease in affinity (Stancanelli et al., 2018). At the same time, the flexibility of IdoA provided unlimited possibilities for the binding of heparin to protein. A recent study used IdoA2S instead of GlcA in the AT III binding sequence (Elli et al., 2020). The results showed that IdoA2S, which replaced GlcA, was in a pure  $^1\text{C}_4$  conformation when bound, and the affinity was tripled, which provided a basis for the application of bovine heparin. The unique structure of bovine heparin also provided unique ideas for the study of the specific mechanism of anticoagulation between heparin and AT III (Naggi et al., 2016). The 3-O- $\text{SO}_3$  and 6-O- $\text{SO}_3$  also had significant effects on the conformational balance of IdoA (Muñoz-García et al., 2012; Guerrini et al., 2013). The contribution of  $\text{A}_2^*$ 's 3-O- $\text{SO}_3$  to binding was in not only the conformation of heparin but also the formation of 'intermediate-activated' AT III (Lindahl et al., 1980;

**TABLE 2 |** GAG binding proteins.

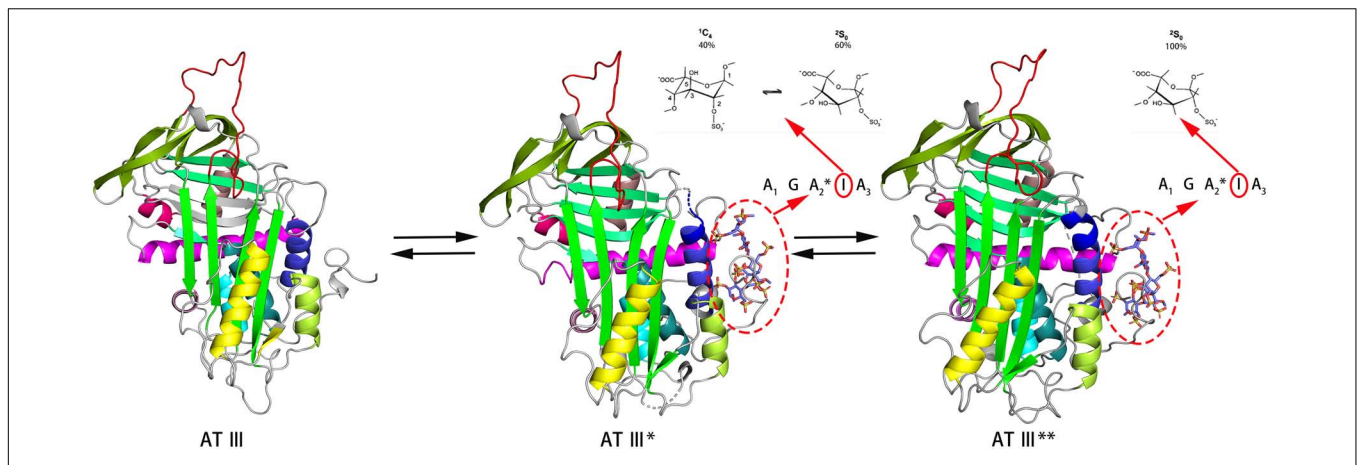
Type of protein	Name of protein	Type of GAG	participating binding residues and Secondary structure	Affinity ( $K_d$ )	References
Chemokine	CCL5	Heparin	40S loop (R <sup>44</sup> KNR <sup>47</sup> ), $\alpha$ helix (K <sup>55</sup> , K <sup>56</sup> )	18 $\mu$ M	Wang et al., 2011
		CS	40S loop(R <sup>44</sup> KNR <sup>47</sup> ), N loop (R <sup>17</sup> , L <sup>19</sup> , I <sup>15</sup> )	0.25 $\mu$ M	Deshauer et al., 2015
	CXCL1	Heparin/HS	N terminus(R <sup>8</sup> ), N-loop (H <sup>19</sup> , K <sup>21</sup> ), 40S turn (K <sup>45</sup> , R <sup>48</sup> ), $\beta$ 3-strand (R <sup>49</sup> ), C-helix (K <sup>60</sup> , K <sup>61</sup> , K <sup>65</sup> )	50 $\mu$ M	Sepuru and Rajarathnam, 2016; Sepuru et al., 2018b; Sepuru and Rajarathnam, 2019
		CS/DS	N terminus(R <sup>8</sup> ), N-loop (H <sup>19</sup> , K <sup>21</sup> ), 40S turn (R <sup>48</sup> ),	4 $\mu$ M	Sepuru and Rajarathnam, 2019
	CXCL2	Heparin	N-loop (R <sup>17</sup> , K <sup>21</sup> ), 40S turn (K <sup>45</sup> ), C-helix (K <sup>61</sup> , K <sup>65</sup> , K <sup>69</sup> )	25 $\mu$ M	Sepuru et al., 2018b
	CXCL5	Heparin/HS	N-loop (H <sup>23</sup> , K <sup>25</sup> ), 40S turn (K <sup>49</sup> ), $\beta$ 3 strand (K <sup>52</sup> ), C-helix (Lys <sup>64</sup> , Lys <sup>65</sup> , Lys <sup>69</sup> , Lys <sup>76</sup> )	30 $\mu$ M	Sepuru et al., 2016; Sepuru and Rajarathnam, 2019
		CS/DS	N-loop (H <sup>23</sup> , K <sup>25</sup> ), 40S turn (K <sup>49</sup> ), $\beta$ 3 strand (K <sup>52</sup> )	3 $\mu$ M	Sepuru and Rajarathnam, 2019
	CXCL7	Heparin	N-loop (H <sup>15</sup> , K <sup>17</sup> ), $\beta$ 3-strand(R <sup>44</sup> ), C-helix (R <sup>54</sup> , K <sup>57</sup> , K <sup>61</sup> )	—	Brown et al., 2017
	CXCL8	Heparin	N-loop (K <sup>15</sup> , H <sup>18</sup> , K <sup>20</sup> , K <sup>23</sup> ), C -helix (R <sup>60</sup> , K <sup>64</sup> , R <sup>68</sup> ), $\beta$ 3-strand (R <sup>47</sup> ), 50S loop (K <sup>54</sup> )	$\mu$ M	Joseph et al., 2015
		Heparin	C- helix (K <sup>57</sup> SKQAR <sup>62</sup> )	—	Severin et al., 2010
	CXCL12	Heparin	C-helix (R <sup>12</sup> , A <sup>40</sup> ), 20S loop (K <sup>24</sup> ), 40S loop (N <sup>46</sup> )	$\mu$ M	Laguri et al., 2011
	CXCL13	HS	C-helix (K <sup>60</sup> , R <sup>64</sup> , R <sup>67</sup> , H <sup>68</sup> ), C-loop (K <sup>84</sup> , R <sup>85</sup> , R <sup>86</sup> )	19 nM	Monneau et al., 2017
	CXCL14	Heparin	10S loop(I <sup>12</sup> ), $\beta$ 2-strand (I <sup>36</sup> , T <sup>37</sup> ), 40S loop(K <sup>54</sup> ), C -helix (R <sup>72</sup> ),	—	Penk et al., 2019
		CS/DS	10S loop(I <sup>12</sup> ), 40S loop(K <sup>54</sup> ), C -helix (R <sup>72</sup> )	—	Penk et al., 2019
Growth factor	FGF1	Heparin	$\beta$ 1– $\beta$ 2 loop (N <sup>18</sup> ), $\beta$ 8– $\beta$ 9 loop (N <sup>92</sup> ), $\beta$ 10– $\beta$ 11 loop (K <sup>113</sup> ), $\beta$ 11 strand (K <sup>118</sup> ), $\beta$ 11– $\beta$ 12 loop (Q <sup>127</sup> , K <sup>128</sup> )	nM	Ogura et al., 1999
	FGF2	Heparin	$\beta$ 1 strand (K <sup>27</sup> ), $\beta$ 1– $\beta$ 2 loop (N <sup>28</sup> ), $\beta$ 8– $\beta$ 9 loop (N <sup>102</sup> ), $\beta$ 10– $\beta$ 11 loop (R <sup>121</sup> ), $\beta$ 11 strand (K <sup>126</sup> ), $\beta$ 11– $\beta$ 12 loop (Q <sup>135</sup> , K <sup>136</sup> )	nM	Faham et al., 1996
	FGF7	Heparin	$\beta$ 3 strand (R <sup>18</sup> ), 40s loop(N <sup>92</sup> ), $\beta$ 10(N <sup>114</sup> ), 110s loop(Q <sup>115</sup> ), 120s loop (V <sup>120</sup> , K <sup>124</sup> , Q <sup>129</sup> , K <sup>130</sup> , T <sup>131</sup> )	—	Ye et al., 2001
Serpin	AT III	Heparin	N-terminal end (K <sup>11</sup> , R <sup>13</sup> ), A helix (R <sup>46</sup> , R <sup>47</sup> ), D helix (K <sup>114</sup> , K <sup>125</sup> , R <sup>129</sup> , R <sup>132</sup> , K <sup>133</sup> , K <sup>136</sup> )	20 nM	Olson et al., 2002
Type II cytokines	IL-10	Heparin/ CS/DS	D helix (K <sup>99</sup> , R <sup>102</sup> , R <sup>104</sup> , R <sup>106</sup> ), 110S loop (R <sup>107</sup> , K <sup>117</sup> , K <sup>119</sup> )	0.41 mM	Künze et al., 2016
	IFN $\gamma$	Heparin	C-terminal end (D1:K <sup>125</sup> TGKRKR <sup>131</sup> , D2:R <sup>137</sup> GRR <sup>140</sup> )	1.63 nM	Saesen et al., 2013
	Roundabout 1	HS	80s loop (K81), 130s loop (V <sup>133</sup> , H <sup>134</sup> , G <sup>135</sup> , R <sup>136</sup> , K <sup>137</sup> ), $\beta$ A strand (I <sup>167</sup> , R <sup>169</sup> )	—	Gao et al., 2016
Cytokine	Pleiotrophin	CS	C-terminal TSR domain $\beta$ -sheet (K <sup>60</sup> , K <sup>61</sup> , K <sup>69</sup> , K <sup>91</sup> , K <sup>92</sup> , K <sup>84</sup> , K <sup>86</sup> , K <sup>107</sup> )	90 $\mu$ M	Ryan et al., 2016
Link protein	CD44	HA	$\beta$ 1 strand (K <sup>38</sup> ), 40s loop (R <sup>41</sup> , Y <sup>42</sup> ), 70s loop (R <sup>78</sup> , Y <sup>79</sup> ), 90s loop (N <sup>100</sup> , N <sup>101</sup> ), 150s loop (R <sup>150</sup> ), $\beta$ 9 strand (R <sup>154</sup> ), 160s loop(R <sup>162</sup> )	$\mu$ M	Banerji et al., 2007
	TSG-6	HA	10s loop (K <sup>11</sup> , Y <sup>12</sup> ), 40s loop (H <sup>45</sup> , C <sup>47</sup> ), $\beta$ 3 strand (A <sup>49</sup> ), $\beta$ 3 strand (Y <sup>59</sup> ), 60s loop (V <sup>62</sup> , K <sup>63</sup> ), 80s loop (Y <sup>78</sup> , R <sup>81</sup> )	$\mu$ M	Higman et al., 2014
Viral pathogen	viral CCL2	Heparin	10s loop(R <sup>18</sup> ), 40s loop (K <sup>45</sup> , R <sup>46</sup> , R <sup>48</sup> )	113 mM	Zhao and LiWang, 2010
Defensins	Human $\beta$ -defensin 2	Heparin/DS	20s loop(R <sup>22</sup> RYK <sup>25</sup> ), $\beta$ 3 strand (K <sup>39</sup> ), 40s loop(K <sup>40</sup> )	5 mM	Seo et al., 2010
RNase A	Eosinophil cationic protein	Heparin	$\alpha$ 1 helix (R <sup>7</sup> , Q <sup>14</sup> , H <sup>15</sup> ), $\beta$ 1 strand (Q <sup>40</sup> ), loop4(H <sup>64</sup> ), $\beta$ 6 strand(H <sup>128</sup> )	15 $\mu$ M	García-Mayoral et al., 2013



Casu et al., 1981). Octa-7 (**Figure 1**), an octasaccharide with extended reducing end, showed that adding an extra 3-O-SO<sub>3</sub> to the A<sub>3</sub> would increase the ratio of <sup>2</sup>S<sub>0</sub> in I by approximately 15%. The additional 3-O-SO<sub>3</sub> formed new ionic bonds with R<sup>46</sup> and R<sup>47</sup>. The extended disaccharide also had a certain contribution to the binding (by interacting with E<sup>113</sup> and R<sup>24</sup>), and the binding force of octasaccharide and AT III was 40% higher than that of the specific pentasaccharide sequence and AT III. In the binding state, I and extended nonreducing end IdoA2S was completely in <sup>2</sup>S<sub>0</sub>. In a similar structure (OCTA-1), due to the lack of 3-O-SO<sub>3</sub> in the reducing end of A<sub>3</sub>, the extended IdoA2S was completely in <sup>1</sup>C<sub>4</sub> when bound, resulting in a substantial decrease in affinity (Guerrini et al., 2013). When extended reducing end IdoA2S's

2-O-SO<sub>3</sub> was removed (OCTA-2), the affinity increased slightly (Guerrini et al., 2008). In addition, there was little interaction between the reducing end extended disaccharide and AT III. In the other two octasaccharides with GlcA or IdoA as the extended nonreducing end (OCTA-3, OCTA-4), there was a significant polarization of affinity. The affinity of octasaccharide with GlcA as the nonreducing end was one order of magnitude higher than that with IdoA, which was in pure <sup>2</sup>S<sub>0</sub>. In recent years, the appearance of low-molecular-weight heparin has become a research hotspot due to its unique fragments produced by cleavage or hydrolysis on anticoagulation. In Guerrini's study, the affinity of two octasaccharides (OCTA-5, OCTA-6) containing specific pentasaccharide sequences derived from enoxaparin





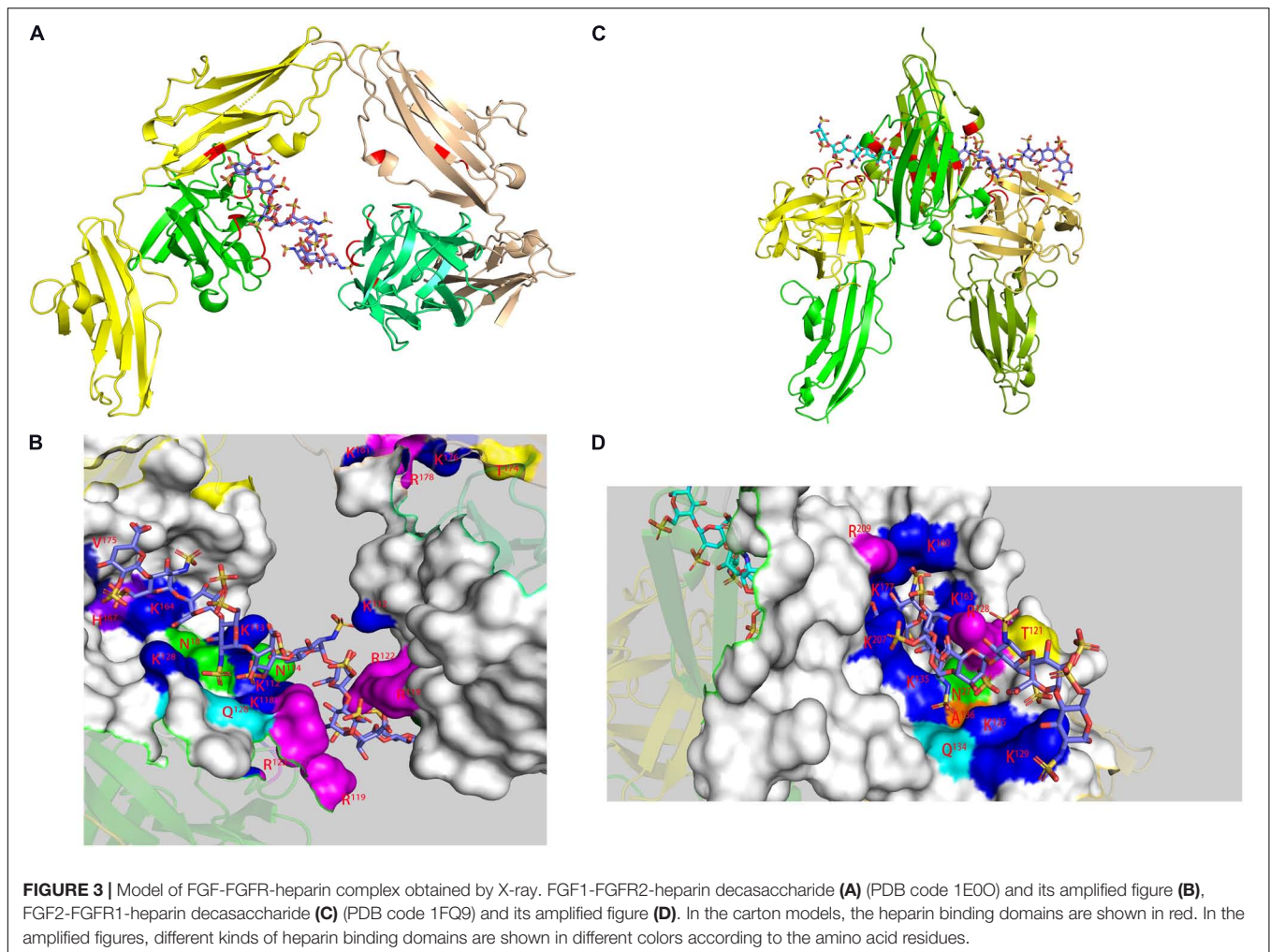
**FIGURE 2 |** Process of heparin binding to AT III. The binding of heparin with AT III is a reversible process. This process involves native unactivated (AT III, PDB code 1E05), intermediate-activated (AT III\*, PDB code 1NQ9) and fully activated (AT III\*\*, 1E03) states. During the binding process, IdoA transforms from conformational equilibrium to a complete  $^2S_0$  conformation (Jimenez-Barbero and Peters, 2003). The models of the three states are derived from X-ray. The reactive center loop (RCL) (red), sheet A (green), and helix D (gray blue) and the helix extension (dark blue) are highlighted in each state.

in binding with AT III decreased by 60-fold compared with the hexasaccharide with a complete pentasaccharide sequence. Because of the special pentasaccharide unit, the binding of the reducing end became weaker (Guerrini et al., 2010). The interaction difference of the octasaccharides with AT III showed that the substitution of different groups on heparin not only affected the binding strength with AT III but also changed the conformation during binding.

Heparin plays a key role in the regional aggregation and oligomerization of fibroblast growth factor (FGF), protecting it from denaturation and degradation and inducing its binding to the receptor (FGFR) (Korsensky and Ron, 2016). FGF is a growth factor family with 23 members, and its structure is highly related (12  $\beta$  strands form the classic  $\beta$ -trefoil structure) (Li et al., 2016). The receptor proteins of FGF include four categories (FGFR1-4), which are composed of three immunoglobulin (Ig)-like domains, which can be subdivided into seven categories according to the difference in Ig3 (Cheng et al., 2017). FGFR Ig2 is a key site for the binding of FGF and FGFR mediated by heparin (Kan et al., 1993). In the study of the effect of FGF and heparin, acidic fibroblast growth factor (aFGF, FGF1) and basic fibroblast growth factor (bFGF, FGF2) were the most classic models (Schlessinger et al., 2000). Studies have shown that the binding of heparin to FGF does not change the FGF conformation, and the binding domain is mainly located at the  $\beta$ 1-2 and  $\beta$ 10-11 strands (Canales-Mayordomo et al., 2006). Although there is clear evidence in the study of Crystallography, in the free state, 116-120 (131-136) of FGF1 (FGF2) constitute  $\beta$ XI structure (Zhu et al., 1991). However, Moy's NMR study on the structure of FGF2 in solution showed that there was no evidence to prove the existence of  $\beta$ XI (Moy et al., 1995). It is speculated that this is the structural change caused by the combination with HSPG, and this change is very important for the combination. This was confirmed in the subsequent NMR structural study of FGF1, Ogura pointed out that in the binding state, the 116-120 sequence has an obvious tendency

of  $\beta$ -chain structure (Ogura et al., 1999). In addition, K<sup>125</sup> in FGF2 and K<sup>118</sup> in FGF1 had high affinity in binding with heparin. Therefore, the  $\beta$ 11 chain was considered to be the key structure for the binding of FGF to heparin. In the combination of FGF2 and heparin, 2-O-SO<sub>3</sub> and N-SO<sub>3</sub> were necessary (Yu et al., 2014), and additional 6-O-SO<sub>3</sub> was required for FGF1 (Guerrini et al., 2002). However, in the study using 48 kinds of heparin disaccharides to bind FGF1, 3-O-SO<sub>3</sub> provided a stronger binding ability, and further C6 sulfation seemed to have a negative effect on the binding (Hu et al., 2012). In the study of the binding of heparin to FGF,  $^1C_4$  might have been the more favorable conformation (Canales et al., 2005; Guglieri et al., 2008). Interestingly, a recent study showed that specific AT-binding sequences can bind to FGFR2 Ig2 as a high-affinity complex, and IdoA remained in a high proportion of  $^2S_0$  (Nieto et al., 2011). Some experiments have shown that the combination of FGF and heparin seem to require a certain regular sequence of monosaccharide units or a special sulfation pattern (Ojeda et al., 2002). The mirror image of the carbohydrate structure also caused a significant reduction or loss of activity (Muñoz-García et al., 2013). For FGF1, only a single 6-sulfated tetrasaccharide was needed to induce its dimerization (Hricovíni et al., 2002). However, for FGF2 to be fully activated, heparin fragments of approximately decasaccharide might be required (Moy et al., 1997), although there was also evidence that tetrasaccharides could induce FGF2 dimerization (Guglieri et al., 2008). Heparin can induce FGF dimerization, but whether it is a critical step is controversial. Some NMR data showed that heparin, which formed a high-affinity complex with FGF, did not induce the dimerization of FGF but still had high activity (Canales et al., 2006).

In the study of the FGF-FGFR-heparin binding model (Figure 3), the crystal study gave two hypotheses: a 2:2:1 trans-binding model and a 2:2:2 cis-binding model (Pellegrini, 2001). NMR research in recent years has explained the formation process of the 2:2:2 model. Nieto used FGF1 and FGFR2 Ig2



and two heparin oligosaccharides to study the mechanism (Nieto et al., 2013). In the activity experiment, FGF1 and FGF2 had different requirements for heparin. In deheparinized cells, FGF2 activity was completely lost. However, after pretreatment of the cells with heparin, the activity recovered. FGF1 requires the presence of an additional heparin-like stabilizer myo-inositol hexasulfate (MIHS). It is speculated that the role of heparin in FGF1 was not limited to mediating the binding of FGF and FGFR. There was a second binding site in the FGF-FGFR complex, which was a clear cis-dimer binding model mark. Subsequent speculation suggested that the signaling pathway should be regarded as follows: FGFR dimerization was initially induced by GAGs, and then FGF and the ternary complex formed a higher-order aggregate and activated the subsequent enzyme cascade. Schieborr investigated the interactions among FGF1/FGF2, FGFR4 Ig2, and three different heparin polysaccharides (Saxena et al., 2010). The experimental results showed that the hexasaccharide could meet all the binding site requirements for inducing FGF dimerization, but the stability of the resulting complex was extremely poor. STD experiments showed that the combination of octasaccharide and FGF2 had a positive synergistic effect, but due to the

lack of heparin structure data, the exact mechanism needs further experimental verification. Heparin was proven to have an extremely low dimerization ability for inducing FGFR4 Ig2, which was clear proof of the trans-dimer model in the description by Pomin (2016). However, the NMR data suggested there was a secondary binding site in the FGF-FGF Ig2 complex, which was again a clear cis-dimer binding model. Schieborr proposed that hexasaccharides and octasaccharide could mediate FGF2 signaling pathways under different mechanisms, and the positive synergistic effect of octasaccharide was due to the different residues involved in the binding. However, while there should theoretically be an FGF/FGFR/heparin 4:2:2 complex in the pathway, there were no data to support its existence. The existence of the FGF/FGFR/heparin 2:2:1 model was clearly supported by Brown's ITC data, but no NMR evidence was obtained (Brown et al., 2013).

CXCL12 has six different splicing variants (CXCL12 $\alpha$ - $\varphi$ ) in humans and is the only CXC chemokine with differential gene splicing (Janssens et al., 2017b). The complex of CXCL12 and the receptor CXCR4 mediates many physiological functions, including physiological processes such as hematopoiesis, embryonic development, vascular repair, and inflammation

(Murphy and Heusinkveld, 2018). CD26, a leukocyte-activating antigen, can be cleaved CXCL12 between the N-terminal P<sup>2</sup> and V<sup>3</sup> residues (Janssens et al., 2017a). The cleaved product has a reduced affinity for CXCR4 and cannot activate it any more. Research on the binding domain of CXCL12 and heparin/HS can be traced back to 1999. The K<sup>24</sup>HLK<sup>27</sup> base sequence in the  $\beta$ 1-strand of the  $\beta$ -sheet, conforming to the BBXB rule, was verified in a mutation experiment (Amara et al., 1999). Sadir believed that R<sup>41</sup> and R<sup>43</sup> in the  $\beta$ 2 strand were additional binding sites, in addition to K<sup>1</sup> at the N-terminus as a potential binding site (Sadir et al., 2001). The binding between heparin/HS and K<sup>1</sup> in CXCL12 was believed to protect CXCL12 from being cleaved by CD26 (Sadir et al., 2004). Murphy first used X-ray crystallography to study the interaction between CXCL12 and heparin/HS and proposed two binding domains in CXCL12: one at the interface of the dimer and the other in the N-loop region and the N-terminal helix similar to the binding domain in CXCL8 (Murphy et al., 2007). Using <sup>13</sup>C-labeled octasaccharides in the NMR experiment, Laguri determined that the heparin-binding sequence was related to the GlcN-3, GlcA-4, and GlcN-5 units of the octasaccharides (Laguri et al., 2011). N-sulfation and 6-O-sulfation are essential for binding. The nonreducing end monosaccharide and reducing end disaccharide of the octasaccharide formed additional contact with the N-terminus of CXCL12 (R<sup>8</sup> and R<sup>12</sup> are the most prominent), and a consistent molecular binding model was constructed. However, Ziarek proposed a controversial molecular model (Ziarek et al., 2013). He believed that heparin and two CXCL12 molecules should drive the formation of the polymer in an almost orthogonal conformation, instead of the previously proposed interface of two CXCL12 molecules (composed of a  $\beta$ 1 strand and the N-terminus). The data indicated that the binding site in CXCL12 should be on the six-strand of the  $\beta$ -sheet, while the N-terminus was not involved. The main residues involved in binding included K<sup>20</sup>, K<sup>24</sup>, K<sup>27</sup>, K<sup>41</sup>, K<sup>43</sup> and R<sup>47</sup>, while A<sup>8</sup> and A<sup>12</sup> provided additional binding. It was proposed that the reason why heparin protected CXCL12 from CD26 cleavage was not the preemptive combination but the coverage of K1 caused by dimerization. Panitz's study proved that the interaction affinity between heparin and CXCL12 was much higher than that of other GAGs, and the degree of sulfation was not the only factor influencing the binding (Panitz et al., 2016). The binding sites in CXCL12 with other GAGs were similar to heparin, with the exception of a second binding site for CS compared to heparin (R<sup>20</sup>, A<sup>21</sup>, N<sup>30</sup>, K<sup>64</sup>).

Type II cytokines have six secondary structure elements (A-F) to form an  $\alpha$ -helical structure, of which A, C, D, and F adopt the classic four-helix topology, while B and E exist as the connecting structure (Pestka et al., 2004). Interleukin-10 (IL-10), interferon- $\gamma$  (IFN $\gamma$ ) and interleukin-26 (IL-26) are the three proteins in this family that exist in the form of dimers. Although IL-10 and IFN $\gamma$  had the same protein folding mode, their binding with heparin split into two completely different manners. STD data indicated that when IL-10 bound to heparin, the degree of sulfation rather than the site had a greater impact on the binding (Künze et al., 2014), although the effect of 6-O-SO<sub>3</sub> on affinity was 2-3 times

greater than the effects of N-SO<sub>3</sub> and 2-O-SO<sub>3</sub>. Data showed that there was a hydrogen bond or strong van der Waals force between IL-10 and the methyl group in the N-acetyl residue of the saccharides. As the heparin chain length increases, the affinity increases. When the chain length reached eight sugars, the affinity suddenly increased. It was calculated using STD data that when IL-10 bound to a heparin oligosaccharide with more than eight sugars, the Hill coefficient was approximately 2. This indicated that heparin and each monomer of the IL-10 dimer were bound, and the binding was synergistically positive. It was speculated that the binding site in IL-10 was located at the C-terminus of the D helix and the basic amino acid cluster L<sup>101</sup>RLRLRRCHRF<sup>111</sup> of the adjacent DE loop. This heparin-binding domain existed in both monomers, which also supported the positive synergistic combination of octasaccharide and IL-10. NOE data showed that the conformation of a tetrasaccharide in the binding center did not change much. Further PCS data confirmed that the binding domain of IL-10 with heparin was in the 101-111 basic amino acid cluster (Gehrcke and Pisabarro, 2015). This domain is absolutely conserved in IL-10 from various sources, and it is also located in the binding domain of IL-10R2 and IL-10. The reason why GAG had an inhibitory effect on IL-10 might be due to the low-affinity IL-10R2 competing with heparin for binding.

Unlike IL-10, the binding domain of IFN $\gamma$  with heparin was located at the C-terminus. IFN $\gamma$  had four clusters of enriched basic amino acids, but only two C-terminal domains, K<sup>125</sup>-R<sup>131</sup> (D1) and R<sup>137</sup>-R<sup>140</sup> (D2), interacted with heparin (Vanhaverbeke et al., 2004). NOE data showed that the interaction between the protein and heparin had no effect on the conformation of the protein, and only the electrostatic force contributed to the binding without any other interaction force. The increase in sugar chain length increased not only the affinity between heparin and IFN $\gamma$  but also the bending degree of the whole sugar chain. The binding of IFN $\gamma$  to heparin protected the D1 domain from protease hydrolysis, and D1 acts as the main binding domain to heparin. ITC experiments have shown that D2 is not necessary for the binding of IFN $\gamma$  to heparin, but removing D2 will increase the binding of IFN $\gamma$  to heparin (Döbeli et al., 1988). Further studies have shown that the combination of D1 with heparin was mainly a thermodynamic process, while the combination of D2 with heparin was a kinetic process (Saesen et al., 2013). The main function of D2 was to strengthen the binding of IFN $\gamma$  with heparin. The binding of the C-terminus of IFN $\gamma$  to heparin is a two-step process. First, D1 bound to heparin, and the binding site was oriented. Then, D2 combined with heparin to strengthen the binding. The binding of IFN $\gamma$  to its receptor includes two domains, one of which is the C-terminus. Therefore, HSPG on the cell surface competed with the IFN $\gamma$  receptor for binding; and the addition of exogenous heparin could also reduce the IFN $\gamma$  concentration on the cell surface. The inhibitory effect of heparin on the activity of certain proteins might be due to its competition with the protein receptor for binding, which led to the decreased or even disappearance of the binding affinity between the receptor and the protein. IL-10 inhibits the activity of IFN $\gamma$ , so its mechanism might be more complicated. Studying the interaction between GAGs and proteins of a specific



sequence may help to develop a more thorough understanding of the mechanism.

## CHONDROITIN SULFATE

According to the type of uronic acid and sulfation, common CS can be divided into five categories: nonsulfated chondroitin sulfate (CS-O), 4-O-sulfated chondroitin sulfate (CS-A), 6-O-sulfated chondroitin sulfate (CS-C), 2, 4-O-disulfated chondroitin sulfate (CS-D), and 4,6-O-disulfated chondroitin sulfate (CS-E) (Yang et al., 2020). CS-B (DS) has all of the sulfation modification types of the above five types of CS, but its uronic acid is epimerized into IdoA. Oversulfated chondroitin sulfate (OSCS) was sulfated at all sites that could be sulfated, and it was one of the culprits that triggered the “heparin crisis” in 2008 (Zhu et al., 2019). There is a special kind of 3-O-sulfated chondroitin sulfate (CS-K) in marine organisms that has a high affinity for growth factors (Palhares et al., 2019).

In the interaction with chemokines, the main function of GAG was to locally aggregate chemokines to increase their binding to G-coupled protein receptors and to form a concentration gradient required for the migration of leukocytes, among which HS was dominant (Rajaratnam et al., 2018). However, CS also played an important role in the interaction with certain chemokines, such as the chemokine CCL5 (regulated upon activation of normal T cell expressed and secreted factor, RANTES). CS plays an important role in a variety of biological pathways mediated by CCL5, such as inducing T cell apoptosis and monocyte blockade. Deshauer studied the interaction between two CS hexasaccharides and CCL5 and used TEMPO to label CS for PRE experiments to study the binding sites in depth (Deshauer et al., 2015). In the titration of CCL5 with CS444 (GlcA-GalNAc4S- GlcA-GalNAc4S- GlcA-GalNAc4S), there were obvious chemical shift changes in the 40S loop, the N-terminus and the N loop (Figure 4). At a ratio of 1:1, the chemical shift had no significant change. When CS644 (GlcA-GalNAc6S- GlcA-GalNAc4S- GlcA-GalNAc4S) is used for titration, there are only small chemical shift disturbances at these three binding sites. However, when the ratio of CS644:CCL5 was more than 1:1, R<sup>17</sup> and L<sup>19</sup> in the N loop showed obvious chemical shift disturbances. In the PRE experiment, CS444 data showed that its reducing end was close to the 40S loop BBXB sequences. However, CS644 had additional chemical shift changes at Y<sup>3</sup>, A<sup>16</sup>, and R<sup>21</sup>, indicating that CS644 was also close to the 20S loop, N-loop and N-terminus, which suggested that the combination of CS644 and CCL5 was more heterogeneous. It can be seen that the type of GAG, the degree of sulfation and the ring conformation had a huge influence on the binding conformation between GAG and protein, which was also reflected in Pichert's CXCL8 and CS hexasaccharide interaction study (Pichert et al., 2012).

Midkine (MK) and pleiotropic protein (PTN) form the MK/PTN cytokine family, which is a heparin-binding nerve growth factor. They are highly similar in structure and share more than 50% of the amino acid sequence (Herradon et al., 2019). They consist of two TSR domains with a hinge connection. Each domain consists of three antiparallel  $\beta$ -strands to form

$\beta$ -sheets. The C-terminal domain (CTD) of PTN is the main CS-binding domain, which has an affinity far greater than that of the N-terminal domain (NTD) (Ryan et al., 2016). CTD has two basic residue clusters (cluster 1: K<sup>69</sup>, K<sup>91</sup>, K<sup>92</sup> and cluster 2: K<sup>84</sup>, K<sup>86</sup>, K<sup>107</sup>). The electrostatic potential diagram showed that the two sides of the  $\beta$ -sheet can be coplanar. According to the PRE data, CS-A preferred cluster 2, while CS-E preferred cluster 1. The data showed that K<sup>54</sup> in NTD was close to the paramagnetic center, but NTD had only a few residues with side chains and HN atom transfer perturbation. The hydrophobic hinge can arrange two lysines (K<sup>60</sup> and K<sup>61</sup>) near CTD cluster 1 to participate in the binding of CS. Although there was no clear reason to prove the effect of the C-terminus on the binding of CS to PTN, the affinity of CS-A, but not CS-E, to the C-terminal truncated PTN was greatly reduced. CS-E had a greater affinity than CS-A, which might be the reason why the PTN/MK family was associated with many tumorous inflammations (Weckbach et al., 2018). Unlike PTN, according to STD data, CS-E can simultaneously bind to the two domains of the midkine (Solera et al., 2016).

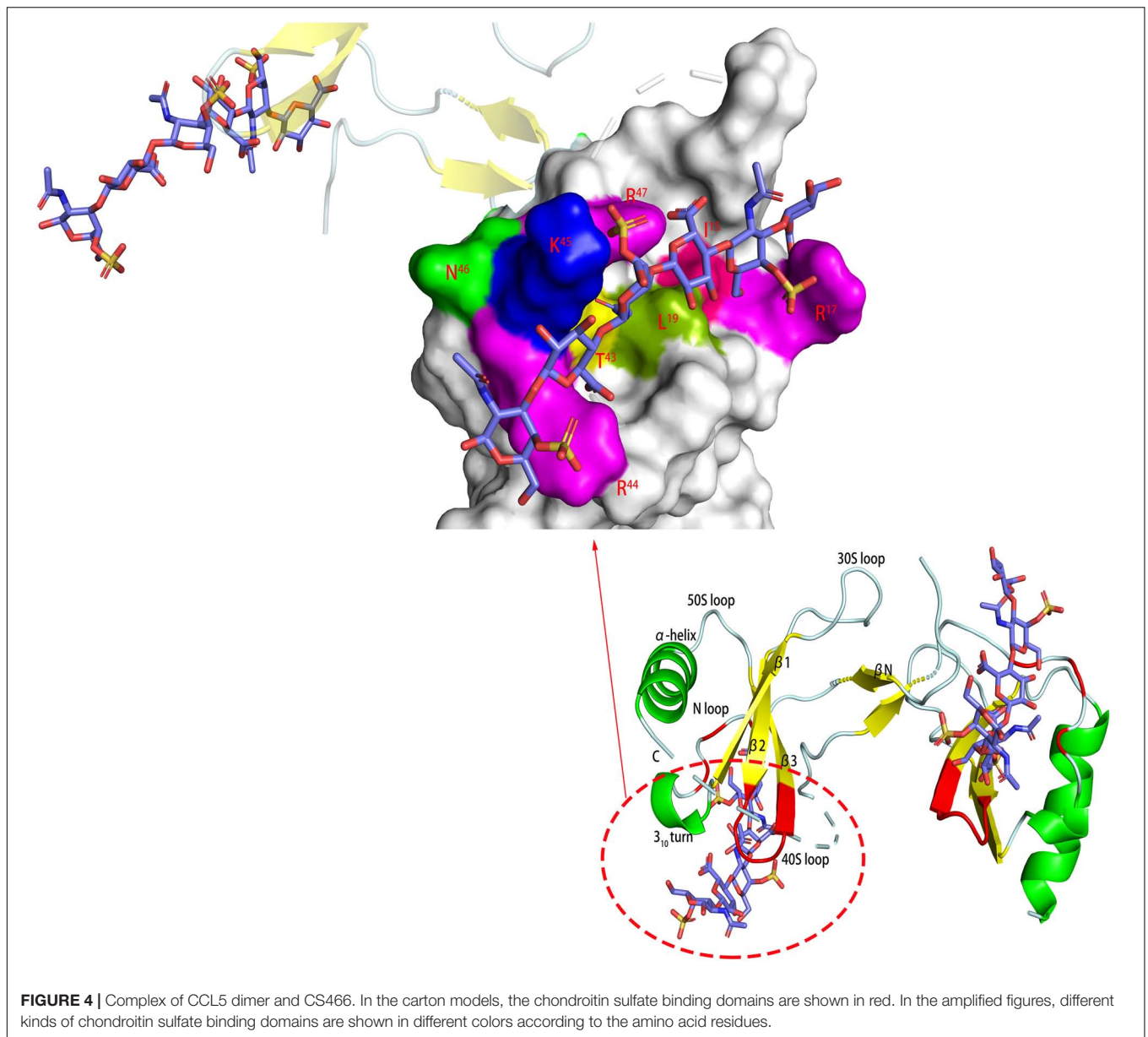
Tumor necrosis factor-stimulated gene-6 (TSG-6) is a classic HA-binding protein that shows different binding modes with CS compared to HA (Park et al., 2016). The combination of CS and Link- TSG-6 had at least two binding sites, and 4-O-sulfation was preferred. The slow exchange site was similar to the HA-binding site, but there were still some differences due to the sulfation pattern of CS. STD data indicated that there was a second group of rapid exchange binding sites, which were close to the heparin-binding site according to the model based on PRE data. The change in the relaxation rate ratio R<sub>2</sub>/R<sub>1</sub> indicated that the initial combination of CS and Link- TSG-6 can induce dimerization. The dimerization interface and the CS binding site were located on opposite sides, so CS plays a neutralizing role rather than functioning as a bridge in inducing dimerization.

## DERMATAN SULFATE

Although DS was similar in structure to CS, the existence of IdoA gave it unparalleled structural flexibility. For example, in combination with hepatocyte growth factor/scattering factor (HGF/SF), the presence or absence of IdoA was the key to the combination of GAG with HGF/SF (Deakin et al., 2009). The binding mode of DS and NK1 (HGF/SF heparin-binding domain) was similar to that of heparin, although the affinity was slightly lower. The binding was concentrated in the N domain. Although crystallographic data proved that the K1 domain was involved in binding, this binding was based on the premise of dimerization. However, the NMR data showed that in solution, the low-molecular-weight GAGs would not induce its dimerization.

Sepuru used medium-length GAG to study the interaction with CXCL1 or CXCL5 in the presence of monomers and dimers through CSP experiments (Sepuru and Rajaratnam, 2019). The two binding sites in CXCL1 with HS were on the opposite sides of the protein, the  $\alpha$ -domain (H<sup>19</sup>, K<sup>21</sup>, K<sup>45</sup>, K<sup>60</sup>, K<sup>61</sup>, K<sup>65</sup>) and the  $\beta$ -domain (R<sup>8</sup>, K<sup>29</sup>, R<sup>48</sup>, K<sup>49</sup>). The results showed that CXCL1 and HS were combined in a ratio of 1:2, and ITC experiments verified this result. The binding sites of CXCL1 with CS and DS





are located in the  $\gamma$ -domain ( $R^8$ ,  $H^{19}$ ,  $K^{21}$ ,  $K^{45}$ ,  $K^{49}$ ). The binding domain of CXCL5 with GAG was similar to that of CXCL1, but there was no obvious specificity for GAG species. Neither CXCL1 nor CXCL5 bound to GAG involved helices, which was different from the previous proposal that helices are an important binding site for the interaction of chemokines that activate CXCR2 with GAG. In the HADDOCK model, the interaction between DS and CXCL1 involved two sulfate groups, two carboxyl groups and two N-acetyl groups, and the interaction model with CXCL5 involved two sulfate groups, one N-acetyl and one hydroxyl group. The molecular docking models of CS and DS with different structures were quite different. They involved different residue-binding groups and positions. This was consistent with the differences in the interaction morphology of GAG with different structures proposed previously. This was also reflected in the combination

of CXCL14 and DS (Penk et al., 2019). The binding of DS and heparin with CXCL14 occurred in the C-terminal helix, part of the N-terminus and the transition between the second and third  $\beta$ -sheets ( $Y^{44}$ - $Q^{47}$ ). However, the maximum perturbation in the combination of DS and CXCL14 was associated with  $R^{72}$ , while  $I^{36}$  and  $T^{37}$  were more affected in terms of heparin. DS and CS also had significant differences in N-terminal disturbances. The interaction between DS and protein was also dependent on chain length and sulfation pattern. In the study of the interaction between tau protein and DS, tau was favored for 6-O-sulfation (Zhao et al., 2017). Disulfated DS had a higher affinity than monosulfated DS, although the affinity of both was less than that of heparin.

Decorin binding protein B (DBPB) bound to DS in a different binding mode than DBPA, mainly through the linker between

helices 1 and 2, the C-terminal tail, and the alkaline patch (Feng and Wang, 2015). In the PRE experiment, there were no clear data indicating that the C-terminal tail was involved in binding. It was speculated that this was because the binding occurs at the nonreducing end of DS, while the TEMPO label was at the reducing end of DS. The mutation data showed that the three sites all had a promoting effect on binding, and the C-terminus played a key role in binding. The most obvious difference between DBPB and DBPA was only the C-terminal disulfide bond, which again emphasized the influence of protein structure on binding. Due to the lack of disulfide bonds, the C-terminus could exist in multiple conformations when combined with DS, which was also thermodynamically favorable. Although the BXBB sequence in DBPA remained highly dynamic in DBPB, it did not contribute much to the binding due to the exposure of the C-terminus and the position of the linker in DBPB.

## HYALURONIC ACID

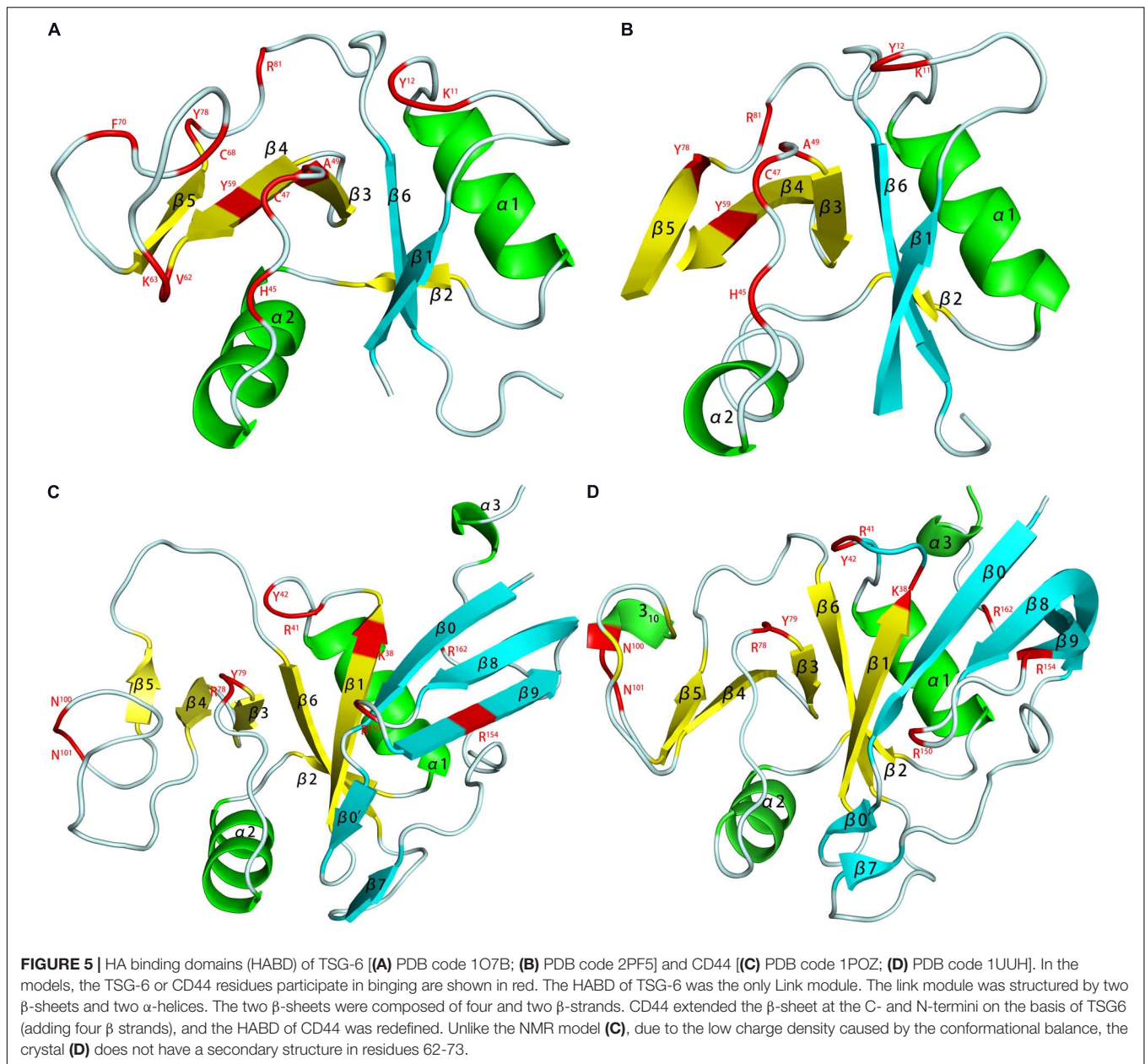
Hyaluronic acid has a different synthesis site (plasma membrane) and a different synthesis form (non-glycoprotein) compared to other GAGs. HA will not undergo further modification; thus, the interaction between it and the protein seems to be structurally specific. The hydrogen bonds and intramolecular hydrogen bonds with water molecules gave it a complex  $\beta$ -sheet structure (Taweechat et al., 2020). In the double helix structure of HA, every two monosaccharide flip  $180^\circ$ . HA, as a structural scaffold, widely exists in the epithelial tissue, connective tissue and nerve tissue of vertebrates and regulates the physical and chemical processes of tissue hydration and penetration. The interaction between HA and HA-binding protein (hyaluroadhesin) mediates various physiological activities, such as cell signal transduction, wound repair, tissue regeneration, leukocyte rolling adhesion and inflammation (Fallacara et al., 2018). Most HA-binding proteins belong to the link protein superfamily. Some other proteins (such as receptor for hyaluronan-mediated motility, RHAMM) and peptides (thymosin  $\alpha 1$ , T $\alpha 1$ ) bound to HA are independent of the link module (Naor, 2016).

The 14 human link proteins can be divided into three categories (A, B, C) according to their structural composition (Kohda et al., 1996). TSG-6 was the most typical type A Link protein, and its HA-binding domain (HABD) was the only Link module (Figure 5; Day and Milner, 2019). The link module was composed of 100 amino acids and structured by two  $\beta$ -sheets and two  $\alpha$ -helices, which were stabilized by two extremely conserved disulfide bonds. The two  $\beta$ -sheets were composed of four and two  $\beta$ -strands. Type B Link protein used CD44 as a template. It extended the  $\beta$ -sheet at the C- and N-termini on the basis of type A (adding four  $\beta$  strands), and the HABD of type B was redefined (Senbanjo and Chellaiah, 2017). The type C link protein was composed of two links in series, both of which participate in binding with HA. This subcategory included aggrecan, versican and HAPLN1-4, but detailed research on its structure is lacking. The binding of HA and protein had very strict requirements on the tertiary structure of the protein. This was most obvious in the type C Link protein, which did not interact with GAGs

other than HA. In one study, three link modules were connected in series, but the binding activity with HA was completely lost (Cai et al., 2004).

Kahmann proposed that the binding of Link-TSG-6 and HA was concentrated in the  $\beta 4/\beta 5$  loop. The association was accompanied by the rearrangement of C<sup>47</sup> and C<sup>68</sup> disulfide bonds (Kahmann et al., 2000). In the previously proposed B(X)<sub>7</sub>B rule motif (R<sup>5</sup>EARS<sup>6</sup>GKYK<sup>13</sup>), R<sup>5</sup> and K<sup>13</sup> had no obvious evidence of involvement in binding, but K<sup>11</sup> was the main binding residue. In Blundell's subsequent research, it was shown that the folding of the link module remains unchanged during the combination (Blundell et al., 2003). The largest structural change was found in  $\beta 4/\beta 5$ . K<sup>11</sup> also changed its orientation and became more oriented. For Y<sup>59</sup> and Y<sup>58</sup>, the benzene rings did not rotate due to ring stacking. Due to the derived polarity of the binding, the two ends of the binding were located at K<sup>11</sup> and R<sup>81</sup>. Higman proposed that in the free state, the  $\beta 4/\beta 5$  loop of TSG-6 was highly dynamic. In this state, there was a conformation that exposes aromatic residues and captured HA by stacking interactions and then rearranged structural elements, such as the  $\beta 4/\beta 5$  loop (Higman et al., 2007). There were two structural elements that were obviously solidified, one of which was G<sup>10</sup> located at the corner of  $\alpha 1/\beta 1$ , and the other was K<sup>54</sup> of  $\beta 3/\beta 4$ . K<sup>54</sup> was far from the HA-binding site but played an important role in the binding of heparin to TSG-6. Its solidification explained the problem that HA and heparin could not bind to TSG-6 at the same time, although they have different binding sites.

In the 2014 study, HA and hybrid HA of different lengths were used to study the interaction with Link-TSG-6 (Higman et al., 2014). Although the heptasaccharide with the reducing end of GlcA (HA<sub>7</sub><sup>AA</sup>) had a complete binding structure, the entropy was unfavorable. Therefore, the octasaccharide with the reducing end of GlcNAc (HA<sub>8</sub><sup>AN</sup>) was defined as the minimum unit required for binding. HSQC data clearly showed that HA<sub>8</sub><sup>NA</sup> and HA<sub>7</sub><sup>AA</sup> had two binding modes, with the reducing end GlcA bound to K<sup>63</sup>/H<sup>45</sup> as the dominant one. The affinity of HA<sub>8</sub><sup>NA</sup> was twice that of HA<sub>8</sub><sup>AN</sup>, while the affinity of the two heptasaccharides had no such difference. The reason for the difference in specific affinity is unknown. In the binding model of HA<sub>8</sub><sup>AN</sup> and TSG-6, H<sup>45</sup> and K<sup>63</sup> appear to be new binding residues. They bound to the reducing terminal disaccharide of the octasaccharide to make the binding tighter. The binding of HA and Link-TSG-6 was mainly through ionic interactions, ring-stacking interactions, hydrogen bonding, van der Waals forces and hydrophobic repulsion. Since the binding occurred on two interfaces, this imposed an inevitable requirement for the distortion of the two glycosidic bonds between the fifth and seventh residues. For heptasaccharides, the significant reduction in the affinity of hexasaccharides might be due to the lack of multiple groups of binding, resulting in instability of the distortion of glycosidic bonds. The CS part of hybrid HA will also be distorted during binding, but due to the lack of structural elements and the lack of hydrogen bonds during binding, the affinity was far lower than that of HA. However, due to the existence of binding, this provided a certain explanation for the chondroprotective function of TSG-6. CS, Heparin and HA

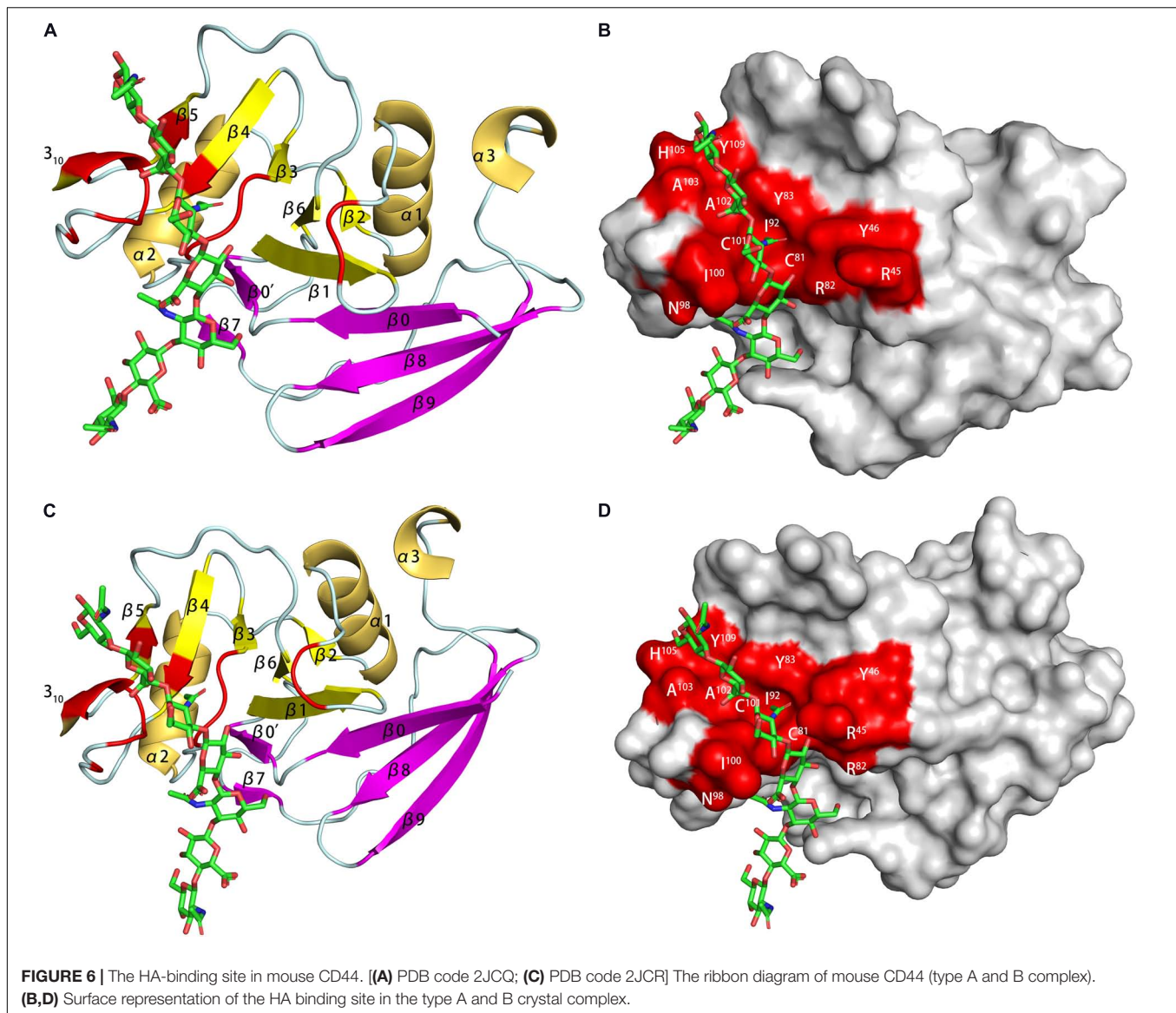


had different binding modes with TSG-6, giving TSG-6 complex biological functions.

The HABD in CD44 was mainly located in the link module, C-terminal extension and  $\alpha 1$ -helix. Two N-linked glycosylation sites (N<sup>25</sup> and N<sup>100</sup>) were also located in the HABD (Takeda et al., 2003). Teriete pointed out that octasaccharide might be the smallest unit that satisfies all binding requirements (Teriete et al., 2004). All binding sites were located on the same plane, but due to the scattered distribution, there might be two incompatible binding modes. One used N<sup>100</sup>/N<sup>101</sup> to R<sup>150</sup>/R<sup>154</sup>, similar to the combination of TSG-6 and HA. The other used K<sup>38</sup>/R<sup>162</sup> as the terminal binding, and the binding was farther away from the charged area. The data showed that the binding is accompanied by a structural rearrangement. Takeda proposed that the parallel

sheets of  $\beta 8$  and  $\beta 0$  involved rearrangement, which might be related to the special structure of  $\beta 8$  (Takeda et al., 2006). More thorough structural changes were located at the C-terminal extensions of  $\alpha 3$  and  $\beta 9$ , and their structure changed from a regular to a randomized structure after the combination. This result was in conflict with crystal studies, which showed that binding did not involve changes in C-terminal extension (Banerji et al., 2007). But unlike other studies, the protein used by Banerji is of mouse origin. And in the model established in this study, the complex is in two conformational equilibrium (type A and B, Figure 6). The difference between the two conformations is the orientation of R<sup>45</sup> (human CD44 R<sup>41</sup>). Ogino also proposed that CD44 was in the balance of two conformations in the unbound or bound state (Ogino et al., 2010). In the unbound state, it had a





regular structure and low HA affinity, which was conducive to cell rolling. In the combined state, it was mainly a random structure with high HA affinity, which was conducive to cell adhesion. The balance of these two states was conducive to the physiological activity of CD44-mediated cell rolling.

In terms of RHAMM, two amino acid clusters were mainly involved in binding with HA: the first was the proposed BX<sub>7</sub>B structure (K<sup>531</sup>-K<sup>541</sup>), and the second was K<sup>553</sup>-K<sup>562</sup> (Ziebell and Prestwich, 2004). Studies have shown that the second binding site plays a major role in binding. Studies on Tα1 indicated that the binding is mainly related to its terminal L<sup>16</sup>KEKK<sup>20</sup> (Mandaliti et al., 2017). The combination of HA and these two substances occurred mainly through electrostatic forces, which was different from the role of HA with TSG-6 and CD44. The combination of HA and CD44 was mainly through hydrogen bonding and van der Waals forces, while the combination with TSG-6 was mainly through electrostatic forces and aromatic accumulation.

## KERTAN SULFATE

Kertan sulfate is the only GAG without any acidic uronic acid residue, and its interaction with proteins mainly depends on structural characteristics and sulfation modification. KS is mainly distributed in the cornea and cartilage tissue and is divided into three categories (I-III) according to the distribution and connection with glycoproteins (Caterson and Melrose, 2018). KS plays an important role in brain development, neurodevelopment and regeneration, implantation and fertilization and maintains the balance of tissue hydration properties (Ota et al., 2018; Melrose, 2019; Miller et al., 2020). KS has many protein partners, including tyrosine protein kinases, inflammatory cytokines, growth factors, chemokines, cytoskeletal cells, and lectins. Only a few studies of the interaction between KS and protein have been investigated using NMR (Huckerby, 2002).



Galectin 3 (Gal-3) seems to be one of KS's most tacit partners, and its distribution is extremely close to that of KS. The interaction between full-length Gal-3 and KS has been studied using HSQC; the disturbance was found to be in the  $\beta 1$ ,  $\beta 3$ ,  $\beta 4$ ,  $\beta 5$ ,  $\beta 6$ , and  $\beta 10$  strands, and the  $\beta 10$  strand was the most important strand. The binding domain can be on the S- and F-faces in Gal-3. When the N-terminal tail of Gal-3 was truncated, KS interaction on the S-face became more obvious. The presence of other negatively charged regions did not affect the binding between KS and the Gal-3 S-face according to MD data. In the binding state, the conformations of the F-face and the N-terminal tail were changed. The binding was mainly concentrated on the left side of the S-face, which facilitated its combination with other proteins or heteropolymerization with other galectins. However, the pulse field gradient NMR data showed that KS did not induce oligomerization of Gal-3. Desulfated KS had far less affinity than KS, and the chemical shift disturbances on the F-face and N-terminal tail were greatly reduced.

## CONCLUSION

Glycosaminoglycans, as common glycoproteins in biological systems, are involved in many physiological and pathological processes. The study of their structure and interaction with proteins has received extensive attention, but the study of molecular perspectives is only the tip of the iceberg. This not only is due to the delay of carbohydrate research but is also related to the limitations of technology. The information produced by NMR is incomparable to all other technologies. For example, it can provide information about the binding

affinity constant, on/off chemical exchange rate, binding site and atomic information, but high-precision research is more demanding for technology. In particular, regarding the special existence of GAG, its highly complex structure not only endows it with rich biological functions but also brings incomparable difficulties for research. The study of the interactions between GAG and proteins using NMR is based on complete structural characterizations of GAG and/or proteins, which face huge obstacles. Biosynthesis carriers of GAG are difficult to find, while chemical and enzymatic syntheses are limited to a few scientists. This in turn makes it difficult to obtain isotope-labeled GAG. Because the binding of GAG and protein has obvious multibinding characteristics, it will cause oligomerization and even precipitation. The application of NMR technology is mainly limited by several factors, including the length of the oligosaccharides, the molecular weight of the proteins, and the concentration range and stability of the complex. However, with the renewal and iteration of technology, the rise of high magnetic flux nuclear magnetic spectrometry and enzymatic chemical synthesis has injected a steady stream of vitality into interaction research. The study of the interaction between GAG and proteins is helpful for understanding various physiological and pathological mechanisms and has a huge impetus for drug development.

## AUTHOR CONTRIBUTIONS

CB and LJ participated in preparation, creation, initial draft writing and review of this article. Both authors contributed to the article and approved the submitted version.

## REFERENCES

- Almond, A. (2018). Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges. *Curr. Opin. Struct. Biol.* 50, 58–64. doi: 10.1016/j.sbi.2017.11.008
- Almond, A., Brass, A., and Sheehan, J. K. (1998). Dynamic exchange between stabilized conformations predicted for hyaluronan tetrasaccharides: comparison of molecular dynamics simulations with available NMR data. *Glycobiology* 8, 973–980. doi: 10.1093/glycob/8.10.973
- Amara, A., Lorthioir, O., Valenzuela, A., Magerus, A., Thelen, M., Montes, M., et al. (1999). Stromal cell-derived factor-1 $\alpha$  associates with heparan sulfates through the first beta-strand of the chemokine. *J. Biol. Chem.* 274, 23916–23925. doi: 10.1074/jbc.274.34.23916
- Banerji, S., Wright, A. J., Noble, M., Mahoney, D. J., Campbell, I. D., Day, A. J., et al. (2007). Structures of the CD44–hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nat. Struct. Mol. Biol.* 14, 234–239. doi: 10.1038/nsmb1201
- Blundell, C. D., Mahoney, D. J., Almond, A., DeAngelis, P. L., Kahmann, J. D., Teriete, P., et al. (2003). The link module from ovulation- and inflammation-associated protein TSG-6 changes conformation on hyaluronan binding. *J. Biol. Chem.* 278, 49261–49270. doi: 10.1074/jbc.M309623200
- Brown, A., Robinson, C. J., Gallagher, J. T., and Blundell, T. L. (2013). Cooperative heparin-mediated oligomerization of fibroblast growth factor-1 (FGF1) precedes recruitment of FGFR2 to ternary complexes. *Biophys. J.* 104, 1720–1730. doi: 10.1016/j.bpj.2013.02.051
- Brown, A., Sepuru, K., and Rajarathnam, K. (2017). Structural basis of native CXCL7 monomer binding to CXCR2 receptor N-domain and glycosaminoglycan heparin. *Int. J. Mol. Sci.* 18:508. doi: 10.3390/ijms18030508
- Cai, S., Dufner-Beattie, J. L., and Prestwich, G. D. (2004). A selective protein sensor for heparin detection. *Analytical Biochem.* 326, 33–41. doi: 10.1016/j.ab.2003.11.017
- Canales, A., Angulo, J., Ojeda, R., Bruix, M., Fayos, R., Lozano, R., et al. (2005). Conformational flexibility of a synthetic glycosylaminoglycan bound to a fibroblast growth factor. FGF-1 recognizes both the 1C4 and 2SO conformations of a bioactive heparin-like hexasaccharide. *J. Am. Chem. Soc.* 127, 5778–5779. doi: 10.1021/ja043363y
- Canales, A., Lozano, R., López Méndez, B., Angulo, J., Ojeda, R., Nieto, P. M., et al. (2006). Solution NMR structure of a human FGF-1 monomer, activated by a hexasaccharide heparin-analogue. *FEBS J.* 273, 4716–4727. doi: 10.1111/j.1742-4658.2006.05474.x
- Canales-Mayordomo, A., Fayos, R., Angulo, J., Ojeda, R., Martín-Pastor, M., Nieto, P. M., et al. (2006). Backbone dynamics of a biologically active human FGF-1 monomer, complexed to a hexasaccharide heparin-analogue, by 15N NMR relaxation methods. *J. Biomol. NMR* 35, 225–239. doi: 10.1007/s10858-006-9024-y
- Cardin, A. D., and Weintraub, H. J. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* 9, 21–32. doi: 10.1161/01.atv.9.1.21
- Casu, B., Oreste, P., Torri, G., Zoppetti, G., Choay, J., Lormeau, J. C., et al. (1981). The structure of heparin oligosaccharide fragments with high anti- (factor Xa) activity containing the minimal antithrombin III-binding sequence. Chemical and 13C nuclear-magnetic-resonance studies. *Biochem. J.* 197, 599–609. doi: 10.1042/bj1970599
- Caterson, B., and Melrose, J. (2018). Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* 28, 182–206. doi: 10.1093/glycob/cwy003

- Cheng, W., Wang, M., Tian, X., and Zhang, X. (2017). An overview of the binding models of FGFR tyrosine kinases in complex with small molecule inhibitors. *Eur. J. Med. Chem.* 126, 476–490. doi: 10.1016/j.ejmech.2016.11.052
- Choay, J., Petitou, M., Lormeau, J. C., Sinaÿ, P., Casu, B., and Gatti, G. (1983). Structure-activity relationship in heparin: a synthetic pentasaccharide with high affinity for antithrombin III and eliciting high anti-factor Xa activity. *Biochem. Biophys. Res. Commun.* 116, 492–499. doi: 10.1016/0006-291X(83)90550-8
- Conrad, H. E. (1997). *Heparin-Binding Proteins*. Illinois: Academic Press.
- Crijns, H., Vanheule, V., and Proost, P. (2020). Targeting chemokine—glycosaminoglycan interactions to inhibit inflammation. *Front. Immunol.* 11:483. doi: 10.3389/fimmu.2020.00483
- Dahms, S. O., Mayer, M. C., Roeser, D., Multhaupt, G., and Than, M. E. (2015). Interaction of the amyloid precursor protein-like protein 1 (APLP1) E2 domain with heparan sulfate involves two distinct binding modes. *Acta Crystallographica Sec. D Biol. Crystallography* 71, 494–504. doi: 10.1107/S1399004714027114
- Das, S. K., Mallet, J. M., Esnault, J., Driguez, P. A., Duchaussoy, P., Sizin, P., et al. (2001). Synthèse konformativ fixierter kohlenhydrate: eine skew-bootkonformation der L-Iduronsäure in Heparin bestimmt dessen antithrombotische Aktivität. *Angewandte Chemie* 113, 1723–1726. doi: 10.1002/1521-3757(20010504)113:93.0.CO;2-N
- Day, A. J., and Milner, C. M. (2019). TSG-6: a multifunctional protein with anti-inflammatory and tissue-protective properties. *Matrix Biol.* 78, 60–83. doi: 10.1016/j.matbio.2018.01.011
- Deakin, J. A., Blaum, B. S., Gallagher, J. T., Uhrin, D., and Lyon, M. (2009). The binding properties of minimal oligosaccharides reveal a common heparan sulfate/dermatan sulfate-binding site in hepatocyte growth factor/scatter factor that can accommodate a wide variety of sulfation patterns. *J. Biol. Chem.* 284, 6311–6321. doi: 10.1074/jbc.M807671200
- Desai, U. R., Petitou, M., Björk, I., and Olson, S. T. (1998). Mechanism of heparin activation of antithrombin: evidence for an induced-fit model of allosteric activation involving two interaction subsites. *Biochemistry* 37, 13033–13041. doi: 10.1021/bi981426h
- Deshauer, C., Morgan, A. M., Ryan, E. O., Handel, T. M., Prestegard, J. H., and Wang, X. (2015). Interactions of the chemokine CCL5/RANTES with medium-sized chondroitin sulfate ligands. *Structure* 23, 1066–1077. doi: 10.1016/j.str.2015.03.024
- Döbeli, H., Gentz, R., Jucker, W., Garotta, G., Hartmann, D. W., and Hochuli, E. (1988). Role of the carboxy-terminal sequence on the biological activity of human immune interferon (IFN- $\gamma$ ). *J. Biotechnol.* 7, 199–216. doi: 10.1016/0168-1656(88)90052-1
- Dyer, D. P., Migliorini, E., Salanga, C. L., Thakar, D., Handel, T. M., and Richter, R. P. (2017). Differential structural remodelling of heparan sulfate by chemokines: the role of chemokine oligomerization. *Open Biol.* 7:160286. doi: 10.1098/rsob.160286
- Elli, S., Stancanelli, E., Wang, Z., Petitou, M., Liu, J., and Guerrini, M. (2020). Degeneracy of the antithrombin binding sequence in heparin: 2-O-sulfated iduronic acid can replace the critical glucuronic acid. *Chem. Eur. J.* 26, 11814–11818. doi: 10.1002/chem.202001346
- Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J., and Rees, D. C. (1996). Heparin structure and interactions with basic fibroblast growth factor. *Science* 271, 1116–1120. doi: 10.1126/science.271.5252.1116
- Fallacara, A., Baldini, E., Manfredini, S., and Vertuani, S. (2018). Hyaluronic acid in the third millennium. *Polymers* 10:701. doi: 10.3390/polym10070701
- Feng, W., and Wang, X. (2015). Structure of decorin binding protein B from *Borrelia burgdorferi* and its interactions with glycosaminoglycans. *Biochim. Biophys. Acta Proteins Proteomics* 1854, 1823–1832. doi: 10.1016/j.bbapap.2015.08.003
- Ferro, D. R., Provasoli, A., Ragazzi, M., Torri, G., Casu, B., Gatti, G., et al. (1987). Evidence for conformational equilibrium of the sulfated L-iduronate residue in heparin and in synthetic heparin mono- and oligosaccharides: NMR and force-field studies. *Cheminform* 108, 6773–6778. doi: 10.1002/chin.198710100
- Gao, Q., Chen, C., Zong, C., Wang, S., Ramiah, A., Prabhakar, P., et al. (2016). Structural aspects of heparan sulfate binding to robo1–Ig1–2. *ACS Chem. Biol.* 11, 3106–3113. doi: 10.1021/acschembio.6b00692
- García-Mayoral, M. F., Canales, Á., Díaz, D., López-Prados, J., Moussaoui, M., de Paz, J. L., et al. (2013). Insights into the glycosaminoglycan-mediated cytotoxic mechanism of eosinophil cationic protein revealed by NMR. *ACS Chem. Biol.* 8, 144–151. doi: 10.1021/cb300386v
- Gargiulo, V., Morando, M. A., Silipo, A., Nurisso, A., Pérez, S., Imberty, A., et al. (2010). Insights on the conformational properties of hyaluronic acid by using NMR residual dipolar couplings and MD simulations. *Glycobiology* 20, 1208–1216. doi: 10.1093/glycob/cwq067
- Gehrcke, J., and Pisabarro, M. T. (2015). Identification and characterization of a glycosaminoglycan binding site on interleukin-10 via molecular simulation methods. *J. Mol. Graphics Modelling* 62, 97–104. doi: 10.1016/j.jmgm.2015.09.003
- Gray, E., Hogwood, J., and Mulloy, B. (2012). The anticoagulant and antithrombotic mechanisms of heparin. *Handb. Exp. Pharmacol.* 207, 43–61. doi: 10.1007/978-3-642-23056-1\_3
- Guerrini, M., Agulles, T., Bisio, A., Hricovini, M., Lay, L., Naggi, A., et al. (2002). Minimal heparin/heparan sulfate sequences for binding to fibroblast growth factor-1. *Biochem. Biophys. Res. Commun.* 292, 222–230. doi: 10.1006/bbrc.2002.6634
- Guerrini, M., Elli, S., Gaudesi, D., Torri, G., Casu, B., Mourier, P., et al. (2010). Effects on molecular conformation and anticoagulant activities of 1,6-anhydrosugars at the reducing terminal of antithrombin-binding octasaccharides isolated from low-molecular-weight heparin enoxaparin. *J. Med. Chem.* 53, 8030–8040. doi: 10.1021/jm100771s
- Guerrini, M., Elli, S., Mourier, P., Rudd, T. R., Gaudesi, D., Casu, B., et al. (2013). An unusual antithrombin-binding heparin octasaccharide with an additional 3-O-sulfated glucosamine in the active pentasaccharide sequence. *Biochem. J.* 449, 343–351. doi: 10.1042/BJ20121309
- Guerrini, M., Guglieri, S., Beccati, D., Torri, G., Viskov, C., and Mourier, P. (2006). Conformational transitions induced in heparin octasaccharides by binding with antithrombin III. *Biochem. J.* 399, 191–198. doi: 10.1042/BJ20060656
- Guerrini, M., Guglieri, S., Casu, B., Torri, G., Mourier, P., Boudier, C., et al. (2008). Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. *J. Biol. Chem.* 283, 26662–26675. doi: 10.1074/jbc.M801102200
- Guglieri, S., Hricovini, M., Raman, R., Polito, L., Torri, G., Casu, B., et al. (2008). Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy. *Biochemistry* 47, 13862–13869. doi: 10.1021/bi801007p
- Haasnoot, C. A. G., de Gelder, R., Kooijman, H., and Kellenbach, E. R. (2020). The conformation of the idopyranose ring revisited: How subtle O-substituent induced changes can be deduced from vicinal  $^1\text{H}$ -NMR coupling constants. *Carbohydrate Res.* 496:108052. doi: 10.1016/j.carres.2020.108052
- Hanske, J., Wawrzinek, R., Geissner, A., Wamhoff, E., Sellrie, K., Schmidt, H., et al. (2017). Calcium-independent activation of an allosteric network in langerin by heparin oligosaccharides. *ChemBioChem.* 18, 1183–1187. doi: 10.1002/cbic.201700027
- Herradon, G., Ramos-Alvarez, M. P., and Gramage, E. (2019). Connecting meta-inflammation and neuroinflammation through the PTN-MK-RPTP $\beta$ / $\zeta$  axis: relevance in therapeutic development. *Front. Pharmacol.* 10:377. doi: 10.3389/fphar.2019.00377
- Higman, V. A., Blundell, C. D., Mahoney, D. J., Redfield, C., Noble, M. E. M., and Day, A. J. (2007). Plasticity of the TSG-6 HA-binding loop and mobility in the TSG-6-HA complex revealed by NMR and X-ray crystallography. *J. Mol. Biol.* 371, 669–684. doi: 10.1016/j.jmb.2007.05.073
- Higman, V. A., Briggs, D. C., Mahoney, D. J., Blundell, C. D., Sattelle, B. M., Dyer, D. P., et al. (2014). A refined model for the TSG-6 link module in complex with hyaluronan. *J. Biol. Chem.* 289, 5619–5634. doi: 10.1074/jbc.M113.542357
- Hricovini, M. (2004). Structural aspects of carbohydrates and the relation with their biological properties. *Curr. Med. Chem.* 11, 2565–2583. doi: 10.2174/0929867043364414
- Hricovini, M., Guerrini, M., Bisio, A., Torri, G., Naggi, A., and Casu, B. (2002). Active conformations of glycosaminoglycans. NMR determination of the conformation of heparin sequences complexed with antithrombin and fibroblast growth factors in solution. *Semin. Thrombosis Hemostasis* 28, 325–334. doi: 10.1055/s-2002-34301
- Hu, Y., Zhong, Y., Chen, Z., Chen, C., Shi, Z., Zulueta, M. M. L., et al. (2012). Divergent synthesis of 48 heparan sulfate-based disaccharides and probing the specific sugar–fibroblast growth factor-1 interaction. *J. Am. Chem. Soc.* 134, 20722–20727. doi: 10.1021/ja3090065

- Huang, R., and Leung, I. K. H. (2019). Protein-small molecule interactions by waterLOGSY. *Methods Enzymol.* 615, 477–500. doi: 10.1016/bs.mie.2018.08.020
- Huckerby, T. (2002). The keratan sulphates: structural investigations using NMR spectroscopy. *Prog. Nuclear Magn. Resonance Spect.* 40, 35–110. doi: 10.1016/S0079-6565(01)00040-1
- Hughes, A., Meneghetti, M., Huang, T., Hung, S., Elli, S., Guerrini, M., et al. (2017). Investigating the relationship between temperature, conformation and calcium binding in heparin model oligosaccharides. *Carbohydrate Res.* 438, 58–64. doi: 10.1016/j.carres.2016.12.002
- Huynh, M. B., Ouidja, M. O., Chantepie, S., Carpentier, G., Maiza, A., Zhang, G., et al. (2019). Glycosaminoglycans from Alzheimer's disease hippocampus have altered capacities to bind and regulate growth factors activities and to bind tau. *PLoS One* 14:e209573. doi: 10.1371/journal.pone.0209573
- Janssens, R., Mortier, A., Boff, D., Ruytinx, P., Gouwy, M., Vantilt, B., et al. (2017a). Truncation of CXCL12 by CD26 reduces its CXC chemokine receptor 4- and atypical chemokine receptor 3-dependent activity on endothelial cells and lymphocytes. *Biochem. Pharmacol.* 132, 92–101. doi: 10.1016/j.bcp.2017.03.009
- Janssens, R., Struyf, S., and Proost, P. (2017b). The unique structural and functional features of CXCL12. *Cell. Mol. Immunol.* 15, 299–311. doi: 10.1038/cmi.2017.107
- Jimenez-Barbero, J., and Peters, T. (2003). *NMR spectroscopy of glycoconjugates*. Weinheim: Wiley-VCH. 74.
- José García-Jiménez, M., Corzana, F., De Paz, J. L., and Nieto, P. M. (2019). Langerin-heparin interaction: analysis of the binding to the non-lectin site. *Nat. Prod. Commun.* 14, 1934578X–1985159X. doi: 10.1177/1934578X19851597
- Joseph, P. R. B., Mosier, P. D., Desai, U. R., and Rajarathnam, K. (2015). Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions. *Biochem. J.* 472, 121–133. doi: 10.1042/BJ20150059
- Kahmann, J. D., O'Brien, R., Werner, J. M., Heinegard, D., Ladbury, J. E., Campbell, I. D., et al. (2000). Localization and characterization of the hyaluronan-binding site on the link module from human TSG-6. *Structure* 8, 763–774. doi: 10.1016/S0969-2126(00)00163-5
- Kan, M., Wang, F., Xu, J., Crabb, J. W., Hou, J., and McKeehan, W. L. (1993). An essential heparin-binding domain in the fibroblast growth factor receptor kinase. *Science* 259, 1918–1921. doi: 10.1126/science.8456318
- Kato, K., and Peters, T. (2017). *NMR in glycoscience and glycotechnology*. London: Royal Society of Chemistry.
- Kohda, D., Morton, C. J., Parkar, A. A., Hatanaka, H., Inagaki, F. M., Campbell, I. D., et al. (1996). Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell* 86, 767–775. doi: 10.1016/S0092-8674(00)80151-8
- Korsensky, L., and Ron, D. (2016). Regulation of FGF signaling: recent insights from studying positive and negative modulators. *Semin. Cell. Dev. Biol.* 53, 101–114. doi: 10.1016/j.semcdb.2016.01.023
- Künze, G., Gehrcke, J., Pisabarro, M. T., and Huster, D. (2014). NMR characterization of the binding properties and conformation of glycosaminoglycans interacting with interleukin-10. *Glycobiology* 24, 1036–1049. doi: 10.1093/glycob/cwu069
- Künze, G., Köhling, S., Vogel, A., Rademann, J., and Huster, D. (2016). Identification of the glycosaminoglycan binding site of interleukin-10 by NMR spectroscopy. *J. Biol. Chem.* 291, 3100–3113. doi: 10.1074/jbc.M115.681759
- Laguri, C., Sapay, N., Simorre, J., Brutscher, B., Imberty, A., Gans, P., et al. (2011). <sup>13</sup>C-labeled heparan sulfate analogue as a tool to study protein/heparan sulfate interactions by NMR Spectroscopy: application to the CXCL12α Chemokine. *J. Am. Chem. Soc.* 133, 9642–9645. doi: 10.1021/ja201753e
- Lane, D. A., Denton, J., Flynn, A. M., Thunberg, L., and Lindahl, U. (1984). Anticoagulant activities of heparin oligosaccharides and their neutralization by platelet factor 4. *Biochem. J.* 218, 725–732. doi: 10.1042/bj2180725
- Ledwith, K. V., Barnes, R. W., and Roberts, A. G. (2016). Unravelling the complex drug-drug interactions of the cardiovascular drugs, verapamil and digoxin, with P-glycoprotein. *Biosci. Rep.* 36:e309. doi: 10.1042/BSR20150317
- Li, X., Wang, C., Xiao, J., McKeehan, W. L., and Wang, F. (2016). Fibroblast growth factors, old kids on the new block. *Semin. Cell Dev. Biol.* 53, 155–167. doi: 10.1016/j.semcdb.2015.12.014
- Lindahl, U., Backstrom, G., Thunberg, L., and Leder, I. G. (1980). Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin. *Proc. Natl. Acad. Sci. U S A.* 77, 6551–6555. doi: 10.1073/pnas.77.11.6551
- Liu, J., Li, J., Arnold, K., Pawlinski, R., and Key, N. S. (2020). Using heparin molecules to manage COVID-2019. *Res. Prac. Thrombosis Haemostasis* 4, 518–523. doi: 10.1002/rth2.12353
- Malmos, K. G., Bjerring, M., Jessen, C. M., Nielsen, E. H. T., Poulsen, E. T., Christiansen, G., et al. (2016). How glycosaminoglycans promote fibrillation of salmon calcitonin. *J. Biol. Chem.* 291, 16849–16862. doi: 10.1074/jbc.M116.715466
- Mandaliti, W., Nepravishta, R., Pica, F., Vallebona, P. S., Garaci, E., and Paci, M. (2017). Thymosin α1 interacts with hyaluronic acid electrostatically by its terminal sequence LKEKK. *Molecules* 22:1843. doi: 10.3390/molecules22111843
- Melrose, J. (2019). Keratan sulfate (KS)-proteoglycans and neuronal regulation in health and disease: the importance of KS-glycodynamics and interactive capability with neuroregulatory ligands. *J. Neurochem.* 149, 170–194. doi: 10.1111/jnc.14652
- Miller, M. C., Cai, C., Wichapong, K., Bhaduri, S., Pohl, N. L. B., Linhardt, R. J., et al. (2020). Structural insight into the binding of human galectins to corneal keratan sulfate, its desulfated form and related saccharides. *Sci. Rep.* 10, 1–18. doi: 10.1038/s41598-020-72645-9
- Monneau, Y. R., Luo, L., Sankaranarayanan, N. V., Nagarajan, B., Vivès, R. R., Baleux, F., et al. (2017). Solution structure of CXCL13 and heparan sulfate binding show that GAG binding site and cellular signalling rely on distinct domains. *Open Biol.* 7:170133. doi: 10.1098/rsob.170133
- Morla, S. (2019). Glycosaminoglycans and glycosaminoglycan mimetics in cancer and inflammation. *Int. J. Mol. Sci.* 20:1963. doi: 10.3390/ijms20081963
- Moy, F. J., Safran, M., Seddon, A. P., Kitchen, D., Bohlen, P., Aviezer, D., et al. (1997). Properly oriented heparin-decasaccharide-induced dimers are the biologically active form of basic fibroblast growth factor. *Biochemistry* 36, 4782–4791. doi: 10.1021/bi9625455
- Moy, F. J., Seddon, A. P., Campbell, E. B., Böhlen, P., and Powers, R. (1995). <sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C and <sup>13</sup>CO assignments and secondary structure determination of basic fibroblast growth factor using 3D heteronuclear NMR spectroscopy. *J. Biomol. NMR* 6, 245–254. doi: 10.1007/BF00197806
- Mulloy, B. (2006). Progress in the structural biology of chondroitin sulfate. *Adv. Pharmacol.* 53:49. doi: 10.1016/S1054-3589(05)53004-7
- Muñoz-García, J. C., Chabrol, E., Vivès, R. R., Thomas, A., de Paz, J. L., Rojo, J., et al. (2015). Langerin-heparin interaction: two binding sites for small and large ligands as revealed by a combination of nmr spectroscopy and cross-linking mapping experiments. *J. Am. Chem. Soc.* 137, 4100–4110. doi: 10.1021/ja511529x
- Muñoz-García, J. C., López-Prados, J., Angulo, J., Díaz-Contreras, I., Reichardt, N., de Paz, J. L., et al. (2012). Effect of the substituents of the neighboring ring in the conformational equilibrium of iduronate in heparin-like trisaccharides. *Chem. Eur. J.* 18, 16319–16331. doi: 10.1002/chem.201202770
- Muñoz-García, J. C., Solera, C., Carrero, P., de Paz, J. L., Angulo, J., and Nieto, P. M. (2013). 3D structure of a heparin mimetic analogue of a FGF-1 activator. A NMR and molecular modelling study. *Organic Biomol. Chem.* 11:8269. doi: 10.1039/c3ob41789a
- Murphy, J. W., Cho, Y., Sachpatzidis, A., Fan, C., Hodsdon, M. E., and Lolis, E. (2007). Structural and functional basis of CXCL12 (stromal cell-derived Factor-1α) binding to heparin. *J. Biol. Chem.* 282, 10018–10027. doi: 10.1074/jbc.M608796200
- Murphy, P. M., and Heusinkveld, L. (2018). Multisystem multitasking by CXCL12 and its receptors CXCR4 and ACKR3. *Cytokine* 109, 2–10. doi: 10.1016/j.cyt.2017.12.022
- Naggi, A., Gardini, C., Pedrinola, G., Mauri, L., Urso, E., Alekseeva, A., et al. (2016). Structural peculiarity and antithrombin binding region profile of mucosal bovine and porcine heparins. *J. Pharmaceutical Biomed. Analysis* 118, 52–63. doi: 10.1016/j.jpba.2015.10.001
- Naor, D. (2016). Editorial: interaction between hyaluronic acid and its receptors (CD44, RHAMM) regulates the activity of inflammation and cancer. *Front. Immunol.* 7:39. doi: 10.3389/fimmu.2016.00039
- Nieto, L., Canales, Á, Fernández, I. S., Santillana, E., González-Corrochano, R., Redondo-Horcajo, M., et al. (2013). Heparin modulates the mitogenic activity of fibroblast growth factor by inducing dimerization of its receptor. A 3D view by using NMR. *Chem. BioChem.* 14, 1732–1744. doi: 10.1002/cbic.201300313



- Nieto, L., Canales, Á., Giménez-Gallego, G., Nieto, P. M., and Jiménez-Barbero, J. (2011). Conformational selection of the AGA\*IAM heparin pentasaccharide when bound to the fibroblast growth factor receptor. *Chem. Eur. J.* 17, 11204–11209. doi: 10.1002/chem.201101000
- Nogueira, A. V., Rossi, G. R., Iacomini, M., Sasaki, G. L., Trindade, E. S., and Cipriani, T. R. (2019). Viscera of fishes as raw material for extraction of glycosaminoglycans of pharmacological interest. *Int. J. Biol. Macromol.* 121, 239–248. doi: 10.1016/j.ijbiomac.2018.09.156
- Oduah, E., Linhardt, R., and Sharfstein, S. (2016). Heparin: past, present, and future. *Pharmaceuticals* 9:38. doi: 10.3390/ph9030038
- Ogino, S., Nishida, N., Umemoto, R., Suzuki, M., Takeda, M., Terasawa, H., et al. (2010). Two-state conformations in the hyaluronan-binding domain regulate CD44 adhesiveness under flow condition. *Structure* 18, 649–656. doi: 10.1016/j.str.2010.02.010
- Ogura, K., Nagata, K., Hatanaka, H., Habuchi, H., Kimata, K., Tate, S., et al. (1999). Solution structure of human acidic fibroblast growth factor and interaction with heparin-derived hexasaccharide. *J. Biomol. NMR* 13, 11–24. doi: 10.1023/A:1008330622467
- Ojeda, R., Angulo, J., Nieto, P. M., and Martín-Lomas, M. (2002). The activation of fibroblast growth factors by heparin: synthesis and structural study of rationally modified heparin-like oligosaccharides. *Can. J. Chem.* 80, 917–936. doi: 10.1139/v02-023
- Olson, S. T., Ingemar, B. R., and Bock, S. C. (2002). Identification of critical molecular interactions mediating heparin activation of antithrombin: implications for the design of improved heparin anticoagulants. *Trends Cardiovas. Med.* 12, 198–205. doi: 10.1016/S1050-1738(02)00160-3
- Olson, S. T., Srinivasan, K. R., Björk, I., and Shore, J. D. (1981). Binding of high affinity heparin to antithrombin III. Stopped flow kinetic studies of the binding interaction. *J. Biol. Chem.* 256, 11073–11079. doi: 10.1016/0165-022X(81)90075-0
- Orton, H. W., Kuprov, I., Loh, C., and Otting, G. (2016). Using paramagnetism to slow down nuclear relaxation in protein NMR. *J. Phys. Chem. Lett.* 7, 4815–4818. doi: 10.1021/acs.jpclett.6b02417
- Ota, F., Hirayama, T., Kizuka, Y., Yamaguchi, Y., Fujinawa, R., Nagata, M., et al. (2018). High affinity sugar ligands of C-type lectin receptor langerin. *Biochim. Biophys. Acta Gen. Subj.* 1862, 1592–1601. doi: 10.1016/j.bbagen.2018.04.004
- Palhares, L. C. G. F., Brito, A. S., de Lima, M. A., Nader, H. B., London, J. A., Barsukov, I. L., et al. (2019). A further unique chondroitin sulfate from the shrimp *litopenaeus vannamei* with antithrombin activity that modulates acute inflammation. *Carbohydrate Polymers* 222:115031. doi: 10.1016/j.carbpol.2019.115031
- Panitz, N., Theissen, S., Samsonov, S. A., Gehrcke, J., Baumann, L., Bellmann-Sickert, K., et al. (2016). The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites. *Glycobiology* 26, 1209–1221. doi: 10.1093/glycob/cwv059
- Park, Y., Jowitt, T. A., Day, A. J., and Prestegard, J. H. (2016). Nuclear magnetic resonance insight into the multiple glycosaminoglycan binding modes of the link module from human TSG-6. *Biochemistry* 55, 262–276. doi: 10.1021/acs.biochem.5b01148
- Pellegrini, L. (2001). Role of heparan sulfate in fibroblast growth factor signalling: a structural view. *Curr. Opin. Struct. Biol.* 11, 629–634. doi: 10.1016/S0959-440X(00)00258-X
- Penk, A., Baumann, L., Huster, D., and Samsonov, S. A. (2019). NMR and molecular modeling reveal specificity of the interactions between CXCL14 and glycosaminoglycans. *Glycobiology* 29, 715–725. doi: 10.1093/glycob/cwz047
- Pestka, S., Krause, C. D., Sarkar, D., Walter, M. R., Shi, Y., and Fisher, P. B. (2004). Interleukin-10 and related cytokines and receptors. *Ann. Rev. Immunol.* 22, 929–979. doi: 10.1146/annurev.immunol.22.012703.104622
- Pichert, A., Samsonov, S. A., Theissen, S., Thomas, L., Baumann, L., Schiller, J., et al. (2012). Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology* 22, 134–145. doi: 10.1093/glycob/cwr120
- Pomin, V., and Mulloy, B. (2018). Glycosaminoglycans and proteoglycans. *Pharmaceuticals* 11:27. doi: 10.3390/ph11010027
- Pomin, V., and Wang, X. (2018a). Glycosaminoglycan-protein interactions by nuclear magnetic resonance (NMR) spectroscopy. *Molecules* 23:2314. doi: 10.3390/molecules23092314
- Pomin, V. H. (2014). Solution NMR conformation of glycosaminoglycans. *Prog. Biophys. Mol. Biol.* 114, 61–68. doi: 10.1016/j.pbiomolbio.2014.01.001
- Pomin, V. H. (2015). Keratan sulfate: an up-to-date review. *Int. J. Biol. Macromol.* 72, 282–289. doi: 10.1016/j.ijbiomac.2014.08.029
- Pomin, V. H. (2016). Paradigms in the structural biology of the mitogenic ternary complex FGF: FGFR: heparin. *Biochimie* 127, 214–226. doi: 10.1016/j.biochi.2016.05.017
- Pomin, V. H., and Wang, X. (2018b). Synthetic oligosaccharide libraries and microarray technology: a powerful combination for the success of current glycosaminoglycan interactomics. *Chem. MedChem.* 13, 648–661. doi: 10.1002/cmdc.201700620
- Przybylski, C., Gonnet, F., Saesen, E., Lortat-Jacob, H., and Daniel, R. (2020). Surface plasmon resonance imaging coupled to on-chip mass spectrometry: a new tool to probe protein-GAG interactions. *Analytical Bioanalytical Chem.* 412, 507–519. doi: 10.1007/s00216-019-02267-2
- Rajarathnam, K., Sepuru, K. M., Joseph, P. R. B., Sawant, K. V., and Brown, A. J. (2018). Glycosaminoglycan interactions fine-tune chemokine-mediated neutrophil trafficking: structural insights and molecular mechanisms. *J. Histochem. Cytochem.* 66, 229–239. doi: 10.1369/0022155417739864
- Rezaie, A. R., and Giri, H. (2020). Antithrombin: an anticoagulant, anti-inflammatory and antibacterial serpin. *J. Thrombosis Haemostasis* 18, 528–533. doi: 10.1111/jth.14724
- Rider, C. C., and Mulloy, B. (2017). Heparin, heparan sulphate and the TGF- $\beta$  cytokine superfamily. *Molecules* 22:713. doi: 10.3390/molecules22050713
- Roth, R., Swanson, R., Izaguirre, G., Bock, S. C., Gettins, P. G. W., and Olson, S. T. (2015). Saturation mutagenesis of the antithrombin reactive center loop P14 residue supports a three-step mechanism of heparin allosteric activation involving intermediate and fully activated states. *J. Biol. Chem.* 290, 28020–28036. doi: 10.1074/jbc.M115.678839
- Rudd, T. R., Preston, M. D., and Yates, E. A. (2017). The nature of the conserved basic amino acid sequences found among 437 heparin binding proteins determined by network analysis. *Mol. BioSyst.* 13, 852–865. doi: 10.1039/C6MB00857G
- Ryan, E., Shen, D., and Wang, X. (2016). Structural studies reveal an important role for the pleiotrophin C-terminus in mediating interactions with chondroitin sulfate. *FEBS J.* 283, 1488–1503. doi: 10.1111/febs.13686
- Sadir, R., Baleux, F., Grosdidier, A., Imbert, A., and Lortat-Jacob, H. (2001). Characterization of the stromal cell-derived factor-1 $\alpha$ -heparin complex. *J. Biol. Chem.* 276, 8288–8296. doi: 10.1074/jbc.M008110200
- Sadir, R., Imbert, A., Baleux, F., and Lortat-Jacob, H. (2004). Heparan sulfate/heparin oligosaccharides protect stromal cell-derived factor-1 (SDF-1)/CXCL12 against proteolysis induced by CD26/dipeptidyl peptidase IV. *J. Biol. Chem.* 279, 43854–43860. doi: 10.1074/jbc.M405392200
- Saesen, E., Sarrazin, S., Laguri, C., Sadir, R., Maurin, D., Thomas, A., et al. (2013). Insights into the mechanism by which interferon- $\gamma$  basic amino acid clusters mediate protein binding to heparan sulfate. *J. Am. Chem. Soc.* 135, 9384–9390. doi: 10.1021/ja4000867
- Sandoval, D. R., Gomez Toledo, A., Painter, C. D., Tota, E. M., Sheikh, M. O., West, A. M. V., et al. (2020). Proteomics-based screening of the endothelial heparan sulfate interactome reveals that C-type lectin 14a (CLEC14A) is a heparin-binding protein. *J. Biol. Chem.* 295, 2804–2821. doi: 10.1074/jbc.RA119.011639
- Sasarmar, F., Maftei, C., Campeau, P. M., Brunel-Guitton, C., Mitchell, G. A., and Allard, P. (2016). Biosynthesis of glycosaminoglycans: associated disorders and biochemical tests. *J. Inherited Metabolic Dis.* 39, 173–188. doi: 10.1007/s10545-015-9903-z
- Saxena, K., Schieborr, U., Anderka, O., Duchardt-Ferner, E., Elshorst, B., Gande, S. L., et al. (2010). Influence of heparin mimetics on assembly of the FGF-FGFR4 signaling complex. *J. Biol. Chem.* 285, 26628–26640. doi: 10.1074/jbc.M109.095109
- Schlessinger, J., Plotnikov, A. N., Ibrahim, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., et al. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell.* 6, 743–750. doi: 10.1016/S1097-2765(00)00073-3
- Senbanjo, L. T., and Chelliah, M. A. (2017). CD44: a multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front. Cell. Dev. Biol.* 5:18. doi: 10.3389/fcell.2017.00018



- Seo, E. S., Blaum, B. S., Vargues, T., De Cecco, M., Deakin, J. A., Lyon, M., et al. (2010). Interaction of Human  $\beta$ -Defensin 2 (HBD2) with glycosaminoglycans. *Biochemistry* 49, 10486–10495. doi: 10.1021/bi1011749
- Sepuru, K. M., Iwahara, J., and Rajarathnam, K. (2018a). Direct detection of lysine side chain NH3+ in protein-heparin complexes using NMR spectroscopy. *Analyst* 143, 635–638. doi: 10.1039/c7an01406f
- Sepuru, K. M., Nagarajan, B., Desai, U. R., and Rajarathnam, K. (2016). Molecular basis of chemokine CXCL5-glycosaminoglycan interactions. *J. Biol. Chem.* 291, 20539–20550. doi: 10.1074/jbc.M116.745265
- Sepuru, K. M., Nagarajan, B., Desai, U. R., and Rajarathnam, K. (2018b). Structural basis, stoichiometry, and thermodynamics of binding of the chemokines KC and MIP2 to the glycosaminoglycan heparin. *J. Biol. Chem.* 293, 17817–17828. doi: 10.1074/jbc.RA118.004866
- Sepuru, K. M., and Rajarathnam, K. (2016). CXCL1/MGSA Is a novel glycosaminoglycan (GAG)-binding chemokine. *J. Biol. Chem.* 291, 4247–4255. doi: 10.1074/jbc.M115.697888
- Sepuru, K. M., and Rajarathnam, K. (2019). Structural basis of chemokine interactions with heparan sulfate, chondroitin sulfate, and dermatan sulfate. *J. Biol. Chem.* 294, 15650–15661. doi: 10.1074/jbc.RA119.009879
- Severin, I. C., Gaudry, J., Johnson, Z., Kungl, A., Jansma, A., Gesslbauer, B., et al. (2010). Characterization of the chemokine CXCL11-heparin interaction suggests two different affinities for glycosaminoglycans. *J. Biol. Chem.* 285, 17713–17724. doi: 10.1074/jbc.M109.082552
- Silipo, A., Zhang, Z., Cañada, F. J., Molinaro, A., Linhardt, R. J., and Jiménez Barbero, J. (2008). Conformational analysis of a dermatan sulfate-derived tetrasaccharide by nmr, molecular modeling, and residual dipolar couplings. *Chembiochem* 9, 240–252. doi: 10.1002/cbic.200700400
- Skidmore, M. A., Guimond, S. E., Rudd, T. R., Fernig, D. G., Turnbull, J. E., and Yates, E. A. (2009). The activities of heparan sulfate and its analogue heparin are dictated by biosynthesis, sequence, and conformation. *Connective Tissue Res.* 49, 140–144. doi: 10.1080/0308200802148595
- Soares, P. A. G., Queiroz, I. N. L., and Pomin, V. H. (2017). NMR structural biology of sulfated glycans. *J. Biomol. Struct. Dynamics* 35, 1069–1084. doi: 10.1080/07391102.2016.1171165
- Solera, C., Macchione, G., Maza, S., Kayser, M. M., Corzana, F., de Paz, J. L., et al. (2016). Chondroitin sulfate tetrasaccharides: synthesis, three-dimensional structure and interaction with midkine. *Chem. Eur. J.* 22, 2356–2369. doi: 10.1002/chem.201504440
- Srb, P., Svoboda, M., Benda, L., Lepšík, M., Tarábek, J., Šícha, V., et al. (2019). Capturing a dynamically interacting inhibitor by paramagnetic NMR spectroscopy. *Phys. Chem. Chem. Phys.* 21, 5661–5673. doi: 10.1039/C9CP00416E
- Stancanelli, E., Elli, S., Hsieh, P., Liu, J., and Guerrini, M. (2018). Recognition and conformational properties of an alternative antithrombin binding sequence obtained by chemoenzymatic synthesis. *Chem. BioChem* 19, 1178–1188. doi: 10.1002/cbic.201800095
- Stewart, K. L., Hughes, E., Yates, E. A., Akien, G. R., Huang, T., Lima, M. A., et al. (2016). Atomic details of the interactions of glycosaminoglycans with amyloid- $\beta$  fibrils. *J. Am. Chem. Soc.* 138, 8328–8331. doi: 10.1021/jacs.6b02816
- Takeda, M., Ogino, S., Umamoto, R., Sakakura, M., Kajiwara, M., Sugahara, K. N., et al. (2006). Ligand-induced structural changes of the CD44 hyaluronan-binding domain revealed by NMR. *J. Biol. Chem.* 281, 40089–40095. doi: 10.1074/jbc.M608425200
- Takeda, M., Terasawa, H., Sakakura, M., Yamaguchi, Y., Kajiwara, M., Kawashima, H., et al. (2003). Hyaluronan recognition mode of CD44 revealed by cross-saturation and chemical shift perturbation experiments. *J. Biol. Chem.* 278, 43550–43555. doi: 10.1074/jbc.M308199200
- Taweechat, P., Pandey, R. B., and Sompornpisut, P. (2020). Conformation, flexibility and hydration of hyaluronic acid by molecular dynamics simulations. *Carbohydrate Res.* 493:108026. doi: 10.1016/j.carres.2020.108026
- Teriete, P., Banerji, S., Noble, M., Blundell, C. D., Wright, A. J., Pickford, A. R., et al. (2004). Structure of the regulatory hyaluronan binding domain in the inflammatory leukocyte homing receptor CD44. *Mol. Cell.* 13, 483–496. doi: 10.1016/s1097-2765(04)00080-2
- Theocharis, A. D., Skandalis, S. S., Gialeli, C., and Karamanos, N. K. (2016). Extracellular matrix structure. *Adv. Drug Delivery Rev.* 97, 4–27. doi: 10.1016/j.addr.2015.11.001
- Thunberg, L., Bäckström, G., and Lindahl, U. (1982). Further characterization of the antithrombin-binding sequence in heparin. *Carbohydrate Res.* 100, 393–410. doi: 10.1016/S0008-6215(00)81050-2
- Tjong, S., Chen, T., Huang, W., and Wu, W. (2007). Structures of heparin-derived tetrasaccharide bound to cobra cardiotoxins: heparin binding at a single protein site with diverse side chain interactions<sup>†</sup>. *Biochemistry* 46, 9941–9952. doi: 10.1021/bi700995v
- Uhl, F. E., Zhang, F., Pouliot, R. A., Uriarte, J. J., Rolandsson, E. S., Han, X., et al. (2020). Functional role of glycosaminoglycans in decellularized lung extracellular matrix. *Acta Biomater* 102, 231–246. doi: 10.1016/j.actbio.2019.11.029
- van Boeckel, C. A., Grootenhuis, P. D., and Visser, A. (1994). A mechanism for heparin-induced potentiation of antithrombin III. *Nat. Struct. Biol.* 1:423. doi: 10.1038/nsb0794-423
- Vanhaverbeke, C., Simorre, J., Sadir, R., Gans, P., and Lortat-Jacob, H. (2004). NMR characterization of the interaction between the C-terminal domain of interferon-gamma and heparin-derived oligosaccharides. *Biochem. J.* 384, 93–99. doi: 10.1042/BJ20040757
- Vasconcelos, A., and Pomin, V. (2017). The sea as a rich source of structurally unique glycosaminoglycans and mimetics. *Microorganisms* 5:51. doi: 10.3390/microorganisms5030051
- Vigetti, D., Passi, A., Luca, G. D., Caon, I., Moretto, P., Angelo, M. L., et al. (2016). Extracellular matrix in atherosclerosis: hyaluronan and proteoglycans insights. *Curr. Med. Chem.* 23, 2958–2971. doi: 10.2174/0929867323666160607104602
- Vignovich, W. P., and Pomin, V. H. (2020). Saturation transfer difference in characterization of glycosaminoglycan-protein interactions. *SLAS Technol.* 25, 307–319. doi: 10.1177/2472630320921130
- Wang, X., Watson, C., Sharp, J. S., Handel, T. M., and Prestegard, J. H. (2011). Oligomeric structure of the chemokine CCL5/RANTES from NMR, MS, and SAXS data. *Structure* 19, 1138–1148. doi: 10.1016/j.str.2011.06.001
- Weckbach, L., Preissner, K., and Deindl, E. (2018). The role of midkine in arteriogenesis, involving mechanosensing, endothelial cell proliferation, and vasodilation. *Int. J. Mol. Sci.* 19:2559. doi: 10.3390/ijms19092559
- Whitmore, E. K., Martin, D., and Guvench, O. (2020). Constructing 3-dimensional atomic-resolution models of nonsulfated glycosaminoglycans with arbitrary lengths using conformations from molecular dynamics. *Int. J. Mol. Sci.* 21:7699. doi: 10.3390/ijms21207699
- Xiao, C., Lian, W., Zhou, L., Gao, N., Xu, L., Chen, J., et al. (2016). Interactions between depolymerized fucosylated glycosaminoglycan and coagulation proteases or inhibitors. *Thrombosis Res.* 146, 59–68. doi: 10.1016/j.thromres.2016.08.027
- Yang, J., and Chi, L. (2017). Characterization of structural motifs for interactions between glycosaminoglycans and proteins. *Carbohydrate Res.* 452, 54–63. doi: 10.1016/j.carres.2017.10.008
- Yang, J., Shen, M., Wen, H., Luo, Y., Huang, R., Rong, L., et al. (2020). Recent advance in delivery system and tissue engineering applications of chondroitin sulfate. *Carbohydrate Polymers* 230:115650. doi: 10.1016/j.carbpol.2019.115650
- Ye, S., Luo, Y., Lu, W., Jones, R. B., Linhardt, R. J., Capila, I., et al. (2001). Structural basis for interaction of FGF-1, FGF-2, and FGF-7 with different heparan sulfate motifs. *Biochemistry* 40, 14429–14439. doi: 10.1021/bi011000u
- Yu, F., Roy, S., Arevalo, E., Schaeck, J., Wang, J., Holte, K., et al. (2014). Characterization of heparin-protein interaction by saturation transfer difference (STD) NMR. *Analytical Bioanalytical Chem.* 406, 3079–3089. doi: 10.1007/s00216-014-7729-4
- Zhao, B., and LiWang, P. J. (2010). Characterization of the interactions of vMIP-II, and a dimeric variant of vMIP-II, with glycosaminoglycans. *Biochemistry* 49, 7012–7022. doi: 10.1021/bi100549y
- Zhao, J., Huvent, I., Lippens, G., Eliez, D., Zhang, A., Li, Q., et al. (2017). Glycan determinants of heparin-tau interaction. *Biophys. J.* 112, 921–932. doi: 10.1016/j.bpj.2017.01.024
- Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G. M., Arakawa, T., et al. (1991). Three-dimensional structures of acidic and basic fibroblast growth factors. *Science* 251, 90–93. doi: 10.1126/science.1702556
- Zhu, Y., Zhang, F., and Linhardt, R. J. (2019). “Heparin Contamination and Issues Related to Raw Materials and Controls,” in *The Science and Regulations of Naturally Derived Complex Drugs*, eds R. Sasisekharan, S. L. Lee, A. Rosenberg,

- and L. A. Walker (Cham: Springer International Publishing), 191–206. doi: 10.1007/978-3-030-11751-1\_11
- Ziarek, J. J., Veldkamp, C. T., Zhang, F., Murray, N. J., Kartz, G. A., Liang, X., et al. (2013). Heparin oligosaccharides inhibit chemokine (CXC Motif) ligand 12 (CXCL12) cardioprotection by binding orthogonal to the dimerization interface, promoting oligomerization, and competing with the chemokine (CXC Motif) receptor 4 (CXCR4) N terminus. *J. Biol. Chem.* 288, 737–746. doi: 10.1074/jbc.M112.394064
- Ziebell, M. R., and Prestwich, G. D. (2004). Interactions of peptide mimics of hyaluronic acid with the receptor for hyaluronan mediated motility (RHAMM). *J. Comp. Aided Mol. Design* 18, 597–614. doi: 10.1007/s10822-004-5433-8
- Zsila, F., Juhász, T., Kohut, G., and Beke-Somfai, T. (2018). Heparin and heparan sulfate binding of the antiparasitic drug imidocarb: circular dichroism spectroscopy, isothermal titration calorimetry, and computational studies. *J. Phys. Chem. B* 122, 1781–1791. doi: 10.1021/acs.jpcc.7b08876
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2021 Bu and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Macrophages Inhibit Ciliary Protein Levels by Secreting BMP-2 Leading to Airway Epithelial Remodeling Under Cigarette Smoke Exposure

Zhigang Wang<sup>1,2†</sup>, Wenzhang Liang<sup>1†</sup>, Cuiqing Ma<sup>1</sup>, Jiachao Wang<sup>1</sup>, Xue Gao<sup>1</sup> and Lin Wei<sup>1\*</sup>

<sup>1</sup> Department of Immunology, Key Laboratory of Immune Mechanism and Intervention on Serious Disease in Hebei Province, Hebei Medical University, Shijiazhuang, China, <sup>2</sup> Department of Intensive Care Unit, Hebei General Hospital, Shijiazhuang, China

## OPEN ACCESS

### Edited by:

Qunye Zhang,  
Shandong University, China

### Reviewed by:

Kurt Henry Piepenbrink,  
University of Nebraska-Lincoln,  
United States  
M. Gabriela Kramer,  
Universidad de la República, Uruguay

### \*Correspondence:

Lin Wei  
weilin@hebmh.edu.cn

<sup>†</sup> These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 04 February 2021

**Accepted:** 29 March 2021

**Published:** 26 April 2021

### Citation:

Wang Z, Liang W, Ma C, Wang J,  
Gao X and Wei L (2021)  
Macrophages Inhibit Ciliary Protein  
Levels by Secreting BMP-2 Leading  
to Airway Epithelial Remodeling Under  
Cigarette Smoke Exposure.  
Front. Mol. Biosci. 8:663987.  
doi: 10.3389/fmolb.2021.663987

Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease with high morbidity and mortality worldwide. So far, smoking is still its leading cause. The characteristics of COPD are emphysema and airway remodeling, as well as chronic inflammation, which were predominated by macrophages. Some studies have reported that macrophages were involved in emphysema and chronic inflammation, but whether there is a link between airway remodeling and macrophages remains unclear. In this study, we found that both acute and chronic cigarette smoke exposure led to an increase of macrophages in the lung and a decrease of ciliated cells in the airway epithelium of a mouse model. The results of *in vitro* experiments showed that the ciliary protein ( $\beta$ -tubulin-IV) levels of BEAS-2B cells could be inhibited when co-cultured with human macrophage line THP-1, and the inhibitory effect was augmented with the stimulation of cigarette smoke extract (CSE). Based on the results of transcriptome sequencing, we focused on the protein, bone morphogenetic protein-2 (BMP-2), secreted by the macrophage, which might mediate this inhibitory effect. Further studies confirmed that BMP-2 protein inhibited  $\beta$ -tubulin-IV protein levels of BEAS-2B cells under the stimulation of CSE. Coincidentally, this inhibitory effect could be nearly blocked by the BMP receptor inhibitor, LDN, or could be interfered with BMP-2 siRNA. This study suggests that activation and infiltration of macrophages in the lung induced by smoke exposure lead to a high expression of BMP-2, which in turn inhibits the ciliary protein levels of the bronchial epithelial cells, contributing to the remodeling of airway epithelium, and aggravates the development of COPD.

**Keywords:** COPD, cigarette smoking, macrophage, bronchial epithelial cell, ciliary protein

## INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease with high morbidity and mortality worldwide. The 2015 Global Burden of Disease (GBD) Study estimated the global morbidity of COPD to be about 174 million (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016), and it has the third ranking mortality after ischemic heart disease and cerebrovascular disease, that is, about 3.2 million deaths per year in 2015 (GBD 2015 Mortality and Cause of Death Collaborators, 2016). COPD has imposed a heavy global burden and will continue to increase in the future because of the

aging population and the persisting air pollution (Mathers and Loncar, 2006). Smoking is still the leading cause of COPD (Mannino and Buist, 2007), although the proportion of male smokers decreased by 28% and that of female smokers decreased by 29% from 1990 to 2015. The WHO data showed that there were still about 1.1 billion people with a smoking habit in 2015 (GBD 2015 Risk Factors Collaborators., 2016), and its consequence would show up in the following decades.

One characteristic of COPD is emphysema (Vogelmeier et al., 2017). Most of the previous studies have been focused on the pathogenesis of emphysema, including gene precondition, early life events (Martinez, 2016), the imbalance of proteolysis/anti-proteolysis (Agusti and Hogg, 2019) and oxidation/anti-oxidation, and an enhanced apoptosis (Tuder et al., 2003). Although the exact mechanism is still unclear, there is a consensus that unquenched chronic inflammation leads to the disease (Barnes, 2013, 2014). Inflammation is characterized by the infiltration of neutrophils, macrophages, and lymphocytes in the airways and lung parenchyma (Hogg et al., 2004; Brusselle et al., 2011; Faner et al., 2013). These inflammatory cells secrete a variety of proteinases, such as, neutrophils elastase (Hunninghake and Crystal, 1983; Ghosh et al., 2019), granulase (Ngan et al., 2009; Kim et al., 2013), matrix metalloproteinase (Ghosh et al., 2019; Jeon et al., 2019), and perforin (Morissette et al., 2007; Zhang et al., 2014), which break down the extracellular matrix, such as collagen and elastin, causing structural damage to the alveoli and small airways, thus leading to emphysema.

Another consequence of COPD is remodeling of the airway wall (Hogg et al., 2004), showing squamous metaplasia, hypertrophy of submucosal glands, hyperplasia of smooth muscle, and fibrosis of adventitia in the airway. Normal airway epithelium consists of basal cells, ciliated cells, secretory cells (goblet cells, plasma cells, and Clara cells), neuroendocrine cells, and a few unclassified or intermediate cells (Mercer et al., 1994). Among them, ciliated cells are the main cell type (accounting for 50–70%) in the human airway (Boers et al., 1998; Montoro et al., 2018). They play a central role in the mucociliary clearance (MCC) function of the lung to get rid of the inhaled xenobiotics. Decreased ciliated cells, shorter cilia, and uncoordinated cilia beating frequency take a large part in the impaired MCC of patients with COPD (Randell, 2006; Rock et al., 2010).

Emphysema and airway remodeling of COPD are caused not only by cigarette smoking itself but also by the inflammation involved. Macrophages are the most prominent inflammatory cells in patients with COPD. Most previous studies reported that macrophage was related to emphysema (Morris et al., 2003; Hume et al., 2020; Xia et al., 2021), while few groups have reported that it was involved in airway remodeling (Ferhani et al., 2010; Bu et al., 2020), and its effect on cilia remains unclear. This study focused on the effect of macrophages on cilia in the airway epithelium of mice and ciliary protein levels of bronchial epithelial cells after exposure to cigarette smoke (CS) and its possible mechanism.

We used a whole-body CS exposure to induce COPD in mice. *In vivo*, we found that both acute and chronic CS exposure in mice could lead to an increase of macrophages in the lung and

a decrease of ciliated cells in the airway epithelium. An *in vitro* study showed that THP-1 cells could inhibit  $\beta$ -tubulin-IV levels of BEAS-2B cells under the stimulation of CS extract (CSE). Further study confirmed that bone morphogenetic protein-2 (BMP-2) secreted by macrophages was responsible for that inhibitory effect.

## MATERIALS AND METHODS

### Mice and CS Exposure

All C57BL/6N mice (6-week-old male) were purchased from Beijing Weitonglihua Laboratory Animal Technology Co., Ltd. The number of the Laboratory Animal Quality Certificate is 11400700223102. All experimental procedures were performed in compliance with the Institutional Animal Welfare Guidelines and were carried out according to the criteria outlined in the Guide for the Care and Use of Laboratory Animals [National Research Council (US) Committee for the Update of the Guide for the Care, and Use of Laboratory Animals, 2011] and with the approval of the Animal Care and Use Committee of Hebei Medical University. The mice were maintained in an animal facility under a 12-h light/dark cycle and were fed standard chow and sterile tap water. After raising them for 2 weeks for adaptation, the mice were divided into control (normal group) and experimental (CS group) groups. The experimental groups were exposed to Hongmei brand CS (tar oil 15 mg, nicotine 1.2 mg, produced by Yunnan Kunming Cigarette Factory, China). The cigarettes were burned in the combustion chamber, and then CS and fresh air were blown into the exposure chamber with a flow rate of 1:2. Acute CS exposure model lighting up two cigarettes every 20 min for five times every morning and every afternoon, respectively, with 6-h intervals, 5 days a week for 4 weeks. Chronic CS exposure model is a similar method to the acute exposure model: light up one cigarette every 20 min and the exposure time lasted for 16 weeks.

### Specimen Tissue Acquisition

Mice were anesthetized by intraperitoneal injection of ketamine (90 mg/kg) and xylazine (9 mg/kg) and were fixed on the animal operating table. The chest and the left ventricle were cut open, and blood was let out. In order to estimate the blood cells in the pulmonary vascular system, the lung was flushed by saline through the pulmonary artery until it became white. The airway and the lung were carefully separated, and then, they were processed differently. Some of the lungs were inflated and fixed for 6 h through intratracheal instillation of 4% paraformaldehyde under 20 cm H<sub>2</sub>O pressure. The trachea and the lung were separately embedded with paraffin and then sectioned. The rest of the fresh lungs were ground gently, sieved with a 70- $\mu$ m nylon BD Falcon cell strainer (BD Biosciences, San Jose, CA, United States), washed with sterile phosphate-buffered saline (PBS) (Gibco-BRL, Gaithersburg, MD, United States) without Ca<sup>2+</sup> and Mg<sup>2+</sup> supplemented with 2 mM EDTA (Sigma-Aldrich, St. Louis, MO, United States), and then suspended in RPMI 1640 culture medium to form single-cell suspension. The rest fresh tracheae were put into liquid nitrogen, and then they were ground to extract the tissue RNA and protein.



## Fluorescence-Activated Cell Sorting (FACS)

The single-cell suspensions were labeled with 5  $\mu$ l of PE anti-mouse F4/80 and 1.25  $\mu$ l of APC anti-mouse/human CD11b as markers for macrophages or 5  $\mu$ l of PE Rat IgG2a and 1.25  $\mu$ l of APC Rat IgG2b as markers for control for 45 min in PBS on ice. All these antibodies were purchased from BioLegend, San Diego, CA, United States. Propidium iodide was added to exclude dead cells, and FACS was performed in the BD flow cytometry facility;  $3 \times 10^4$  events were recorded.

## Reverse Transcription PCR (RT-PCR)

RNA was extracted with the RNeasy micro kit (QIAGEN, Beijing, China), and cDNA was synthesized with PrimeScript IV 1st strand cDNA Synthesis Mix (Tarkara, Kusatsu, Japan). PCR was performed with SYBR green chemistry in a Step One Plus (Applied Biosystems, Grand Island, NY, United States), and data were analyzed using the  $2^{-\Delta\Delta C_t}$  method. The primers shown in Table 1 were synthesized by Invitrogen.

## Western Blot

For airway analysis, the tracheae of the mice were homogenized in RIPA buffer (1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS in PBS) containing phosphatase inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, United States). BEAS-2B cells were collected in cell lysis buffer containing 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton-X 100, 1.0  $\mu$ g/ml leupeptin, 10  $\mu$ g/ml aprotinin, 0.2 mM phenylmethylsulfonyl fluoride, 1 mM sodium orthovanadate, 0.1 mM sodium fluoride, 2.5 mM sodium pyrophosphate, and 1 mM  $\beta$ -glycerophosphate. Tissue lysates or cell lysates were centrifuged, and supernatant proteins were separated on 10% gradient SDS-PAGE and transferred to PVDF membranes (Millipore, Billerica, MA, United States). The membranes were blotted against antibodies to  $\beta$ -tubulin-IV (Abcam, Cambridge, MA, United States), and  $\beta$ -actin (Cell Signaling Technology, Danvers, MA, United States) or GAPDH (Cell Signaling Technology, Danvers, MA, United States). Primary antibody binding was detected with secondary antibodies conjugated with horseradish peroxidase and enhanced chemiluminescence (Amersham Pharmacia Biotech, Amersham, United Kingdom).

## Immunohistochemistry

Sections were deparaffinized, rehydrated, and subjected to antigen retrieval by autoclaving (10 min, 120°C, 30 psi) for

10 min in the citrate target retrieval solution. Subsequently, endogenous peroxidase was quenched with 3%  $H_2O_2$  and blocked for 20 min with 10% goat serum. Primary rabbit anti-mouse CD68 polyclonal antibody (Abclonal, Wuhan, China) was added overnight at 4°C in 10% BSA-PBS (1:200). Sections were washed with PBS and then incubated with a biotinylated goat anti-rabbit secondary antibody (1:100) for 60 min followed by a 15-min treatment with streptavidin-horseradish peroxidase (Dako, Glostrup, Denmark). The antigen of interest was visualized using the brown chromogen 3,3'-diaminobenzidine (Dako, Glostrup, Denmark) and counterstained with Harris' hematoxylin solution (Sigma-Aldrich, St. Louis, MO, United States). Sections were then dehydrated and mounted with Cytoseal 60 (Richard-Allan Scientific, Kalamazoo, MI, United States). Antibody dilutions and all washes were immersed in Tris-buffered saline solution. The section was scanned by TissueFAXS and analyzed by TissueFAXS Cytometry.

## Immunofluorescent Staining

Sections were deparaffinized, rehydrated, and subjected to antigen retrieval by autoclaving (10 min, 120°C, 30 psi) for 10 min in the citrate target retrieval solution. Primary rabbit anti- $\beta$ -tubulin-IV antibody (Abcam) was added overnight at 4°C in 10% BSA-PBS (1:100). Sections were washed with PBS and then were incubated in dark with FITC-labeled goat anti-rabbit IgG (KPL, Gaithersburg, MD, United States) for 2 h at room temperature. Sections were washed three times in dark with PBS and sealed with ProLong Gold Antifade Reagent with DAPI (Cell Signaling Technology, Danvers, MA, United States). The section was scanned by Tissue FAXS and analyzed with StrataQuest (Tissue Gnostics GmbH, Vienna, Austria) and Tissue quest software.

## Cigarette Smoke Extract Preparation

Cigarette smoke extract was prepared by bubbling the CS from one commercially available cigarette (Hongmei, China; tar oil 15 mg, nicotine 1.2 mg) into 4 mL of RPMI 1640 medium (Life Technologies, Carlsbad, CA, United States), containing no serum or growth factors, using a modified method described previously. This was considered as a 100% CSE solution and then sterilized and stored at -80°C (Xu et al., 2012; Kratzer et al., 2013a; Sakhatysky et al., 2014).

## Cell Lines, Culture Media, and Growth Conditions

The BEAS-2B, simian virus 40-transformed, immortalized bronchial epithelial cell line used for this study (Reddel et al., 1988) was preserved in our laboratory. This cell line has been cultured continuously for >100 passages. In this study, passages 55 and 65 were used. The cells were cultured in extracellular matrix (ECM)-coated dishes in RPMI 1640 medium supplemented with 10% heat-inactivated-FCS, 100 U/mL penicillin, and 100 mg/mL streptomycin (Gibco, Carlsbad, CA, United States) at 37°C in a humidified atmosphere of 5%  $CO_2$  in air. ECM-coated dishes were prepared by incubating wells with 1 ml of 50  $\mu$ g/ml of collagen IV (Sigma, St. Louis,

TABLE 1 | Primers used for RT-qPCR.

Primer	Sequence
ACTB sense	5'-GTTG GTTG GAGC AAAC ATCC C-3'
ACTB antisense	5'-TTAG GAGT GGGG GTGG CTTT-3'
Foxj1 sense	5'-GGGT CGCA GAAT GGAA GTGA-3'
Foxj1 antisense	5'-GAGC CTTG GCGT TGAG AATG-3'
GAPDH sense	5'-CCTC TGAC TTCA ACAG CGAC AC-3'
GAPDH antisense	5'-CACC ACCC TGTT GCTG TAGC CA-3'
BMP-2 sense	5'-CTGC GGTC TCCT AAAG GTCG-3'
BMP-2 antisense	5'-GGGG TGGG TCTC TGTT TGAG-3'

MO, United States) per well overnight and washed with sterile PBS before cells were seeded;  $1 \times 10^3$  cells were seeded in each well overnight, then cultured with the stimulation of CSE and/or recombinant BMP2 protein (PeproTech, Rocky Hill, NJ, United States) for 48 h or pre-treated with LDN193189 (MCE, Monmouth, NJ, United States) for 24 h, and then co-cultured with  $5 \times 10^5$  THP-1 cells in Transwell with the stimulation of CSE for 48 h.

Human monocytic leukemia cell line THP-1 (American Type Culture Collection, Manassas, VA, United States) cells were maintained in RPMI 1640 media with the addition of 10% (v/v) heat-inactivated-FBS at 37°C in a 5% CO<sub>2</sub> humidified atmosphere. Cells were seeded at  $5 \times 10^5$  cells per well on Transwell filters and treated with phorbol myristate acetate (PMA 30 ng/ml, Multi Sciences, China) for 24 h to induce differentiation and filter attachment, and co-cultured with Beas-2B cells or pre-incubated with siRNA for 24 h, and then co-cultured with Beas-2B cells with the stimulation of CSE.

## Co-culture

BEAS-2B cells were seeded at  $1 \times 10^3$  cells in 6-well cell culture plates for 24 h. At the same time, THP-1 cells were seeded at  $5 \times 10^5$  cells per well on Transwell filters and treated with PMA for 24 h. The next day, the Transwell inserts were placed in the 6-well cell culture plate and were stimulated with CSE for 48 h before harvesting.

## RNA-Seq Analysis

Total RNA was isolated from THP-1 cells in three different conditions, including THP-1 cells activated by PMA (group 1), THP-1 cells co-cultured with BEAS-2B cells (group 2), and THP-1 cells co-cultured with BEAS-2B cells and stimulated by CSE (group 3). Construction of the cDNA library and sequencing were performed by Sinotech Genomics, ShanHai, China using the Illumina Novaseq 6000 sequencing platform. High-quality reads were mapped with Homo sapiens GRCh38, using Hisat2 version 2.0.4. The expression level of each gene was standardized to fragments per kilobase of exon model per million mapped reads (FPKM) using StringTie version 1.3.0 and trimmed mean of *M* values (TMM) (*Q* value < 0.05, fold change  $\geq 2$  times).

## SiRNA

THP-1 cells were seeded at  $5 \times 10^5$  cells per well on Transwell filters and treated with PMA for 24 h; siBMP2 was transfected into cells with Lipofectamine 2000 (Invitrogen, Carlsbad, CA, United States) following the instructions of the manufacturer for 24 h and co-cultured with Beas-2B cells for 48 h with the stimulation of CSE. The sequence of siBMP2 (Ribobiotech, GuangZhou, China) was listed as:

sense: 5'UCAACUCUGUUAACUCUAA3'  
antisense: 5'UUAGAGUUAACAGAGUUGA3'

## Statistical Analysis

SPSS statistical software (version 16.0) was used for statistical analysis. Data were expressed as mean plus SD. The significance between the two groups is determined using *t*-test. \**P* < 0.05

was considered statistically significant. All experiments were performed at least three times.

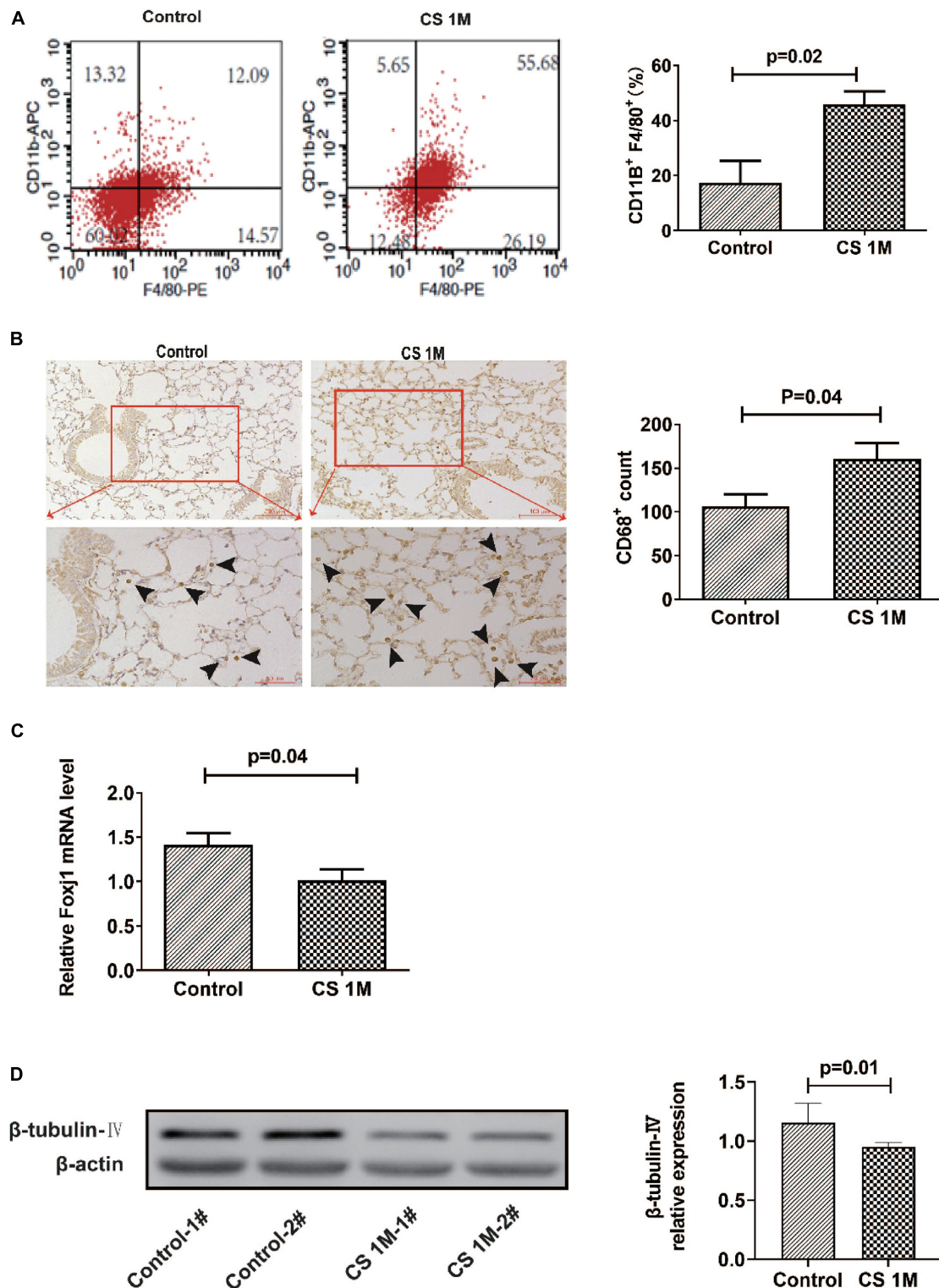
## RESULTS

### Acute CS Exposure Led to an Increase of Macrophages in the Lung and a Decrease of $\beta$ -Tubulin-IV Levels in the Airway

In our experiment, we first explored the effect of acute high-dose CS exposure on macrophages in the lung and ciliary protein levels in the airway of mice. We took CS exposure (20 cigarettes per day) for 4 weeks as the acute exposure condition because the restoration of epithelial histology is about 2 weeks after its extensive damage caused by inhalation of SO<sub>2</sub> in mice (Rawlins and Hogan, 2008). After acute CS exposure, mice were anesthetized and sacrificed, the lungs were freshly made into single-cell suspension for FACS or were fixed and sectioned for immunohistochemistry (IHC), and the airways were homogenized to extract RNA and protein for RT-PCT and Western blot. The results showed that the percentage of CD11b<sup>+</sup>F4/80<sup>+</sup> macrophages detected by FACS increased significantly in the CS group ( $45.89 \pm 4.73\%$ ) compared to the control group ( $17.43 \pm 7.90\%$ ) (Figure 1A). Also, the number of CD68<sup>+</sup> macrophages in the lung section detected by IHC staining also increased evidently from 106.3 per 10 random HPFs (high power fields) in the control group to 160.2 in the CS group (Figure 1B). After acute CS exposure, the expression of both foxj1 mRNA and  $\beta$ -tubulin-IV protein of the airway decreased significantly in the CS group compared to that of the control group (Figures 1C,D).

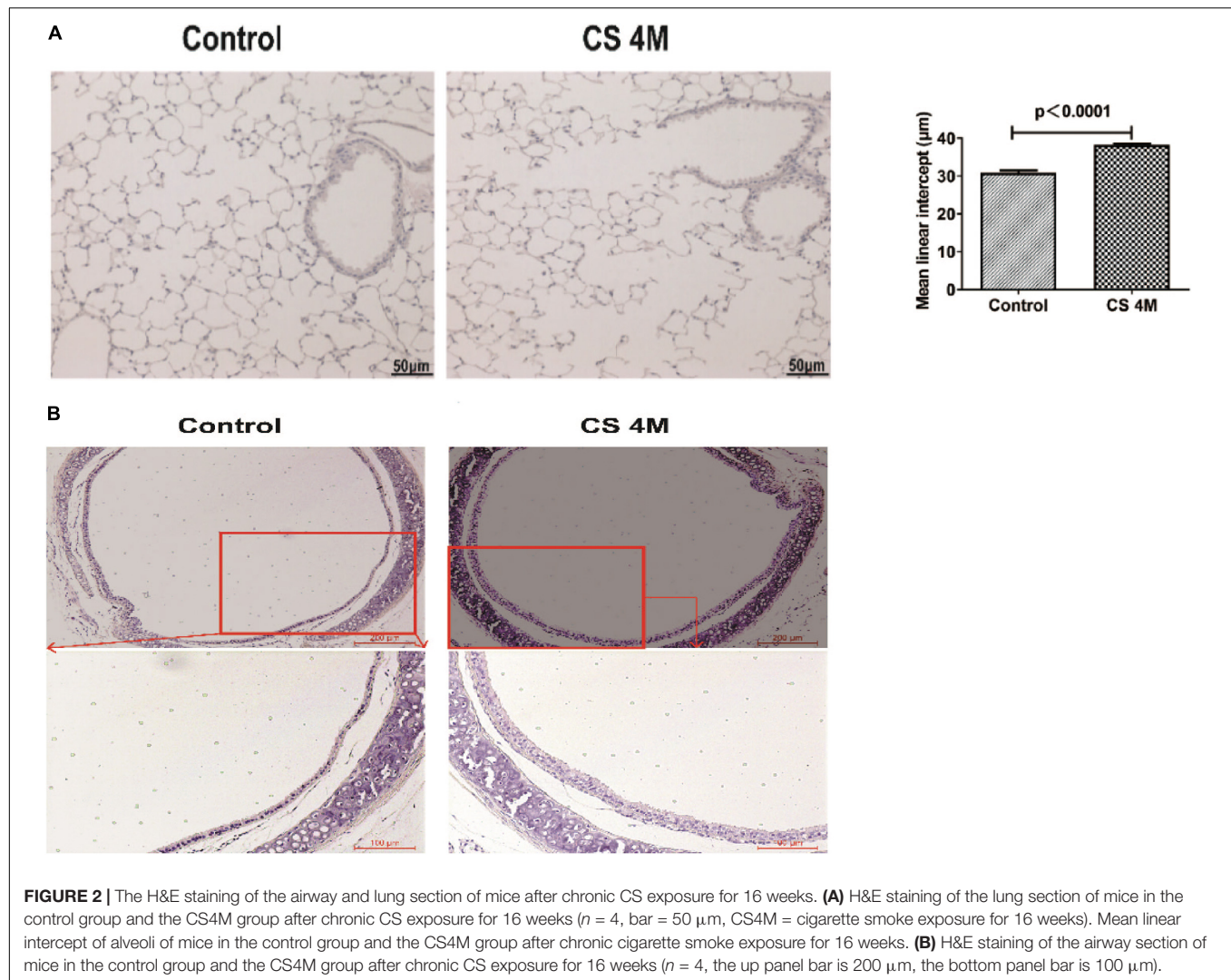
### Chronic CS Exposure Resulted in the Development of Emphysema and Remodeling of the Airway Wall in Mice

According to the commonly used methods in a previous study, animal models usually took 4–6 months to display signs of disease (Ofulue and Ko, 1999; Kratzer et al., 2013b). Different from the acute exposure condition, mice were exposed to CS (10 cigarettes per day) for 16 weeks to induce emphysema in chronic exposure mode and were successfully induced for the development of emphysema in the CS group. H&E staining of lung sections showed that, in the CS group, most of the integrity of small airways was destroyed, the alveolar structure was seriously damaged, alveolar fusion was evident, and the mean alveolar intercept was  $38.20 \pm 0.40 \mu\text{m}$ , which was significantly enlarged compared to that of the control group, indicating that emphysema was successfully induced by long-term CS exposure. While, in the control group, small airway integrity was retained and just a small amount of alveolar destruction was observed, alveolar fusion was mild, and the mean alveolar intercept was  $30.34 \pm 0.44 \mu\text{m}$  (Figure 2A). As we know, apart from emphysema, the other characteristic of COPD is remodeling of the airway wall. Then, we examined the airway histological profile of mice. The results showed that the airway epithelium of



**FIGURE 1 |** The changes of macrophages in the lung and  $\beta$ -tubulin-IV in the airway of mice after acute cigarette smoke (CS) exposure for 4 weeks. **(A)** The percentage of CD11b<sup>+</sup>F4/80<sup>+</sup> macrophages in the lung detected by fluorescence-activated cell sorting (FACS) in the control group and the CS1M group ( $n = 4$ , CS1M = cigarette smoke exposure for 4 weeks). **(B)** The CD68<sup>+</sup> macrophages in the lung detected by immunohistochemical (IHC) staining in the control group and the CS1M group (the up panel bar is 100  $\mu$ m, the bottom panel bar is 50  $\mu$ m, black arrow indicates CD68<sup>+</sup> macrophages,  $n = 6$ ). **(C)** The relative expression of foxj1 mRNA in the airway of mice was determined by qRT-PCR after acute CS exposure for 4 weeks ( $n = 6$ ). **(D)** The relative levels of  $\beta$ -tubulin-IV protein in the airway of mice were quantitated by densitometry and normalized to  $\beta$ -actin after acute CS exposure for 4 weeks ( $n = 6$ ).





the CS group was thicker than that of the control group, the cell layers were increased by three to four layers, and its arrangement was disordered in the CS group, while it was one or two aligned cell layers in the control group (Figure 2B). These epithelial histological changes indicated that there was the remodeling of the airway wall after chronic CS exposure for 16 weeks.

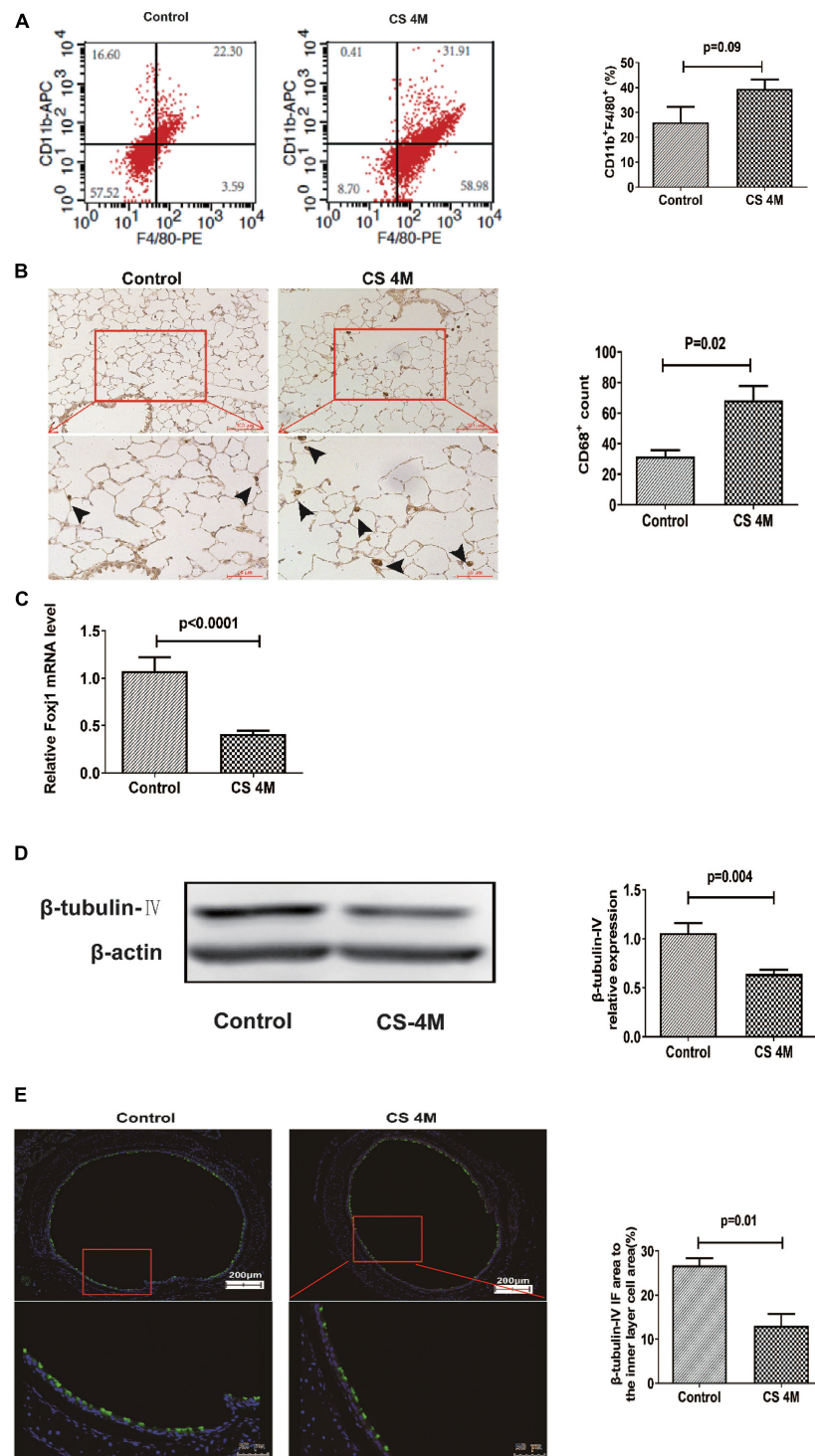
### Chronic CS Exposure Led to an Increase of Macrophages in the Lung and a Decrease of Ciliated Cells in the Airway of Mice

In this study, after chronic CS exposure for 16 weeks in the mouse model, macrophages increased in the CS group compared to the control group, but there was a little difference between different marker performances. The percentage of CD11b<sup>+</sup>F4/80<sup>+</sup> macrophages in the lung detected by FACS was  $39.49 \pm 3.70\%$  in the CS group and  $26.01 \pm 6.22\%$  in the control group. Although the number of macrophages in the CS group increased, there was no significant difference

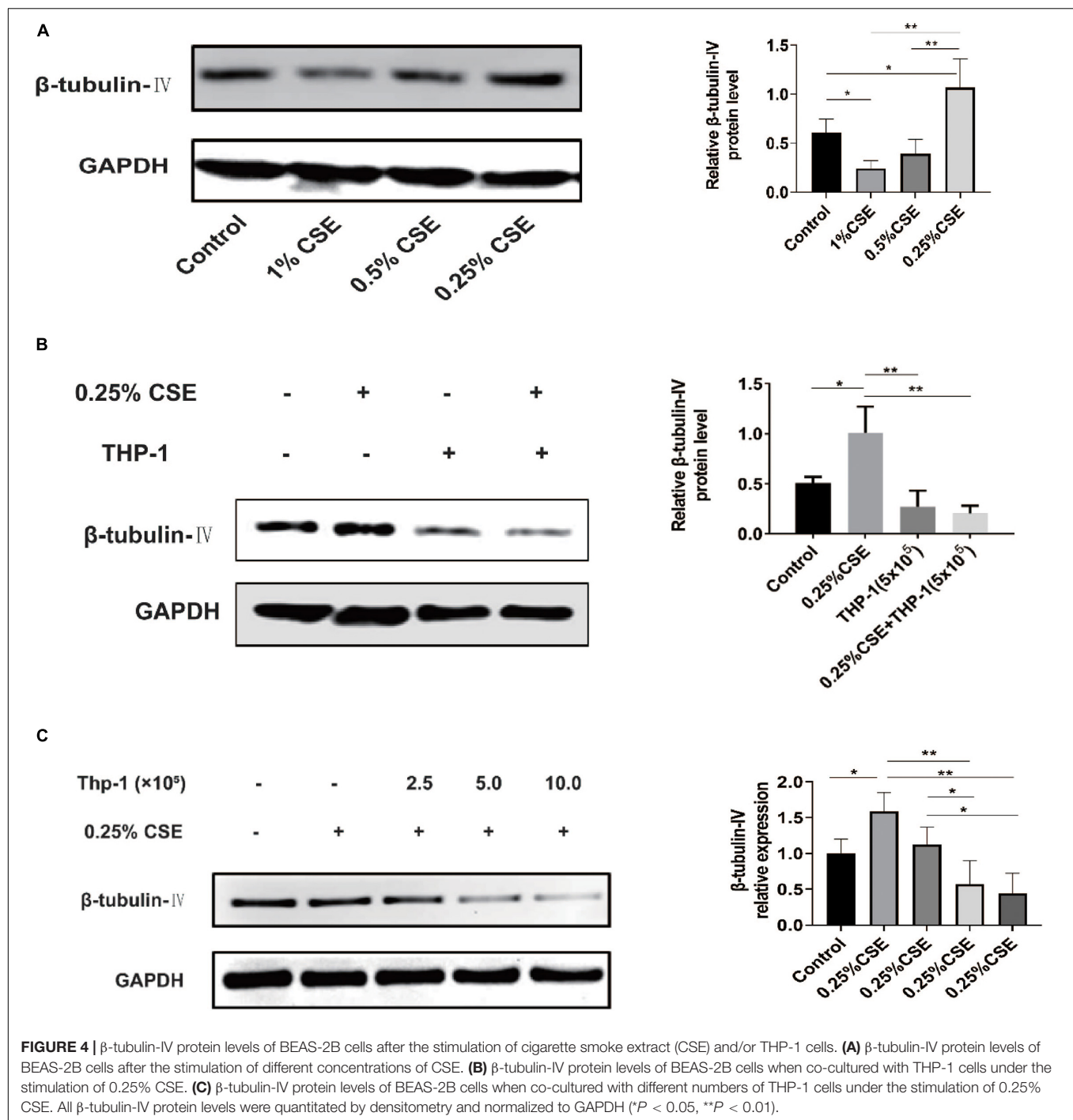
between the two groups (Figure 3A). However, the number of macrophages detected by IHC was significantly different between the two groups, from  $31.67 \pm 4.05$  per 10 random HPFs in the control group to  $68.50 \pm 9.37$  in the CS group (Figure 3B).

Although there was a slight difference in the macrophage performance, it was consistent with the change of ciliated cells in the airways after acute and chronic CS exposure. After chronic CS exposure, the expression of foxj1 mRNA detected by RT-PCR (Figure 3C) and  $\beta$ -tubulin-IV protein detected by Western blot (Figure 3D) in the airways both decreased significantly in the CS group compared with that in the control group. We also detected  $\beta$ -tubulin-IV protein of cilia in the epithelium by IF, and the images showed that the cilia dyed with green color were interspersed in the inner layer of the airway, its distribution was non-uniform, and the intensity was weakened or even lost in some area in the CS group, while it was uniform in the control group. The ratio of cilia area to the inner layer cell area was 15.6% in the CS group and 23.4% in the control group, and there was a significant difference between the two groups





**FIGURE 3 |** The changes of macrophages in the lung and  $\beta$ -tubulin-IV in the airway of mice after chronic CS exposure for 16 weeks. **(A)** The percentage of CD11b<sup>+</sup>F4/80<sup>+</sup> macrophages in the lung detected by FACS in the control group and the CS4M group ( $n = 4$ , CS4M = cigarette smoke exposure for 16 weeks). **(B)** The CD68<sup>+</sup> macrophages in the lung detected by IHC staining in the control group and the CS4M group (the upper panel bar is 100  $\mu$ m, the bottom panel bar is 50  $\mu$ m, black arrow indicates CD68<sup>+</sup> macrophages,  $n = 6$ ). **(C)** The relative expression of foxj1 mRNA in the airway of mice was determined by qRT-PCR after chronic CS exposure for 16 weeks ( $n = 6$ ). **(D)** The relative levels of  $\beta$ -tubulin-IV protein in the airway of mice were quantitated by densitometry and normalized to  $\beta$ -actin after chronic CS exposure for 16 weeks ( $n = 6$ ). **(E)** The  $\beta$ -tubulin-IV protein levels in the airway epithelium of mice stained by IF in the control group and the CS4M group (the upper panel bar is 200  $\mu$ m, the bottom panel bar is 50  $\mu$ m, nuclear was dyed with DAPI, and cilia was dyed with FITC,  $n = 4$ ). The ratio of cilia area to the inner layer cell area in the control group and the CS4M group.



(Figure 3E). These results indicated that CS exposure led to a decrease of ciliated cells.

### THP-1 Cells Inhibited β-Tubulin-IV Levels of BEAS-2B Cells Under the Stimulation of CSE

*In vivo*, we confirmed that acute and chronic CS exposure both lead to macrophage infiltration and ciliated cell reduction.

*In vitro*, we tried to explore whether there is some relationship between macrophages and ciliated cells. First, we tested the effect of CSE on β-tubulin-IV levels of bronchial epithelial cells. The results showed that the levels of β-tubulin-IV of BEAS-2B cells decreased with the stimulation of 1% CSE for 48 h, while they increased with 0.25% CSE stimulation for 48 h (Figure 4A). Next, we studied the effect of macrophages on β-tubulin-IV levels of bronchial epithelial cells. The results showed that THP-1 cells could inhibit β-tubulin-IV levels of BEAS-2B cells, and then when

0.25% CSE was added into the medium, the inhibitory effect was augmented (**Figure 4B**). Furthermore, this inhibitory effect was further enhanced with increased numbers of THP-1 cells from  $2.5 \times 10^5$  to  $10 \times 10^5$  (**Figure 4C**).

## Bone Morphogenetic Protein-2 Was Screened Out and Verified to Have an Inhibitory Effect on $\beta$ -Tubulin-IV Levels

Now that THP-1 cells could inhibit  $\beta$ -tubulin-IV levels of BEAS-2B cells when they were co-cultured in Transwell without direct contact, this inhibitory effect might come from certain soluble factors secreted by activated THP-1 cells. Transcriptome array technology was used to test RNA expression profiles of THP-1 cells with different conditions, namely, THP-1 cells (group 1), THP-1 cells co-cultured with BEAS-2B cells (group 2), and THP-1 cells co-cultured with BEAS-2B cells under the stimulation of CSE (group 3). Taking RNA differential expression of more than two-fold as standard, compared with group 1, the differentially expressed genes were 374 (172 upregulated and 202 downregulated) in group 2 and were 1,465 (594 upregulated and 871 downregulated) in group 3, which suggested that the effect of co-culture under the stimulation of CSE was far more complicated than that of co-culture alone (**Figure 5A**). We speculated that genes that have an inhibitory effect on the  $\beta$ -tubulin-IV protein levels of BEAS-2B should be soluble substances and should be upregulated in group 2 and group 3 at the same time. We noticed that many of these differentially expressed genes belonged to the TGF- $\beta$  superfamily, which was associated with airway epithelial cell differentiation. Eventually, BMP-2 was screened out as the target (**Figure 5B**).

In order to verify if BMP-2 could affect the  $\beta$ -tubulin-IV levels, different concentrations (25, 50, 100, and 200 ng/ml) of recombinant human BMP-2 protein was added into the medium when BEAS-2B cells were cultured under the stimulation of 0.25% CSE. Results showed that the  $\beta$ -tubulin-IV levels were inhibited in a concentration-dependent manner from 25 to 200 ng/ml. When the concentration reached 50 ng/ml,  $\beta$ -tubulin-IV levels were inhibited down below the normal level (**Figure 5C**), indicating that BMP-2 could inhibit the  $\beta$ -tubulin-IV levels of BEAS-2B cells. In the following experiments, the concentration of recombinant human BMP-2 protein was fixed to 50 ng/ml, because at this point, the  $\beta$ -tubulin-IV levels of BEAS-2B was not affected without the stimulation of CSE (**Figure 5D**).

To further verify that BMP-2 could inhibit the  $\beta$ -tubulin-IV levels of BEAS-2B cells, we interfered with the BMP signal using BMP-2 blockage LDN or reducing BMP-2 production with BMP-2 siRNA in THP-1 cells.

First, BEAS-2B cells were pre-treated with BMP blockage LDN for 24 h before CSE and recombinant human BMP-2 stimulation; the inhibitory effect of BMP-2 on  $\beta$ -tubulin-IV levels of BEAS-2B cells could be completely blocked at a very lower dose of LDN (5 nM), which was similar to the other experimental dosages of LDN from 10 to 100 nM (**Figure 5E**).

Then, we tested whether LDN would block the inhibitory effect of THP-1 cells on  $\beta$ -tubulin-IV levels of BEAS-2B cells. The results showed that the  $\beta$ -tubulin-IV levels of BEAS-2B cells

could be mostly restored when the BEAS-2B cells were pre-treated with LDN for 24 h before they were co-cultured with THP-1 cells and stimulated by CSE (**Figure 5F**).

Subsequently, we interfered THP-1 cells with BMP-2 siRNA, which turned out to decrease BMP-2 mRNA expression by over 70% (**Figure 5G**). The results showed that the inhibitory effect of THP-1 cells on  $\beta$ -tubulin-IV levels of BEAS-2B cells could also be mostly blocked when THP-1 cells interfered with BMP-2 siRNA for 24 h before they were co-cultured with BEAS-2B cells and stimulated by CSE (**Figure 5H**).

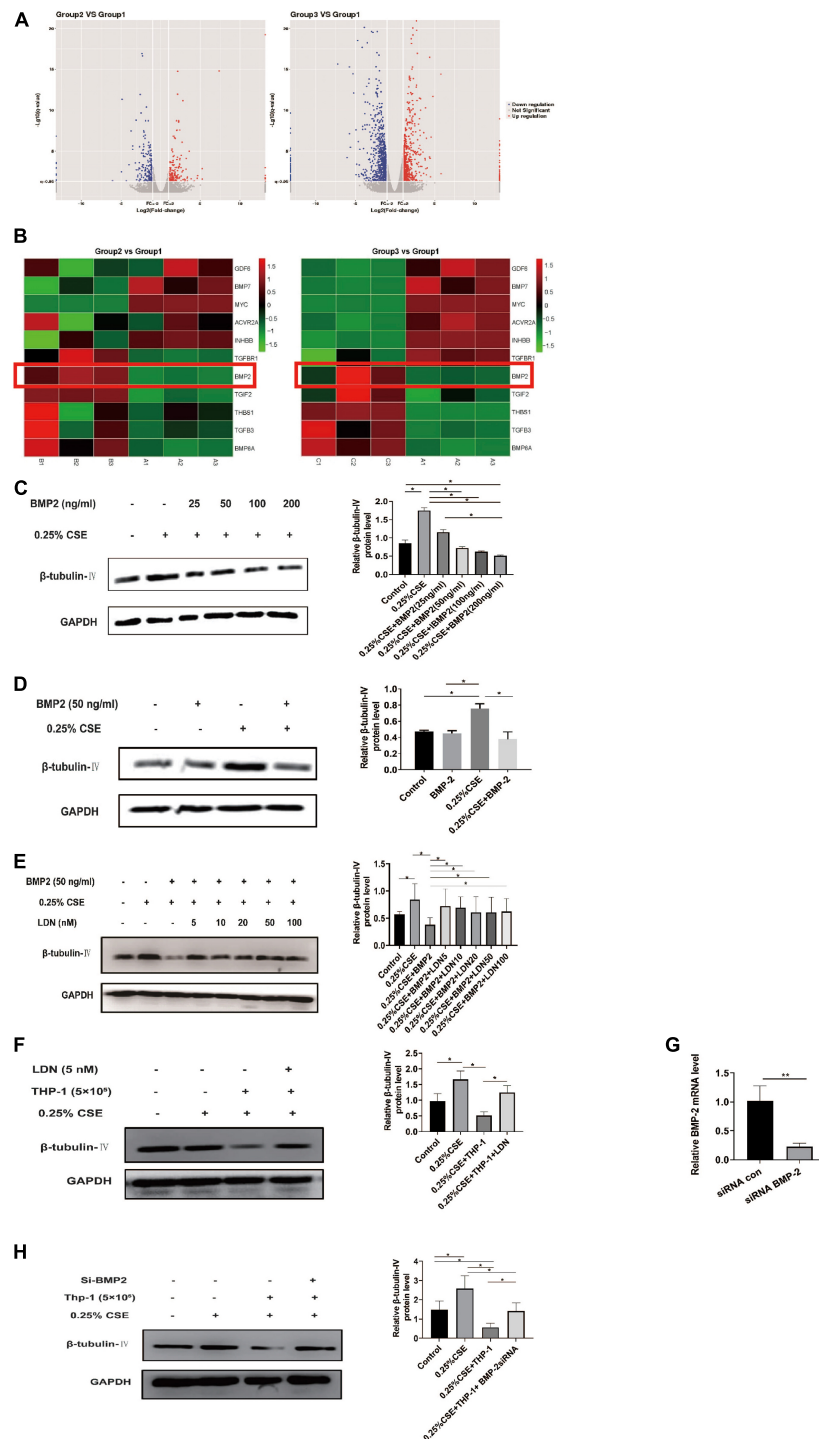
Both LDN and BMP-2 siRNA mostly block the inhibitory effect of THP-1 cells on the  $\beta$ -tubulin-IV levels of BEAS-2B cells, demonstrating that BMP-2 produced by THP-1 cells plays an essential role in inhibiting the  $\beta$ -tubulin-IV levels of BEAS-2B cells.

## DISCUSSION

Airway epithelium, which plays a key role in the mucosal defense response of the host to pathogens, consists of about 50–70% ciliated cells, 30% basal cells, secretory cells (up to 25% goblet cells, 11% Clara cells and some serous cells), neuroendocrine cells, and a few unclassified or intermediate cells (Mercer et al., 1994; Boers et al., 1998; Montoro et al., 2018). It is clear that ciliated cells are the main cell types of human airways. Each ciliated cell has about 300 motile cilia. Normal amounts of cilia and coordinated cilia beating frequency are necessary for normal mucociliary clearance (Wanner et al., 1996). Goblet cells mainly secrete mucin (Roy et al., 2014), which is coated on the surface of the airway, and capture inhaled particles. Then, these exotic particles can be expelled out of the airway through the beating of cilia. Clara cells secrete a 10 kD protein, called CC10 or CCSP, which is an anti-inflammatory and immunoregulatory protein (Chen et al., 2001; Mandal et al., 2004; Liu et al., 2013). Normal cilium structure and function, the appropriate amount of mucus as well as physicochemical properties, and the appropriate amount of lining fluid layer around the cilia constitute the normal MCC function. Impaired MCC function means that the ability to get rid of pathogens and other exotics is weakened. Apart from viral or bacterial pathogens, cigarette smoking has a profound impact on health.

Chemical analysis has identified more than 3,800 compounds in CS, which contain many harmful substances (Lofroth, 1989; Brunneemann and Hoffmann, 1991; Pryor and Stone, 1993; Hasday et al., 1999). Long-term chronic cigarette smoking exposure is known to cause the development of COPD, which has some conventional characteristics, such as emphysema and remodeling of the airway wall.

Cigarette smoke exposure is the most appropriate model to study COPD in mice, and several exposure methods are available (Tanner and Single, 2019). The protocols for CS exposure in mice vary greatly in length, frequency, and numbers of CS exposures, as well as in the exposure mode and cigarette being used (Fricker et al., 2014). Although these different methods of smoke generation affect smoke components and its concentrations, they seemingly do not significantly influence the animal disease state



**FIGURE 5 |** Bone morphogenetic protein-2 (BMP-2) protein has an inhibitory effect on  $\beta$ -tubulin-IV levels of Beas-2B cells. **(A)** The volcano images of genes whose expression difference was more than two-fold in THP-1 cells co-cultured with BEAS-2B cells (group 2, **B**) vs. THP-1 cells (group 1, **A**) and THP-1 cells co-cultured with BEAS-2B cells under stimulation of CSE (group 3, **C**) vs. THP-1 cells (group 1, **A**). **(B)** The heat map of genes related to TGF- $\beta$  superfamily in group 2 vs. group 1 and group 3 vs. group 1. **(C)**  $\beta$ -tubulin-IV levels of BEAS-2B after being stimulated with 0.25% CSE and different concentrations of recombinant human BMP-2 protein. **(D)**  $\beta$ -tubulin-IV levels of BEAS-2B after being stimulated with 0.25% CSE and 50 ng/ml recombinant human BMP-2 protein. **(E)**  $\beta$ -tubulin-IV levels of BEAS-2B pre-treated with different concentrations of BMP receptor inhibitor LDN before being stimulated with 0.25% CSE and 50 ng/ml recombinant human BMP-2 protein. **(F)**  $\beta$ -tubulin-IV levels of BEAS-2B pre-treated with 5 nM LDN before co-cultured with THP-1 cells and being stimulated with 0.25% CSE. **(G)** BMP2 mRNA expression of THP-1 cells pre-incubated with siBMP2 for 24 h. **(H)**  $\beta$ -tubulin-IV levels of BEAS-2B after being stimulated with 0.25% CSE and co-cultured with THP-1 cells pre-incubated with siBMP2. All  $\beta$ -tubulin-IV protein levels were quantitated by densitometry and normalized to GAPDH (\* $P$  < 0.05, \*\* $P$  < 0.01). Red box indicates the result of BMP-2 in heatmap of RNA-seq analysis between groups.



(Leberl et al., 2013). Differences of interspecies in animal models influence the development time of emphysematous phenotypes following CS exposure. Generally, it takes 4–6 months to display signs of disease in most animal models (Ofulue and Ko, 1999; Kratzer et al., 2013b). In this study, we used a whole-body exposure for 4 and 16 weeks, respectively, to establish mouse models of acute and chronic respiratory diseases.

Long-term chronic CS exposure causes significant changes in the airway epithelium. A link between smoking and decreased ciliated cells had already been established 60 years ago, which showed that the more the number of cigarettes, the more is the area of cilia absence (Auerbach et al., 1961). Human airway biopsy specimens showed that cilia were approximately 13% to 15% shorter in smokers than in healthy people who had never smoked (Leopold et al., 2009). Long-term chronic CS exposure in DAB/2 and C57BL/6J mice led to an increase in the area without cilia, disordered arrangement, and shortening of cilia in the area with cilia (Bartalesi et al., 2005). It is reported that CS exposure could affect the ciliogenesis by inhibiting the essential genes for ciliogenesis, such as MCIDAS and FOXJ1 (You et al., 2004; Tamashiro et al., 2009; Stubbs et al., 2012; Didon et al., 2013; Brekman et al., 2014). Their downstream protein, acetylated  $\alpha$ -tubulin (Griggs et al., 2017) or  $\beta$ -tubulin-IV (Li et al., 2017), is often used as a specific symbol of ciliated cells. Some research showed that when primary human bronchial epithelial cells were cultured at air-liquid interface, CS exposure could decrease the number of ciliated cells and acetylated tubulin protein levels (Schamberger et al., 2015), while roflumilast N-oxide could resume  $\beta$ -tubulin-IV protein levels (Milara et al., 2012). In this experiment, both acute and chronic CS exposure in mice could lead to decreased expression of foxj1 mRNA and of  $\beta$ -tubulin-IV protein in the airway. In chronic CS exposure mode, we also detected  $\beta$ -tubulin-IV protein in the airway through immunofluorescence. Because the immunofluorescence staining was continuous, it was hard to tell which cilia came from which cell, so it was impossible to distinguish  $\beta$ -tubulin-IV-positive cells from negative ones. We had to quantify the ratio of the immunofluorescence staining area to the inner layer cells area. The results showed that this proportion was 15.6% in the CS group and 23.4% in the control group, verifying that CS exposure led to a decrease of ciliated cells. Compared with the data from previous studies that ciliated cells account for more than half of the epithelial cells, the proportion of ciliated cells in this study was reduced. There might be two reasons: One reason is that the ratio of the fluorescence area to the inner layer cells area itself leads to a decrease in the proportion of ciliated cells; the other reason might be due to the deviation of immunofluorescence staining of paraffin sections. Anyway, the decrease of ciliated cells caused by cigarette smoking is consistent.

Patients with COPD experience significant changes in the airway epithelium that not only impede pathogen clearance but also trigger an inflammatory response (Rock and Hogan, 2011). This inflammation is evident throughout the tracheal tree and lung tissue, presenting with increased infiltration of neutrophils, macrophages, and lymphocytes (Hogg et al., 2004; Brusselle et al., 2011; Faner et al., 2013). Macrophages and neutrophils could release proteases, namely, neutral elastin, MMP, and cathepsin, by

degrading the ECM, leading to lung tissue damage (Chapman and Shi, 2000). Lymphocyte infiltration was prominent with CD8<sup>+</sup>T cells (Saetta et al., 1999; Caramori et al., 2016). CD8<sup>+</sup>T cells secrete various soluble cytokines and chemokines to recruit and activate macrophages producing more cytokines. The produced elastin fragments act as a monocyte chemokine to enhance macrophage-mediated lung injury (Small et al., 2001; Grumelli et al., 2004). Lymphocytes can also secrete macrophage migration inhibitors. Therefore, to some extent, macrophage aggregation depends on the presence of lymphocytes.

The results showed that, after acute CS exposure in mice, the number of macrophages in the lung of the CS group increased significantly compared with that of the control group detected by FACS and IHC, which was consistent with previous studies, suggesting that acute CS exposure could induce the accumulation of macrophages in the lung. After chronic CS exposure in mice, the number of CD68<sup>+</sup> macrophages in the lung of the CS group increased significantly compared with that of the control group detected by IHC. But there was only an increasing trend of CD11b<sup>+</sup>F4/80<sup>+</sup> macrophages in the lung of the CS group detected by FACS and without significant difference between the two groups. This mild difference might be caused by the specificity of different markers.

After chronic CS exposure, we successfully induced emphysema development in the mouse model. The pathogenesis of emphysema is complex and heterogeneous, that is, several mechanisms coexist and interact (Agusti and Faner, 2018), including, the imbalance of proteolytic/antiproteolytic enzyme and oxidation/anti-oxidation, increased epithelial cell apoptosis, and innate and adaptive immune abnormalities (Tuder and Petrache, 2012). CS exposure could induce emphysema in lac gene-deficient or SCID-deficient mice, indicating that the development of emphysema did not require adaptive immunity (D'hulst et al., 2005; De Cunto et al., 2016). However, other studies have shown that CD8<sup>+</sup> T lymphocytes play a protective role in emphysema development induced by CS exposure (Maeno et al., 2007; Motz et al., 2010). This suggests that the role of adaptive immunity in COPD development remains controversial, but it is indisputable that innate immunity plays a role throughout the development of COPD. Increased alveolar macrophage numbers are clinically correlated with COPD severity (Di Stefano et al., 1998; Retamales et al., 2001). It has been suggested that persistent intrinsic immune activation contributes to the persistence of chronic airway inflammation. Airway epithelial progenitor cells provide upstream stimulation signals for chronic intrinsic immune activation, suggesting some interaction between undifferentiated airway epithelial cells and innate immunity, especially macrophages (Byers et al., 2013). How macrophages interact with airway epithelium is now a hot topic in the research.

We want to know if there is a relationship between macrophage and ciliary protein levels of airway epithelium.

The majority of previous literature reported that CS could lead to decreased ciliated cells (Milara et al., 2012; Schamberger et al., 2015; Amatngalim et al., 2018) or shorter cilia (Bartalesi et al., 2005). We first detected the impact of CSE on ciliary protein levels of BEAS-2B cells. Unexpectedly, the results showed

that ciliary protein levels of BEAS-2B cells decreased with the stimulation of 1% CSE but increased with the stimulation of 0.25% CSE. This discrepancy may be due to fact that primary bronchial epithelial cells (PBECs) are far more sensitive to CSE stimulation than BEAS-2B cells, and even a lower concentration of CSE could result in decreased ciliated cells. Another possibility is that previous PBECs experiments did not study the effect of different concentrations of CSE on ciliated cells. It is due to the fact that mild exotic stimulation could increase ciliary protein levels or promote PBECs differentiated toward ciliated cells from a physiological protective perspective. In the following tests, we used 0.25% CSE as the stimulation condition.

When BEAS-2B cells and activated THP-1 cells were co-cultured in Transwell with or without the stimulation of CSE, the results showed that THP-1 cells could inhibit the ciliary protein levels of BEAS-2B cells. This inhibitory effect was significantly enhanced with CSE stimulation. Therefore, it is speculated that activated macrophages may affect the ciliation, which will be aggravated during cigarette smoking exposure in humans. Subsequently, we tested the transcriptome of three different kinds of THP-1 cells, namely, activated THP-1 cells by PMA, activated THP-1 cells co-cultured with BEAS-2B, and activated THP-1 cells co-cultured with BEAS-2B following the stimulation of CSE.

The number of differentially expressed genes was much higher when THP-1 and BEAS-2B cells were co-cultured under the stimulation of CSE than that of co-culture alone, which suggests that CSE causes a more complicated situation. Comparison among groups showed that there were 11 differentially expressed genes that belong to the TGF- $\beta$  superfamily, which has more than 40 ligand members, involving in embryonic development, airway epithelial differentiation, tissue balance, and many disease states (Derynck and Zhang, 2003; Guo and Wang, 2009; Massague, 2012). We speculated that genes that have an inhibitory effect on  $\beta$ -tubulin-IV protein levels of BEAS-2B should be soluble substances and upregulated in group 2 and group 3 at the same time. Ruling out the other nine genes and TGIF2, which has a similar performance with BMP-2 but not stable between group 1 and group 3, BMP-2 protein was finally screened out as a probable target.

Bone morphogenetic protein signaling plays an essential role in the maintenance of mature lung tissue. In adult mice, defect in BMP/Smad signaling leads to abnormal pulmonary vascular remodeling and pulmonary hypertension (Huang et al., 2009), while the BMP-2 receptor inhibitor LDN could reverse squamous metaplasia and increase the number of ciliated cells and secretory cells (Lee et al., 2015). In primary bronchial epithelial cells from healthy volunteers and patients with cystic fibrosis, BMP activity inhibition could promote cell differentiation and increase the population of ciliated cells. Following stimulation with recombinant BMP, cell differentiation was blocked and ciliated cells decreased significantly (Cibois et al., 2015).

The results showed that, under the stimulation of CSE, recombinant human BMP-2 protein could decrease the ciliary protein levels of BEAS-2B cells in a dose-dependent manner. In order to further verify whether it was BMP-2 secreted by activated THP-1 cells that inhibits ciliary protein levels, we interfered with BMP signal either through blocking BMP

signal transduction by BMP receptor blocker LDN or through decreasing the production of BMP-2 secreted from THP-1 cells by siRNA interference. The results showed that ciliary protein levels of BEAS-2B cells mostly reversed after LDN pre-treatment. Then, we used BMP-2 siRNA to interfere with THP-1 cells, and BMP-2 expression of THP-1 cells was inhibited by more than 70%, and ciliary protein levels of BEAS-2B cells were also restored. These results implied that the BMP-2 from activated macrophages will activate the BMP signal in bronchial epithelial cells, resulting in inhibition of ciliary protein levels.

In conclusion, enhancing BMP signal by exogenous recombinant human BMP-2 protein could inhibit the ciliary protein levels of BEAS-2B cells, and weakening BMP signal using BMP receptor blocker or decreasing the production of BMP-2 from THP-1 cells could reverse the inhibitory effect of THP-1 cells on the ciliary protein levels of BEAS-2B cells. This indicated that macrophages activated by CSE could secrete more BMP-2, which activates the BMP signal, leading to inhibition of ciliary protein levels.

What is the exact mechanism of BMP-2 inhibiting on the ciliary protein levels? It has been reported that, during the differentiation of neuroepithelial cells, BMP-2 could activate Smad1, which mediates Notch signal enhancement, inhibiting neuronal differentiation (Takizawa et al., 2003), suggesting that there is an interaction between BMP and Notch pathway, which is closely related to the differentiation of airway basal cell (Whitsett and Kalinichenko, 2011; Firth et al., 2014). Whether BMP-2 affects the ciliary protein levels through the Notch pathway will be further discussed in follow-up studies. Additionally, what is in the CSE mixture? Although we did not explore the exact substances that played an essential role in this study, generally, CSE compositions were as follows: 18.7% acetonitrile, 18.0% acetone, 12.5% 2-hydroxy-2-methyl-propanenitrile, 8.98% nicotine, and 5.86% nicotyrine (Kim et al., 2018). Without a doubt, knowing the exact composition of CSE will be helpful to explore the molecular mechanism of its effect on airway remodeling in future research.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care and Use Committee of Hebei Medical University.

## AUTHOR CONTRIBUTIONS

LW, ZW, and WL conceived and designed the research. ZW and WL performed the experiments, analyzed the data, and

interpreted the results of the experiments. JW and XG prepared the figures and drafted the manuscript. LW and CM edited and revised the manuscript, and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Agusti, A., and Faner, R. (2018). COPD beyond smoking: new paradigm, novel opportunities. *Lancet Respir. Med.* 6, 324–326. doi: 10.1016/s2213-2600(18)30060-2
- Agusti, A., and Hogg, J. C. (2019). Update on the pathogenesis of chronic obstructive pulmonary disease. *N. Engl. J. Med.* 381, 1248–1256.
- Amatngalim, G. D., Schrumpf, J. A., Dishchekian, F., Mertens, T. C. J., Ninaber, D. K., van der Linden, A. C., et al. (2018). Aberrant epithelial differentiation by cigarette smoke dysregulates respiratory host defence. *Eur. Respir. J.* 51:1701009. doi: 10.1183/13993003.01009-2017
- Auerbach, O., Stout, A. P., Hammond, E. C., and Garfinkel, L. (1961). Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N. Engl. J. Med.* 265, 253–267. doi: 10.1056/nejm196108102650601
- Barnes, P. J. (2013). New anti-inflammatory targets for chronic obstructive pulmonary disease. *Nat. Rev. Drug Discov.* 12, 543–559. doi: 10.1038/nrd4025
- Barnes, P. J. (2014). Cellular and molecular mechanisms of chronic obstructive pulmonary disease. *Clin. Chest Med.* 35, 71–86.
- Bartalesi, B., Cavarra, E., Fineschi, S., Lucattelli, M., Lunghi, B., Martorana, P. A., et al. (2005). Different lung responses to cigarette smoke in two strains of mice sensitive to oxidants. *Eur. Respir. J.* 25, 15–22. doi: 10.1183/09031936.04.00067204
- Boers, J. E., Ambergen, A. W., and Thunissen, F. B. (1998). Number and proliferation of basal and parabasal cells in normal human airway epithelium. *Am. J. Respir. Crit. Care Med.* 157(6 Pt 1), 2000–2006. doi: 10.1164/ajrccm.157.6.9707011
- Brekman, A., Walters, M. S., Tilley, A. E., and Crystal, R. G. (2014). FOXJ1 prevents cilia growth inhibition by cigarette smoke in human airway epithelium in vitro. *Am. J. Respir. Cell Mol. Biol.* 51, 688–700. doi: 10.1165/rcmb.2013-0363oc
- Brunnemann, K. D., and Hoffmann, D. (1991). Analytical studies on tobacco-specific N-nitrosamines in tobacco and tobacco smoke. *Crit. Rev. Toxicol.* 21, 235–240. doi: 10.3109/10408449109017910
- Brusselle, G. G., Joos, G. F., and Bracke, K. R. (2011). New insights into the immunology of chronic obstructive pulmonary disease. *Lancet* 378, 1015–1026.
- Bu, T., Wang, L. F., and Yin, Y. Q. (2020). How do innate immune cells contribute to airway remodeling in COPD progression? *Int. J. Chron. Obstruct. Pulmon. Dis.* 15, 107–116. doi: 10.2147/copd.s235054
- Byers, D. E., Alexander-Brett, J., Patel, A. C., Agapov, E., Dang-Vu, G., Jin, X., et al. (2013). Long-term IL-33-producing epithelial progenitor cells in chronic obstructive lung disease. *J. Clin. Invest.* 123, 3967–3982. doi: 10.1172/jci65570
- Caramori, G., Casolari, P., Barczyk, A., Durham, A. L., Stefano, A. D., Adcock, I., et al. (2016). COPD immunopathology. *Semin. Immunopathol.* 38, 497–515. doi: 10.1007/s00281-016-0561-5
- Chapman, H. A., and Shi, G. P. (2000). Protease injury in the development of COPD: Thomas A. Neff Lecture. *Chest* 117(5 Suppl. 1), 295S–299S.
- Chen, L. C., Zhang, Z., Myers, A. C., and Huang, S. K. (2001). Cutting edge: altered pulmonary eosinophilic inflammation in mice deficient for Clara cell secretory 10-kDa protein. *J. Immunol.* 167, 3025–3028. doi: 10.4049/jimmunol.167.6.3025
- Cibois, M., Luxardi, G., Chevalier, B., Thome, V., Mercey, O., Zaraosi, L.-E., et al. (2015). BMP signalling controls the construction of vertebrate mucociliary epithelia. *Development* 142, 2352–2363. doi: 10.1242/dev.118679
- De Cunto, G., Lunghi, B., Bartalesi, B., Cavarra, E., Fineschi, S., Olivieri, C., et al. (2016). Severe reduction in number and function of peripheral T cells does not afford protection toward emphysema and bronchial remodeling induced in mice by cigarette smoke. *Am. J. Pathol.* 186, 1814–1824. doi: 10.1016/j.ajpath.2016.03.002
- Derynck, R., and Zhang, Y. E. (2003). Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* 425, 577–584. doi: 10.1038/nature02006
- D'hulst, A. I., Maes, T., Bracke, K. R., Demedts, I. K., Tournoy, K. G., Joos, G. F., et al. (2005). Cigarette smoke-induced pulmonary emphysema in scid-mice. Is the acquired immune system required? *Respir. Res.* 6:147.
- Di Stefano, A., Capelli, A., Lusuadi, M., Balbo, P., Vecchio, C., Maestrelli, P., et al. (1998). Severity of airflow limitation is associated with severity of airway inflammation in smokers. *Am. J. Respir. Crit. Care Med.* 158, 1277–1285. doi: 10.1164/ajrccm.158.4.9802078
- Didon, L., Zwick, R. K., Chao, I. W., Walters, M. S., Wang, R., Hackett, N. R., et al. (2013). RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium. *Respir. Res.* 14:70. doi: 10.1186/1465-9921-14-70
- Faner, R., Cruz, T., and Agusti, A. (2013). Immune response in chronic obstructive pulmonary disease. *Expert Rev. Clin. Immunol.* 9, 821–833.
- Ferhani, N., Letuve, S., Kozhich, A., Thibaudau, O., Grandsaigne, M., Maret, M., et al. (2010). Expression of high-mobility group box 1, and of receptor for advanced glycation end products in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 181, 917–927. doi: 10.1164/rccm.200903-0340oc
- Firth, A. L., Dargitz, C. T., Qualls, S. J., Menon, T., Wright, R., Singer, O., et al. (2014). Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, E1723–E1730.
- Fricker, M., Deane, A., and Hansbro, P. M. (2014). Animal models of chronic obstructive pulmonary disease. *Expert Opin. Drug Discov.* 9, 629–645.
- GBD 2015 Disease and Injury Incidence and Prevalence Collaborators (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1545–1602.
- GBD 2015 Mortality and Cause of Death Collaborators (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1459–1544.
- GBD 2015 Risk Factors Collaborators. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1659–1724.
- Ghosh, A., Coakley, R. D., Ghio, A. J., Muhlebach, M. S., Esther, C. R., Alexis, N. E., et al. (2019). Chronic E-cigarette use increases neutrophil elastase and matrix metalloprotease levels in the lung. *Am. J. Respir. Crit. Care Med.* 200, 1392–1401. doi: 10.1164/rccm.201903-0615oc
- Griggs, T. F., Bochkov, Y. A., Basnet, S., Pasic, T. R., Brockman-Schneider, R. A., Palmenberg, A. C., et al. (2017). Rhinovirus C targets ciliated airway epithelial cells. *Respir. Res.* 18:84.
- Grumelli, S., Corry, D. B., Song, L. Z., Song, L., Green, L., Huh, J., et al. (2004). An immune basis for lung parenchymal destruction in chronic obstructive pulmonary disease and emphysema. *PLoS Med.* 1:e8. doi: 10.1371/journal.pmed.0010008
- Guo, X., and Wang, X. F. (2009). Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res.* 19, 71–88. doi: 10.1038/cr.2008.302
- Hasday, J. D., Bascom, R., Costa, J. J., Fitzgerald, T., and Dubin, W. (1999). Bacterial endotoxin is an active component of cigarette smoke. *Chest* 115, 829–835. doi: 10.1378/chest.115.3.829
- Hogg, J. C., Chu, F., Utokaparch, S., Woods, R., Elliott, W. M., Buzatu, L., et al. (2004). The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N. Engl. J. Med.* 350, 2645–2653.

## FUNDING

This work was supported by grants from the National Natural Science Foundation of China (grant nos. 81971474, 31770971, 31370914, and 81802014) and the Scientific and the Key R&D Projects of Hebei Province (grant no. 20277738D).



- Huang, Z., Wang, D., Ihida-Stansbury, K., Jones, P. L., and Martin, J. F. (2009). Defective pulmonary vascular remodeling in Smad8 mutant mice. *Hum. Mol. Genet.* 18, 2791–2801. doi: 10.1093/hmg/ddp214
- Hume, P. S., Gibbings, S. L., Jakubczak, C. V., Tuder, R. M., Curran-Everett, D., Henson, P. M., et al. (2020). Localization of macrophages in the human lung via design-based stereology. *Am. J. Respir. Crit. Care Med.* 201, 1209–1217. doi: 10.1164/rccm.201911-2105oc
- Hunninghake, G. W., and Crystal, R. G. (1983). Cigarette smoking and lung destruction. accumulation of neutrophils in the lungs of cigarette smokers. *Am. Rev. Respir. Dis.* 128, 833–838.
- Jeon, B. N., Song, J. Y., Huh, J. W., Yang, W. I., and Hur, M. W. (2019). Derepression of matrix metalloproteinase gene transcription and an emphysema-like phenotype in transcription factor Zbtb7c knockout mouse lungs. *FEBS Lett.* 593, 2665–2674. doi: 10.1002/1873-3468.13501
- Kim, W. D., Chi, H. S., Choe, K. H., Oh, Y. M., Lee, S. D., Kim, K. R., et al. (2013). A possible role for CD8+ and non-CD8+ cell granzyme B in early small airway wall remodeling in centrilobular emphysema. *Respirology* 18, 688–696. doi: 10.1111/resp.12069
- Kim, Y. H., An, Y. J., Jo, S., Lee, S. H., Lee, S. J., Choi, S. J., et al. (2018). Comparison of volatile organic compounds between cigarette smoke condensate (CSC) and extract (CSE) samples. *Environ. Health Toxicol.* 33:e2018012-0.
- Kratzer, A., Chu, H. W., Salys, J., Moumen, Z., Leberl, M., Bowler, R., et al. (2013a). Endothelial cell adhesion molecule CD146: implications for its role in the pathogenesis of COPD. *J. Pathol.* 230, 388–398. doi: 10.1002/path.4197
- Kratzer, A., Salys, J., Nold-Petry, C., Cool, C., Zamora, M., Bowler, R., et al. (2013b). Role of IL-18 in second-hand smoke-induced emphysema. *Am. J. Respir. Cell Mol. Biol.* 48, 725–732. doi: 10.1165/rcmb.2012-0173oc
- Leberl, M., Kratzer, A., and Taraseviciene-Stewart, L. (2013). Tobacco smoke induced COPD/emphysema in the animal model—are we all on the same page? *Front. Physiol.* 4:91.
- Lee, S. N., Lee, D. H., Lee, M. G., and Yoon, J. H. (2015). Proprotein convertase 5/6a is associated with bone morphogenetic protein-2-induced squamous cell differentiation. *Am. J. Respir. Cell Mol. Biol.* 52, 749–761. doi: 10.1165/rcmb.2014-0029oc
- Leopold, P. L., O'mahony, M. J., Lian, X. J., Tilley, A. E., Harvey, B. G., Crystal, R. G., et al. (2009). Smoking is associated with shortened airway cilia. *PLoS One* 4:e8157. doi: 10.1371/journal.pone.0008157
- Li, Y. Y., Liu, J., Li, C. W., Subramaniam, S., Chao, S. S., Yu, F. G., et al. (2017). Myrtol standardized affects mucociliary clearance. *Int. Forum Allergy Rhinol.* 7, 304–311. doi: 10.1002/alr.21878
- Liu, Y., Yu, H. J., Wang, N., Zhang, Y. N., Huang, S. K., Cui, Y. H., et al. (2013). Clara cell 10-kDa protein inhibits T(H)17 responses through modulating dendritic cells in the setting of allergic rhinitis. *J. Allergy Clin. Immunol.* 131, 387–394.e1–e12.
- Lofroth, G. (1989). Environmental tobacco smoke: overview of chemical composition and genotoxic components. *Mutat. Res.* 222, 73–80. doi: 10.1016/0165-1218(89)90021-9
- Maeno, T., Houghton, A. M., Quintero, P. A., Grumelli, S., Owen, C. A., Shapiro, S. D., et al. (2007). CD8+ T Cells are required for inflammation and destruction in cigarette smoke-induced emphysema in mice. *J. Immunol.* 178, 8090–8096. doi: 10.4049/jimmunol.178.12.8090
- Mandal, A. K., Zhang, Z., Ray, R., Choi, M. S., Chowdhury, B., Pattabiraman, N., et al. (2004). Uteroglobin represses allergen-induced inflammatory response by blocking PGD2 receptor-mediated functions. *J. Exp. Med.* 199, 1317–1330. doi: 10.1084/jem.20031666
- Mannino, D. M., and Buist, A. S. (2007). Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* 370, 765–773. doi: 10.1016/s0140-6736(07)61380-4
- Martinez, F. D. (2016). Early-life origins of chronic obstructive pulmonary disease. *N. Engl. J. Med.* 375, 871–878. doi: 10.1056/nejmra1603287
- Massague, J. (2012). TGFbeta signalling in context. *Nat. Rev. Mol. Cell Biol.* 13, 616–630.
- Mathers, C. D., and Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 3:e442. doi: 10.1371/journal.pmed.0030442
- Mercer, R. R., Russell, M. L., Roggli, V. L., and Crapo, J. D. (1994). Cell number and distribution in human and rat airways. *Am. J. Respir. Cell Mol. Biol.* 10, 613–624. doi: 10.1165/ajrcmb.10.6.8003339
- Milara, J., Armengot, M., Banuls, P., Tenor, H., Beume, R., and Artigues, E. (2012). Roflumilast N-oxide, a PDE4 inhibitor, improves cilia motility and ciliated human bronchial epithelial cells compromised by cigarette smoke in vitro. *Br. J. Pharmacol.* 166, 2243–2262. doi: 10.1111/j.1476-5381.2012.01929.x
- Montoro, D. T., Haber, A. L., Biton, M., Vinarsky, V., Lin, B., Birket, S. E., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 560, 319–324. doi: 10.1038/s41586-018-0393-7
- Morissette, M. C., Parent, J., and Milot, J. (2007). Perforin, granzyme B, and FasL expression by peripheral blood T lymphocytes in emphysema. *Respir. Res.* 8:62.
- Morris, D. G., Huang, X., Kaminski, N., Wang, Y., Shapiro, S. D., Dolganov, G., et al. (2003). Loss of integrin alpha(v)beta6-mediated TGF-beta activation causes Mmp12-dependent emphysema. *Nature* 422, 169–173. doi: 10.1038/nature01413
- Motz, G. T., Eppert, B. L., Wesselkamper, S. C., Flury, J. L., and Borchers, M. T. (2010). Chronic cigarette smoke exposure generates pathogenic T cells capable of driving COPD-like disease in Rag2-/- mice. *Am. J. Respir. Crit. Care Med.* 181, 1223–1233. doi: 10.1164/rccm.200910-1485oc
- National Research Council (US) Committee for the Update of the Guide for the Care, and Use of Laboratory Animals (2011). *The Guide for the Care and Use of Laboratory Animals*. Washington, DC: National Academies Press.
- Ngan, D. A., Vickerman, S. V., Granville, D. J., Man, S. F., and Sin, D. D. (2009). The possible role of granzyme B in the pathogenesis of chronic obstructive pulmonary disease. *Ther. Adv. Respir. Dis.* 3, 113–129. doi: 10.1177/1753465809341965
- Ofulue, A. F., and Ko, M. (1999). Effects of depletion of neutrophils or macrophages on development of cigarette smoke-induced emphysema. *Am. J. Physiol.* 277, L97–L105.
- Pryor, W. A., and Stone, K. (1993). Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxyxynitrate, and peroxyxynitrite. *Ann. N. Y. Acad. Sci.* 686, 12–27. doi: 10.1111/j.1749-6632.1993.tb39148.x
- Randell, S. H. (2006). Airway epithelial stem cells and the pathophysiology of chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* 3, 718–725. doi: 10.1513/pats.200605-117sf
- Rawlins, E. L., and Hogan, B. L. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. *Am. J. Physiol. Lung Cell Mol. Physiol.* 295, L231–L234.
- Reddel, R. R., Ke, Y., Gerwin, B. I., McMenamin, M. G., Lechner, J. F., Su, R. T., et al. (1988). Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection via strontium phosphate coprecipitation with a plasmid containing SV40 early region genes. *Cancer Res.* 48, 1904–1909.
- Retamales, I., Elliott, W. M., Meshi, B., Coxson, H. O., Pare, P. D., and Scierba, F. C. (2001). Amplification of inflammation in emphysema and its association with latent adenoviral infection. *Am. J. Respir. Crit. Care Med.* 164, 469–473. doi: 10.1164/ajrcm.164.3.2007149
- Rock, J. R., and Hogan, B. L. (2011). Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell Dev. Biol.* 27, 493–512. doi: 10.1146/annurev-cellbio-100109-104040
- Rock, J. R., Randell, S. H., and Hogan, B. L. (2010). Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis. Model. Mech.* 3, 545–556. doi: 10.1242/dmm.006031
- Roy, M. G., Livraghi-Butrico, A., Fletcher, A. A., McElwee, M. M., Evans, S. E., Boerner, R. M., et al. (2014). Muc5b is required for airway defence. *Nature* 505, 412–416.
- Saetta, M., Baraldo, S., Corbino, L., Turato, G., Braccioni, F., Rea, F., et al. (1999). CD8+ve cells in the lungs of smokers with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 160, 711–717. doi: 10.1164/ajrcm.160.2.9812020
- Sakhatskyy, P., Gabino Miranda, G. A., Newton, J., Lee, C. G., Choudhary, G., Vang, A., et al. (2014). Cigarette smoke-induced lung endothelial apoptosis and emphysema are associated with impairment of FAK and eIF2alpha. *Microvasc. Res.* 94, 80–89. doi: 10.1016/j.mvr.2014.05.003
- Schamberger, A. C., Staab-Weijnitz, C. A., Mise-Racek, N., and Eickelberg, O. (2015). Cigarette smoke alters primary human bronchial epithelial cell differentiation at the air-liquid interface. *Sci. Rep.* 5:8163.



- Small, B. A., Dressel, S. A., Lawrence, C. W., Darke, D. R., Stoler, M. H., and Enelow, R. I. (2001). CD8(+) T cell-mediated injury in vivo progresses in the absence of effector T cells. *J. Exp. Med.* 194, 1835–1846. doi: 10.1084/jem.194.12.1835
- Stubbs, J. L., Vadar, E. K., Axelrod, J. D., and Kintner, C. (2012). Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation. *Nat. Cell Biol.* 14, 140–147. doi: 10.1038/ncb2406
- Takizawa, T., Ochiai, W., Nakashima, K., and Taga, T. (2003). Enhanced gene activation by Notch and BMP signaling cross-talk. *Nucleic Acids Res.* 31, 5723–5731. doi: 10.1093/nar/gkg778
- Tamashiro, E., Xiong, G., Anselmo-Lima, W. T., Kreindler, J. L., Palmer, J. N., Cohen, N. A., et al. (2009). Cigarette smoke exposure impairs respiratory epithelial ciliogenesis. *Am. J. Rhinol. Allergy* 23, 117–122. doi: 10.2500/ajra.2009.23.3280
- Tanner, L., and Single, A. B. (2019). Animal models reflecting chronic obstructive pulmonary disease and related respiratory disorders: translating pre-clinical data into clinical relevance. *J. Innate Immun.* 12, 203–225. doi: 10.1159/000502489
- Tuder, R. M., and Petrache, I. (2012). Pathogenesis of chronic obstructive pulmonary disease. *J. Clin. Invest.* 122, 2749–2755.
- Tuder, R. M., Petrache, I., Elias, J. A., Voelkel, N. F., and Henson, P. M. (2003). Apoptosis and emphysema: the missing link. *Am. J. Respir. Cell Mol. Biol.* 28, 551–554. doi: 10.1165/rcmb.f269
- Vogelmeier, C. F., Criner, G. J., Martinez, F. J., Anzueto, A., Barnes, P. J., Bourbeau, J., et al. (2017). Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am. J. Respir. Crit. Care Med.* 195, 557–582.
- Wanner, A., Salathe, M., and O’rordan, T. G. (1996). Mucociliary clearance in the airways. *Am. J. Respir. Crit. Care Med.* 154(6 Pt 1), 1868–1902. doi: 10.1164/ajrccm.154.6.8970383
- Whitsett, J. A., and Kalinichenko, V. V. (2011). Notch and basal cells take center stage during airway epithelial regeneration. *Cell Stem Cell* 8, 597–598. doi: 10.1016/j.stem.2011.05.008
- Xia, H., Wu, Y., Zhao, J., Li, W., Lu, L., Ma, H., et al. (2021). The aberrant cross-talk of epithelium-macrophages via METTL3-regulated extracellular vesicle miR-93 in smoking-induced emphysema. *Cell Biol. Toxicol.* doi: 10.1007/s10565-021-09585-1 Online ahead of print.
- Xu, F. H., Xiong, D., Xu, Y. F., Cao, S. M., Xue, W. Q., Qin, H. D., et al. (2012). An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein-Barr virus activation. *J. Natl. Cancer Inst.* 104, 1396–1410.
- You, Y., Huang, T., Richer, E. J., Schmidt, J. E., Zabner, J., Borok, Z., et al. (2004). Role of f-box factor foxj1 in differentiation of ciliated airway epithelial cells. *Am. J. Physiol. Lung Cell Mol. Physiol.* 286, L650–L657.
- Zhang, C., Yan, M. Y., Lu, P., Chen, P., Yang, M., Ye, X. W., et al. (2014). Hypomethylation of perforin regulatory elements in CD4+ T cells from rat spleens contributes to the development of autoimmune emphysema. *Respirology* 19, 376–381. doi: 10.1111/resp.12240

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Liang, Ma, Wang, Gao and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Terminal Epitope-Dependent Branch Preference of Siglecs Toward N-Glycans

Shuaishuai Wang<sup>1†</sup>, Congcong Chen<sup>1†</sup>, Minhui Guan<sup>2,3,4</sup>, Ding Liu<sup>1</sup>, Xiu-Feng Wan<sup>2,3,4,5</sup> and Lei Li<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, Georgia State University, Atlanta, GA, United States, <sup>2</sup>MU Center for Influenza and Emerging Infectious Diseases, University of Missouri, Columbia, MO, United States, <sup>3</sup>Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, MO, United States, <sup>4</sup>Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, <sup>5</sup>Department of Electrical Engineering and Computer Science, College of Engineering, University of Missouri, Columbia, MO, United States

## OPEN ACCESS

### Edited by:

Fuming Zhang,  
Rensselaer Polytechnic Institute,  
United States

### Reviewed by:

Takashi Angata,  
Academia Sinica, Taiwan  
Matthew S. Macauley,  
University of Alberta, Canada  
Corwin Nycholat,  
The Scripps Research Institute,  
United States

### \*Correspondence:

Lei Li  
ll22@gsu.edu

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 24 December 2020

**Accepted:** 18 February 2021

**Published:** 29 April 2021

### Citation:

Wang S, Chen C, Guan M, Liu D,  
Wan X-F and Li L (2021) Terminal  
Epitope-Dependent Branch  
Preference of Siglecs Toward N-  
Glycans.  
Front. Mol. Biosci. 8:645999.  
doi: 10.3389/fmolb.2021.645999

Siglecs are sialic acid-binding immunoglobulin-like lectins that play vital roles in immune cell signaling. Siglecs help the immune system distinguish between self and nonself through the recognition of glycan ligands. While the primary binding specificities of Siglecs are known to be divergent, their specificities for complex glycans remain unclear. Herein, we determined N-glycan binding profiles of a set of Siglecs by using a complex asymmetric N-glycan microarray. Our results showed that Siglecs had unique terminal epitope-dependent branch preference when recognizing asymmetric N-glycans. Specifically, human Siglec-3, -9, and -10 prefer the  $\alpha$ 1-3 branch when Sia $\alpha$ 2-6Gal $\beta$ 1-4GlcNAc terminal epitope serves as the binding ligand but prefer the opposite  $\alpha$ 1-6 branch when Sia $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc epitope serves as the ligand. Interestingly, Siglec-10 exhibited dramatic binding divergence toward a pair of Neu5Ac-containing asymmetric N-glycan isomers, as well as their Neu5Gc-containing counterparts. This new information on complex glycan recognition by Siglecs provides insights into their biological roles and applications.

**Keywords:** Siglecs, asymmetric N-glycan, Neu5Gc, Neu5Ac, microarray

## INTRODUCTION

Sialic acid-binding immunoglobulin-like lectins (Siglecs) are cell-surface transmembrane receptors that are differentially expressed on immune cells (Läubli and Varki, 2020). They play critical roles in immune cell signaling and help the immune system to distinguish self and nonself (Macauley et al., 2014). Most Siglecs, with the only exception being sialoadhesin/Siglec-1, have C-terminal regulatory motifs in their cytoplasmic domains that participate in the regulation of immune systems. On the N-terminal, each Siglec has a V-set immunoglobulin (Ig) domain that recognizes sialic acid-containing glycans (Duan and Paulson, 2020). There are 15 human Siglecs and 9 murine Siglecs. Among those, four are conserved across mammals (Siglec-1, 2, 4, and 15). All remaining Siglecs are named CD33-related Siglecs as they contain less conserved structure between humans and other vertebrates, but all have high homologies to CD33.

Siglecs are immune-modulatory receptors within the mammalian immune system. Most Siglecs have intracellular immunoreceptor tyrosine inhibitory motifs (ITIMs) that can, in principle, participate in inhibitory or activating signals. The binding of anti-Siglec antibodies or

multivalent trans-ligand with inhibitory Siglecs can activate/phosphorylate the ITIMs and produce negative signals (Duan and Paulson, 2020). Additionally, some Siglecs are specifically expressed on certain types of immune cells and presented as endocytic receptors. Hence, they were utilized as the desired target for drug development. For example, Siglec-3, also called CD33, is an inhibitory receptor that is relatively specifically expressed on myeloid lineage and endocytosed upon antibody binding, thus serving as a specific target for developing therapeutic antibodies. Gemtuzumab ozogamicin is the first approved CD33-targeting antibody-drug conjugate (ADC) and was used for induction therapy of acute myeloid leukemia (AML) (Laszlo et al., 2014).

Despite the diverse roles that Siglecs play in immune cell regulation and disease processes, their natural ligands, especially the fine binding specificity, toward complex glycans are relatively underinvestigated. Glycan microarray was developed for identifying interactions between glycans and glycan-binding proteins (GBPs) 2 decades ago (Fukui et al., 2002; Palma and Chai, 2019). It enabled simultaneous binding analysis of GBPs to hundreds of glycan structures and had become a major tool to unveil glycan–protein interactions (Gao et al., 2019b). Various versions of glycan microarray were used to investigate interactions between glycans and Siglecs (Blixt et al., 2003; Bochner et al., 2005; Campanero-Rhodes et al., 2006; Rillahan et al., 2012; Rillahan et al., 2013; Gao et al., 2019a). However, the fine specificity details of Siglecs toward natural complex glycans remain largely unknown.

Herein, we investigated the binding specificity of Siglec-3, -9, -10, and -F using a unique glycan microarray containing 98 structurally well-defined complex glycans, revealing a unique terminal epitope-dependent branch preference toward asymmetric *N*-glycans. Particularly, a dramatic binding divergence of Siglec-10 toward a pair of *N*-glycan isomers was observed and further confirmed by synthesized Neu5Gc-containing counterparts. Later, quantitative assay by biolayer interferometry analyses suggested a 67-fold avidity difference among the Neu5Gc-containing isomers.

## MATERIALS AND METHODS

Unless otherwise stated, all chemicals were purchased and used without further purification. The 98 *N*-glycan microarray was prepared as described previously (Supplementary Figure S1) (Li et al., 2019). Sugar nucleotides, including uridine 5'-diphosphogalactose (UDP-Gal) (Muthana et al., 2012), were prepared as described previously. Enzymes including *Neisseria meningitidis*  $\beta$ 1-4galactosyltransferase (NmLgtB) (Lau et al., 2010), *N. meningitidis* CMP-sialic acid synthetase (NmCSS) (Yu et al., 2004), *Pasteurella multocida*  $\alpha$ 2-3sialyltransferase mutant M144D (PmST1-M144D) (Sugiarto et al., 2012), and *Photobacterium damsela*  $\alpha$ 2-6sialyltransferase (Pd26ST) (Yu et al., 2006) were expressed and purified as previously described.

### Chemoenzymatic Synthesis of *N*-Glycans

*N*-glycans **38** and **54** were prepared as previously reported (Li et al., 2015). For the  $\alpha$ 2-6sialylation of **38**, **100**, **54**, and **104**, reactions were carried out in reaction systems containing

Tris-HCl (100 mM, pH 8.0), an acceptor glycan (10 mM), CTP (15 mM), *N*-acetylneuraminic acid (Neu5Ac) or *N*-glycolylneuraminic acid (Neu5Gc) (15 mM), MgCl<sub>2</sub> (10 mM), and appropriate amounts of NmCSS and Pd26ST. Reactions were incubated at 37°C for 3 h and monitored by HPLC. After over 95% acceptor was converted, reactions were quenched by the addition of equal volumes of ice-cold ethanol, concentrated, and subject to HPLC separation to afford compounds **99**, **101**, **103**, and **105**. Product-containing fractions were pooled and lyophilized for characterization and next step modular assembly. For the  $\beta$ 1-4galactosylation of **99** and **103**, reactions were performed in mixtures containing Tris-HCl (100 mM, pH 7.5), an acceptor glycan (10 mM), UDP-Gal (15 mM), MgCl<sub>2</sub> (10 mM), and an appropriate amount of NmLgtB. Reactions were incubated at 37°C overnight and monitored by HPLC. After over 95% acceptor was converted, reactions were quenched, concentrated, and subject to HPLC separation of compounds **100** and **104**. Product-containing fractions were pooled and lyophilized for characterization and subsequent synthesis. The  $\alpha$ 2-3sialylation of **100** and **104** was carried out in reaction systems containing Tris-HCl (100 mM, pH 8.0), an acceptor glycan (10 mM), CTP (15 mM), Neu5Gc (15 mM), MgCl<sub>2</sub> (10 mM), and appropriate amounts of NmCSS and PmST1-M144D. PmST1-M144D-catalyzed reactions were incubated at 37°C for 3 h and monitored by HPLC. After over 90% acceptor was converted, the reaction was quenched, concentrated, and subject to HPLC separation to afford compounds **102** and **106**. Product-containing fractions were then pooled and lyophilized for characterization.

Newly synthesized *N*-glycans were purified by HPLC using a Waters XBridge BEH amide column (130 Å, 5  $\mu$ m, 10 mm  $\times$  250 mm) under a gradient running condition (solvent A: water or 100 mM ammonium formate; solvent B: acetonitrile; flow rate: 4.5 ml/min, B%: 65–50% in 30 min) and monitored by UV absorbance at 210 nm. MALDI-TOF MS analyses were performed on UltrafleXtreme MALDI TOF/TOF Mass Spectrometer (Bruker). Scan range of MS was set according to molecular weight, and reflector mode was used for analysis. Mass spectra were obtained in negative extraction mode with the following voltage settings: ion source 1 (19.0 kV), ion source 2 (15.9 kV), and lens (9.3 kV). The reflector voltage was set to 20 kV. The laser was pulsed at 7 Hz and the pulsed ion extraction time was set at 400 ns. The laser power was kept in the range of 40–90%. <sup>1</sup>H NMR spectra were recorded on a Bruker AVANCE 600 (600 MHz) spectrometer at 25°C. All <sup>1</sup>H Chemical shifts (in ppm) were assigned according to D<sub>2</sub>O ( $\delta$  = 4.79 ppm).

Compound **99**, white power (0.92 mg). <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O)  $\delta$  5.05 (s, 1H), 4.83 (d, *J* = 1.7 Hz, 1H), 4.52 (dd, *J* = 8.0, 2.8 Hz, 2H), 4.47 (d, *J* = 8.4 Hz, 1H), 4.36 (d, *J* = 7.9 Hz, 1H), 4.17 (d, *J* = 1.9 Hz, 1H), 4.11 (dd, *J* = 3.5, 1.6 Hz, 1H), 4.02 (d, *J* = 7.3 Hz, 2H), 3.95–3.29 (m, 42H), 2.60 (dd, *J* = 12.4, 4.7 Hz, 1H), 2.02–1.92 (m, 9H), 1.65 (t, *J* = 12.2 Hz, 1H). MALDI-MS: C<sub>67</sub>H<sub>111</sub>N<sub>5</sub>O<sub>50</sub>, calc. for 1785.6297, found [M-H]<sup>−</sup> 1784.765.

Compound **100**, white power (0.74 mg). <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O)  $\delta$  5.10 (d, *J* = 2.7 Hz, 1H), 5.05 (s, 1H), 4.85 (d, *J* = 1.7 Hz, 1H), 4.54–4.46 (m, 3H), 4.38 (dd, *J* = 13.8, 7.9 Hz, 2H), 4.17 (s, 1H), 4.11 (dd, *J* = 3.6, 1.6 Hz, 1H), 4.03 (d, *J* = 4.4 Hz, 3H),

3.97–3.35 (m, 56H), 2.60 (dd,  $J = 12.4, 4.7$  Hz, 1H), 2.03–1.91 (m, 11H), 1.65 (t,  $J = 12.2$  Hz, 1H). MALDI-MS:  $C_{73}H_{121}N_5O_{55}$ , calc. for 1947.6825, found  $[M-H]^-$  1946.883.

Compound **101**, white power (0.61 mg).  $^1H$  NMR (600 MHz,  $D_2O$ )  $\delta$  5.10 (d,  $J = 2.8$  Hz, 1H), 5.05 (s, 1H), 4.86 (s, 1H), 4.55–4.48 (m, 3H), 4.36 (d,  $J = 7.9$  Hz, 2H), 4.17 (s, 1H), 4.11 (dd,  $J = 3.6, 1.5$  Hz, 1H), 4.03 (s, 3H), 3.95–3.35 (m, 65H), 2.59 (ddd,  $J = 12.2, 7.4, 4.7$  Hz, 2H), 2.03–1.91 (m, 15H), 1.64 (q,  $J = 11.8$  Hz, 2H). MALDI-MS:  $C_{84}H_{138}N_6O_{63}$ , calc. for 2238.7779, found  $[M-2H + Na]^-$  2259.918.

Compound **102**, white power (0.36 mg).  $^1H$  NMR (600 MHz,  $D_2O$ )  $\delta$  5.05 (s, 1H), 4.84 (s, 1H), 4.55–4.43 (m, 4H), 4.36 (d,  $J = 7.9$  Hz, 1H), 4.17 (d,  $J = 2.8$  Hz, 1H), 4.11 (d,  $J = 3.4$  Hz, 1H), 4.03 (s, 6H), 3.95–3.36 (m, 72H), 2.69 (dd,  $J = 12.4, 4.6$  Hz, 1H), 2.60 (dd,  $J = 12.4, 4.7$  Hz, 1H), 2.01–1.91 (m, 12H), 1.73 (t,  $J = 12.2$  Hz, 1H), 1.65 (t,  $J = 12.2$  Hz, 1H). MALDI-MS:  $C_{84}H_{138}N_6O_{64}$ , calc. for 2254.7728, found  $[M-2H + Na]^-$  2276.097.

Compound **103**, white power (1.21 mg).  $^1H$  NMR (600 MHz,  $D_2O$ )  $\delta$  5.10 (d,  $J = 2.6$  Hz, 1H), 5.03 (s, 1H), 4.86 (s, 1H), 4.52 (t,  $J = 6.4$  Hz, 2H), 4.47 (d,  $J = 8.4$  Hz, 1H), 4.37 (d,  $J = 7.9$  Hz, 1H), 4.17 (s, 1H), 4.10 (dd,  $J = 3.4, 1.6$  Hz, 1H), 4.03 (s, 3H), 3.95–3.32 (m, 56H), 2.60 (dd,  $J = 12.3, 4.7$  Hz, 1H), 2.02–1.91 (m, 12H), 1.65 (t,  $J = 12.2$  Hz, 1H). MALDI-MS:  $C_{67}H_{111}N_5O_{50}$ , calc. for 1785.6297, found  $[M-H]^-$  1784.867.

Compound **104**, white power (0.87 mg).  $^1H$  NMR (600 MHz,  $D_2O$ )  $\delta$  5.10 (d,  $J = 2.7$  Hz, 1H), 5.03 (s, 1H), 4.86 (d,  $J = 1.8$  Hz, 1H), 4.55–4.48 (m, 3H), 4.37 (t,  $J = 7.8$  Hz, 2H), 4.16 (d,  $J = 2.5$  Hz, 1H), 4.11 (dd,  $J = 3.4, 1.6$  Hz, 1H), 4.03 (d,  $J = 3.6$  Hz, 3H), 3.95–3.37 (m, 55H), 2.60 (dd,  $J = 12.4, 4.7$  Hz, 1H), 2.03–1.93 (m, 12H), 1.65 (t,  $J = 12.2$  Hz, 1H), 1.24 (d,  $J = 6.9$  Hz, 1H). MALDI-MS:  $C_{73}H_{121}N_5O_{55}$ , calc. for 1947.6825, found  $[M-H]^-$  1947.017.

Compound **105**, white power (0.59 mg).  $^1H$  NMR (600 MHz,  $D_2O$ )  $\delta$  5.10 (d,  $J = 2.8$  Hz, 1H), 5.05 (s, 1H), 4.86 (s, 1H), 4.56–4.48 (m, 3H), 4.39–4.32 (m, 2H), 4.17 (s, 1H), 4.11 (dd,  $J = 3.5, 1.6$  Hz, 1H), 4.03 (s, 3H), 3.95–3.38 (m, 58H), 2.59 (td,  $J = 12.9, 4.6$  Hz, 2H), 2.04–1.90 (m, 12H), 1.64 (td,  $J = 12.2, 9.2$  Hz, 2H). MALDI-MS:  $C_{84}H_{138}N_6O_{63}$ , calc. for 2238.7779, found  $[M-2H + Na]^-$  2260.103.

Compound **106**, white power (0.43 mg).  $^1H$  NMR (600 MHz,  $D_2O$ )  $\delta$  5.10 (d,  $J = 2.7$  Hz, 1H), 5.03 (s, 1H), 4.86 (s, 1H), 4.54–4.44 (m, 4H), 4.37 (d,  $J = 7.9$  Hz, 1H), 4.17 (d,  $J = 2.5$  Hz, 1H), 4.14–4.09 (m, 1H), 4.03 (s, 6H), 3.94–3.38 (m, 40H), 2.69 (dd,  $J = 12.4, 4.6$  Hz, 1H), 2.60 (dd,  $J = 12.4, 4.7$  Hz, 1H), 2.01–1.94 (m, 12H), 1.73 (t,  $J = 12.2$  Hz, 1H), 1.65 (t,  $J = 12.2$  Hz, 1H). MALDI-MS:  $C_{84}H_{138}N_6O_{64}$ , calc. for 2254.7728, found  $[M-2H + Na]^-$  2276.203.

## Glycan Derivatization and Quantification

All synthesized glycans with free reducing-end were derivatized by reductive amination using 2-amino-*N*-(2-aminoethyl)-benzamide (AEAB) as previously described (Song et al., 2009). Labeled glycans were further purified by HPLC to homogeneity using a porous graphitic carbon column (5  $\mu$ m, 4.6 mm  $\times$  150 mm) under a gradient running condition (solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile; flow rate: 1 ml/min, B%: 15–45% in 30 min) and monitored by UV absorbance at 330 nm. Product-containing fractions were pooled and

lyophilized. The quantifications of AEAB-labeled glycans were conducted as previously described (Li et al., 2019).

## Neu5Gc-*N*-glycan Microarray Fabrication

The AEAB labeled-glycans were prepared at a concentration of 100  $\mu$ M in the printing buffer (150 mM phosphate, pH 8.5), and printed on multivalent NHS-derivatized microscope-glass slides (Z Biotech, LLC), each for 400 pL in replicates of six, as described previously (Heimburg-Molinari et al., 2011). Noncontact printing was performed at room temperature with a humidity of 60% by a sciFLEXARRAYER S3 spotter (Scienion) with two PDC 80 Piezo Dispense Capillary. After overnight dehumidification under room temperature, the slides were washed with MilliQ water and subsequently blocked with 50 mM ethanolamine in 100 mM Tris-HCl (pH 9.0) for 2 h. The blocked slides were then washed with MilliQ water twice, dried, and stored desiccated at  $-20^\circ\text{C}$  until use.

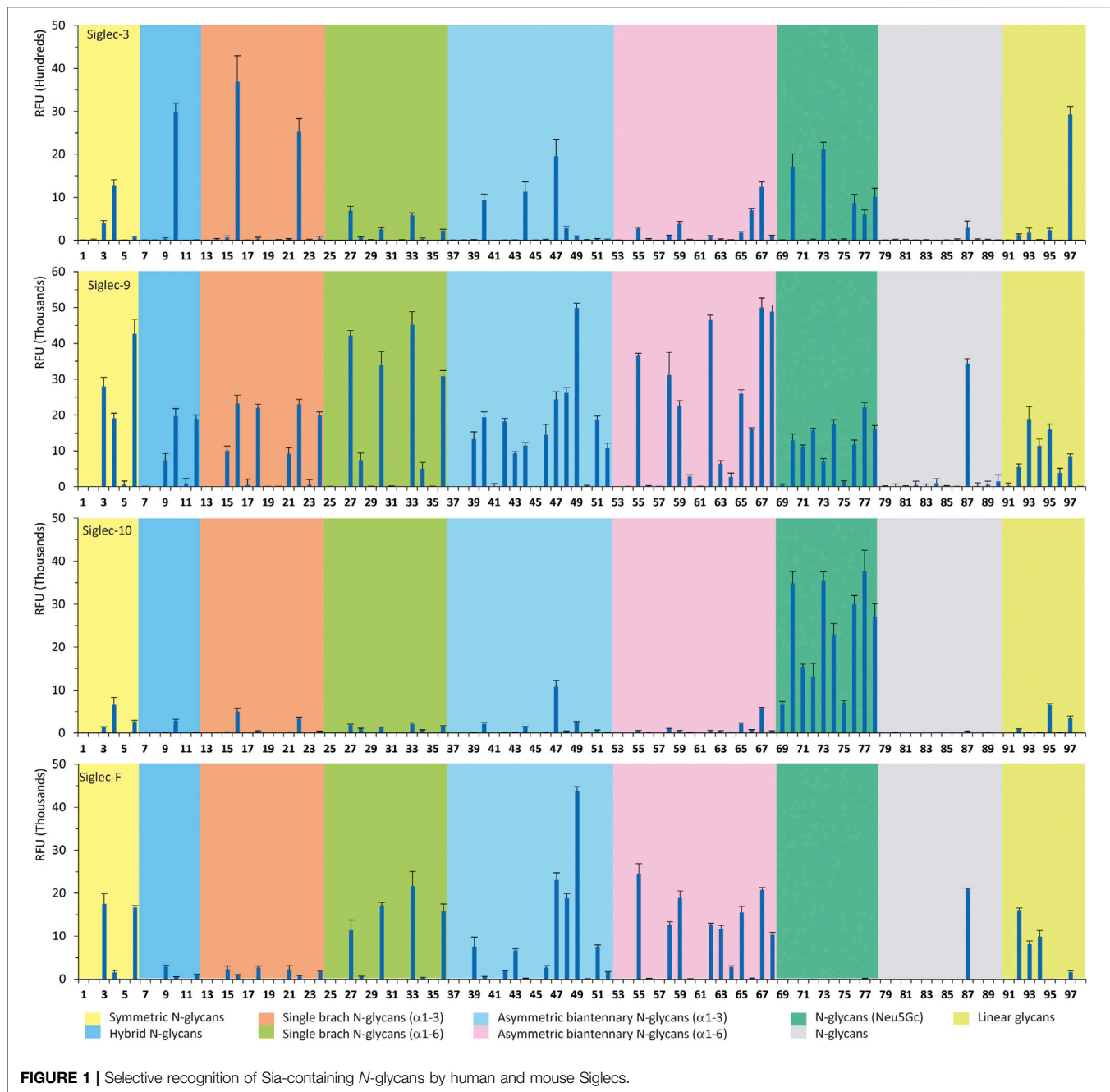
## Microarray Assay of Siglecs With *N*-Glycan Microarrays

The 98 *N*-glycan microarray slide (Li et al., 2019) was fitted with a ProPlate 8-well microarray module (Sigma-Aldrich), and the subarrays were then rehydrated for 10 min with 200  $\mu$ L of Buffer TSMTB (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM  $CaCl_2$ , 2 mM  $MgCl_2$ , 0.05% (v/v) Tween-20, and 1% (w/v) BSA) at room temperature. Then, the buffer was drained and 200  $\mu$ L of Siglec-3, -8, -9, -10, and -F (R&D Systems) (20  $\mu$ g/ml) in TSMTB were added into each subarray, sealed, and incubated at room temperature for 1 h with gentle shaking. Slides were then washed with Buffer TSMT (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM  $CaCl_2$ , 2 mM  $MgCl_2$ , and 0.05% (v/v) Tween-20) for four times. Next, 200  $\mu$ L of 5  $\mu$ g/ml goat anti-human IgG Fc antibody cross-adsorbed, DyLight<sup>®</sup> 650 (Thermo Fisher) was added into each subarray, sealed, and incubated at room temperature for 1 h with gentle shaking. Finally, slides were washed with TSMT, TSM (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM  $CaCl_2$ , and 2 mM  $MgCl_2$ ) and MilliQ water, four times for each buffer, respectively, and dried by brief centrifugation. Slides were scanned at a resolution of 10  $\mu$ m using a Genepix 4100 A microarray scanner (Molecular Devices Corp) with 500 or 600 PMT gains and 80% power. Image analyses were carried out using Genepix Pro 6.0 as previously reported (Li et al., 2019). Spots were defined as circular features with a variable radius as determined by the Genepix scanning software, and local background subtraction was performed. Similarly, Siglec-10 was analyzed using the newly fabricated Neu5Gc-*N*-glycan array at concentrations of 1 and 5  $\mu$ g/ml.

## Biolayer Interferometry Receptor Binding Assay and Data Analysis

The AEAB-labeled glycan **102** and **106** were labeled with Biotin by using the reagent EZ-Link<sup>™</sup> NHS-Biotin (Thermo Fisher). In detail, 1 mM AEAB-labeled glycan was incubated with 10 mM NHS-Biotin at room temperature for 10 min. Then, labeled glycans were purified by HPLC to homogeneity using an





ODS4 column (5  $\mu$ m, 4.6 mm  $\times$  150 mm) under a gradient running condition (solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile; flow rate: 1 ml/min, B%: 5–50% in 30 min), monitored by UV absorbance at 330 nm. Product-containing fractions were pooled and lyophilized for storage. The purified Biotin-labeled glycans were quantified by HPLC as described above.

Avidities were measured by biolayer interferometry using an Octet RED instrument (Pall FortéBio, Fremont, CA, United States). The prepared biotinylated glycans were preloaded onto streptavidin-coated biosensors at up to 100 nM for 3 min in 1 $\times$  kinetic buffer (Pall FortéBio, Menlo Park, CA,

United States). Siglec-10 was diluted to concentrations of 1  $\mu$ M, 500 nM, and 250 nM with 1 $\times$  kinetic buffer, respectively. The glycan-loaded biosensors were submerged in wells containing different concentrations of Siglec-10 for 5 min followed by 15 min of dissociation in 1 $\times$  kinetic buffer at 25°C with the orbital shake speed of 1000 rpm. As a reference control for subtraction, glycan-loaded biosensors were also dipped in wells containing 1 $\times$  kinetic buffer. The binding kinetics data were processed by the ForteBio data analysis software (version 11.1). The association and dissociation curves were fitted, and the avidity values were calculated by using a heterogeneous ligand (2:1) model.

## RESULTS

### Fine Specificity of Human Siglecs Toward the 98 *N*-Glycan Microarray

The primary glycan ligands of Siglecs were reported and well summarized (Macauley et al., 2014; Duan and Paulson, 2020). Human Siglec-3, -8, -9, and -10 and mouse Siglec-F recognize Neu5Ac $\alpha$ 2-6Gal $\beta$ 1-4GlcNAc (Ac6LN) and/or Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc (Ac3LN), which are often identified as terminal epitopes on complex glycans found on mammalian cells. To explore fine binding specificities of Siglecs, these Siglecs were analyzed against a previously fabricated microarray containing 98 structurally well-defined complex glycans (**Supplementary Figure S1**) (Li et al., 2019).

As shown in **Figure 1**, human Siglec-3 gave lower binding signals toward *N*-glycans compared with Siglec-9, -10, and -F. Siglec-3, which is found on myeloid cells, is associated with acute myeloid leukemia (AML) and Alzheimer's disease (Freeman et al., 1995; Zhao, 2019) and was reported to prefer the Ac6LN trisaccharide, plus relatively weak affinity to Gc6LN (Neu5Gc $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc) and Ac3LN (Blixt et al., 2003). Recently, Rodrigues et al. (2020) reported that Siglec-3 could recognize both  $\alpha$ 2-3 and  $\alpha$ 2-6sialosides in solution and on cells. This is consistent with our results that it bound to *N*-glycans with terminal epitopes Ac3LN (compound 3), Ac6LN (compound 4), Gc3LN (compound 77), and Gc6LN (compound 78). In addition, the RFUs of Siglec-3 to  $\alpha$ 2-6sialosides (4 and 78) are higher than those of  $\alpha$ 2-3sialosides (3 and 77), again consistent with a previous report toward O-mannosyl glycans (Meng et al., 2018). In addition, a slight preference toward Neu5Ac over Neu5Gc was observed, as binding signals of glycans with Neu5Ac residues (4, 10, and 16) were greater than those of their Neu5Gc-containing counterparts (70, 73, and 78). Furthermore, high to moderate bindings were observed toward *N*-glycans carrying Ac3LN on the  $\alpha$ 1-3 branch (16, 33, 40, 44, 47, 66, and 67), whereas no meaningful binding signals were observed to their positional isomers (28, 34, 50, 51, 56, 60, and 63). These data suggested that Siglec-3 had an apparent preference toward the  $\alpha$ 1-3 branch when terminal epitope Ac6LN serves as the binding ligand. On the other hand, Siglec-3 exhibited an opposite branch preference toward the  $\alpha$ 1-6 branch when terminal epitope Ac3LN served as the binding ligand; for example, it bound to 27, 33, 55, and 59, but failed to bind their positional isomers (15, 21, 39, and 43). Such a unique terminal epitope-dependent branch preference was double evidenced by strong binding to 47, which presents terminal epitopes on preferred branches (Ac6LN on the  $\alpha$ 1-3 branch and Ac3LN on the  $\alpha$ 1-6 branch), but no binding to 63 that presents terminal epitopes on nonpreferred branches (Ac3LN on the  $\alpha$ 1-3 branch and Ac6LN on the  $\alpha$ 1-6 branch).

Siglec-9 was reported to bind to both  $\alpha$ 2-3 and  $\alpha$ 2-6sialosides, with a high affinity to epitope Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4(6-sulfo)GlcNAc (Zhang et al., 2000; Rillahan et al., 2012; Duan and Paulson, 2020). Our microarray results are consistent with previous reports as all related *N*-glycans showed binding signals. *N*-glycans with sialyl Lewis X (sLe<sup>X</sup>) epitopes showed the highest binding signals, including compounds 6, 27, 33, 49, 67, and 68. Additionally, glycans with the Ac3LN epitope (3, 27, and 33) exhibited higher bindings than those with the Ac6LN epitope (4, 28, and 34). Interestingly, the same

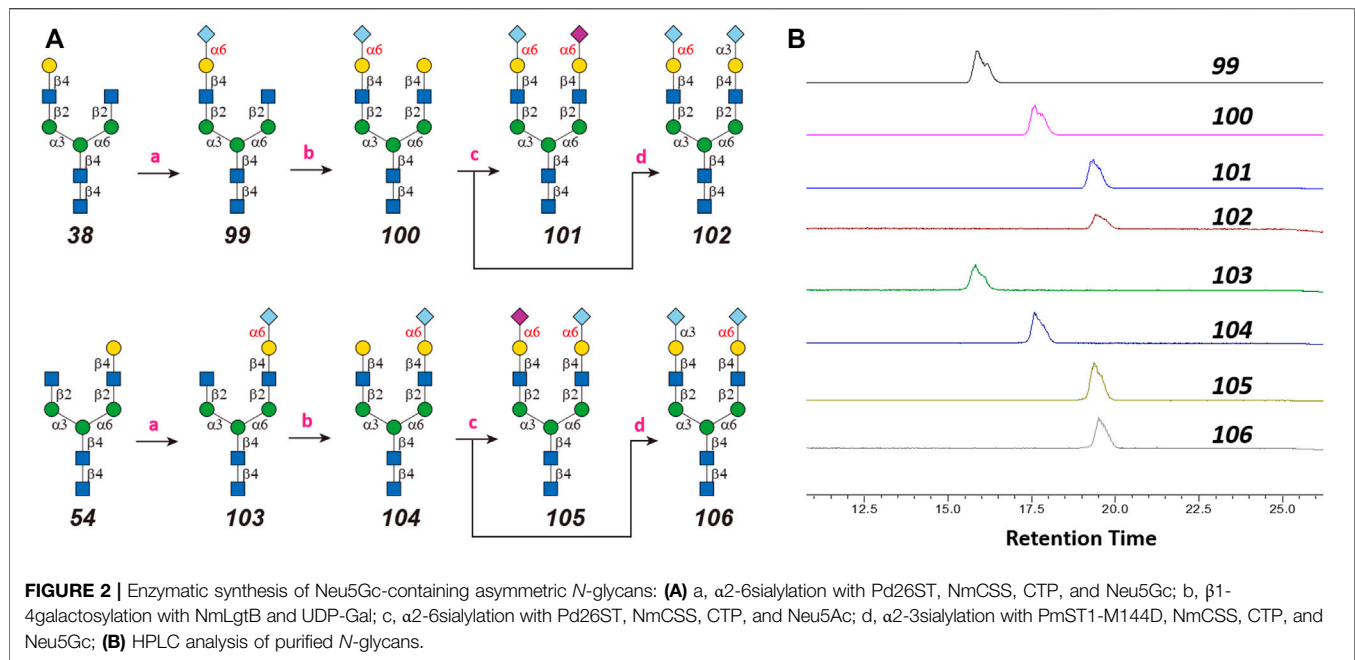
terminal epitope-dependent branch preference for Siglec-3 was also observed for Siglec-9. When bound to glycans with the terminal epitope Ac6LN, Siglec-9 showed an apparent preference toward the  $\alpha$ 1-3 branch (16, 22, 40, and 44) over the  $\alpha$ 1-6 branch (28, 34, 66, and 60) of *N*-glycans. In contrast, an opposite branch preference was found when bound to glycans with the terminal epitope Ac3LN and sLe<sup>X</sup>. Lastly, it is observed that Siglec-9 has a slight preference to Neu5Ac-containing *N*-glycans (9, 10, 12, 15, 16, 18, 39, and 40) over their Neu5Gc-containing counterparts (69–76).

The specificity of Siglec-10 was previously profiled as having a high affinity to Gc6LN, with moderate and weak affinity to Ac6LN and Ac3LN, respectively (Crocker et al., 2007). In our array, as expected, Siglec-10 exhibited strong bindings to the majority of Neu5Gc-terminated *N*-glycans (69–78). In addition, bindings of Siglec-10 toward various glycan ligands showed gradient diminished signals (Gc6LN >> Gc3LN > Ac6LN >> Ac3LN). For example, the binding signals of Siglec-10 to complex type *N*-glycans 76 (Gc6LN) >> 75 (Gc3LC) > 40 (Ac6LC) >> 39 (Ac3LN) and hybrid type *N*-glycans 70 (Gc6LN) >> 69 (Gc3LC) > 10 (Ac6LC) >> 9 (Ac3LN) clearly showed this trend. Siglec-10 also exhibited a terminal epitope-dependent branch preference toward asymmetric *N*-glycans. For example, high to moderate bindings were observed when epitope Ac6LN was presented on the terminal of the  $\alpha$ 1-3 branch (16, 22, 66, 40, and 44), while no meaningful signals could be observed to their positional isomers where Ac6LN was presented on the  $\alpha$ 1-6 branch (28, 34, 50, 56, and 60). This result suggested that Siglec-10 had an apparent preference toward the  $\alpha$ 1-3 branch when  $\alpha$ 2-6sialylated glycans served as ligands. In contrast, Siglec-10 exhibited an opposite branch preference toward the  $\alpha$ 1-6 branch when terminal epitope Ac3LN or sLe<sup>X</sup> served as the binding ligand. This is evidenced by relatively weak bindings toward compounds 27, 30, 33, and 36, which present Ac3LN or sLe<sup>X</sup> on the  $\alpha$ 1-6 branch but no bindings to their positional isomers. This terminal epitope-dependent branch preference is identical to that of Siglec-3 and Siglec-9.

Human Siglec-8 did not bind to any glycans on the array (data not shown), which is consistent with previous observations that Siglec-8 specifically recognizes Neu5Ac $\alpha$ 2-3 (6-sulfo) Gal $\beta$ 1-4GlcNAc (6-sulfo-sLe<sup>X</sup>) (Bochner et al., 2005). Mouse Siglec-F is a functional paralogue of human Siglec-8, and it was reported to bind to Ac3LN and 6'-sulfo-sLe<sup>X</sup> (Tateno et al., 2005). As depicted in **Figure 1**, Siglec-F could recognize *N*-glycans with Ac3LN and sLe<sup>X</sup> epitopes, such as 3, 6, and 49, whereas no binding was observed toward any Neu5Gc-containing glycans, suggesting a strict preference toward Neu5Ac. In addition, Siglec-F showed an apparent  $\alpha$ 1-6 branch preference. For example, high binding signals were observed for glycans 27, 30, 33, and 36, but very low bindings were observed for their positional isomers 15, 18, 21, and 24. Interestingly, compound 49, which contains sLe<sup>X</sup> on the  $\alpha$ 1-3 branch and Ac3LN on the  $\alpha$ 1-3 branch, showed the strongest binding signals.

### Chemoenzymatic Synthesis of Neu5Gc-Containing *N*-Glycans

One interesting observation is that Siglec-10 showed high binding to an asymmetric *N*-glycan 47 but no binding to its positional isomer 63 (**Figure 1**). Such a dramatic binding divergence can be



explained by its terminal epitope-dependent branch preference, as both ligands (Ac3LN and Ac6LN) on **47** are located on the terminal of favored branches, whereas both ligands are located on the unfavored branches of **63**. Because Siglec-10 strongly prefers Neu5Gc-containing *N*-glycans, we speculate that a Neu5Gc-modified counterpart of **47** (**Figure 2A**, compound **102**) may be of higher affinity and a most favorable *N*-glycan ligand of Siglec-10. To test this hypothesis and to further validate the terminal epitope-dependent branch preference of Siglec-10, we enzymatically synthesized eight Neu5Gc-containing *N*-glycans (**Figure 2A**). In detail, compounds **99** to **102** were assembled starting from previously prepared glycan **38** (Li et al., 2015). First,  $\alpha$ 2-6Neu5Gc was installed onto the  $\alpha$ 1-3 branch to achieve **99** by Pd26ST-catalyzed  $\alpha$ 2-6sialylation in the presence of cytidine-5'-triphosphate (CTP), Neu5Gc, and NmCSS for the *in situ* generation of the sugar donor CMP-Neu5Gc. Then,  $\beta$ 1-4Gal was installed onto the  $\alpha$ 1-6 branch by NmLgtB-catalyzed reaction in the presence of UDP-Gal to provide **100**. The addition of  $\alpha$ 2-6Neu5Ac to the  $\alpha$ 1-6 branch of **100** by Pd26ST then provided **101**. On the other hand, the addition of  $\alpha$ 2-3Neu5Gc to this branch by PmST1-M144D-catalyzed  $\alpha$ 2-3sialylation gave the desired asymmetric *N*-glycan **102**. In the same synthetic manner, another four asymmetric *N*-glycans **103**, **104**, **105**, and **106** were assembled starting from *N*-glycan **54**. All compounds were purified and characterized by HPLC (**Figure 2B**), mass spectrometry, and NMR (supporting information).

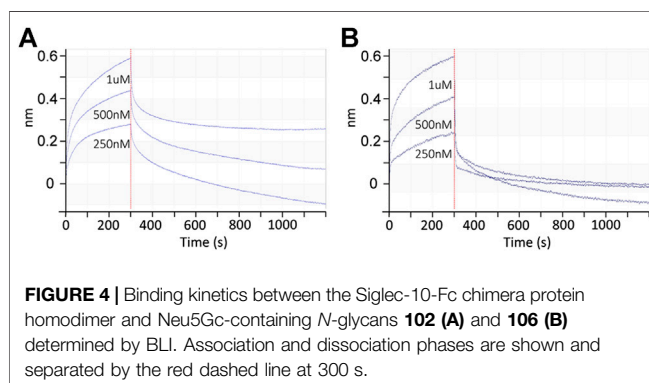
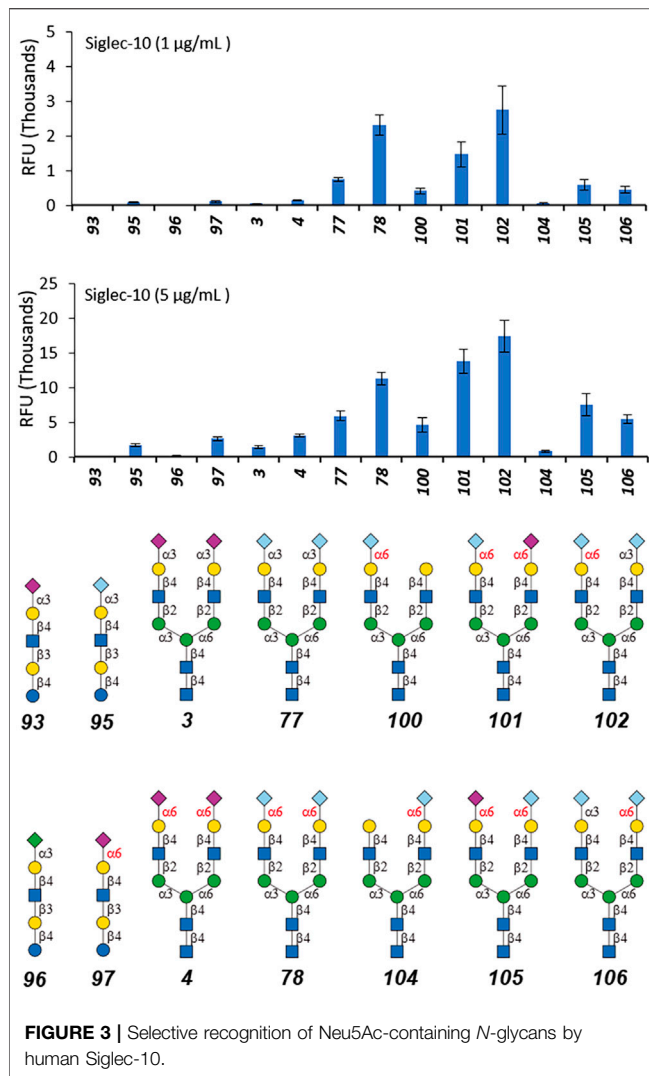
## Neu5Gc *N*-Glycan Microarray Fabrication and Assay With Human Siglec-10

The Neu5Gc-containing *N*-glycans were labeled with AEAB to provide an amino group for microarray fabrication as previously reported (Li et al., 2019). The Neu5Gc *N*-glycan microarray was

then constructed on NHS glass slides with four additional sialylated *N*-glycans (**3**, **4**, **77**, and **78**) and four sialylated linear glycans (**93**, **95**, **96**, and **97**). The recognition by Siglec-10 was then assayed at the concentrations of 1  $\mu$ g/ml and 5  $\mu$ g/ml (**Figure 3**). As shown in **Figure 3**, di-sialylated glycans (**78**, **101**, **102**, and **105**) showed high RFU compared with mono-sialylated glycans (**100** and **104**) and linear glycan (**93**, **95**, **96**, and **97**). The preferences of Siglec-10 toward Neu5Gc (**77**, **78**) over Neu5Ac (**3** and **4**) and  $\alpha$ 2-6sialosides (**4**, **78**) over  $\alpha$ 2-3sialosides (**3**, **77**) were further confirmed by this focused array. In addition, when terminal epitope Gc6LN served as the binding ligand, Siglec-10 preferred the  $\alpha$ 1-3 branch (**100**) over the  $\alpha$ 1-6 branch (**104**). And as expected, Siglec-10 showed the highest binding signal to glycan **102** (the Neu5Gc modified counterpart of **47**), whereas its positional isomer **106** only showed comparable bindings as that of mono-sialylated **100**. These results further confirmed the terminal epitope-dependent branch preference; that is, Siglec-10 prefers  $\alpha$ 2-6sialosides on the  $\alpha$ 1-3 branch and  $\alpha$ 2-3sialosides on the  $\alpha$ 1-6 branch of *N*-glycans. Interesting, the binding signals of Siglec-10 toward the four  $\alpha$ 2-6sialylated *N*-glycans with Neu5Ac/Neu5Gc chimeras (**4**, Neu5Ac on both branches; **78**, Neu5Gc on both branches; **101**, Neu5Ac on  $\alpha$ 1-6 branch, Neu5Gc on  $\alpha$ 1-3 branch; **105**, Neu5Gc on  $\alpha$ 1-6 branch, and Neu5Ac on  $\alpha$ 1-3 branch) are distinct, indicating that minor structural divergence in complex glycan may cause substantial changes in glycan-protein interactions.

## Avidity of Siglec-10 to *N*-Glycans **102** and **106**

As shown in **Figure 3**, the binding signals of Siglec-10 to **102** is around 5-fold stronger than to its positional isomer **106** and 20-fold stronger than to linear glycans **97**, suggesting compound **102** as a potential



high-affinity ligand of Siglec-10. The avidity of Siglec-10 toward **102** and **106** was thus measured by biolayer interferometry (BLI). AEAB labeled **102** and **106** were further conjugated with NHS-Biotin and purified with HPLC, and then immobilized onto streptavidin-coated biosensors for BLI assay (Figure 4). The association and dissociation

curves were fitted, and the avidity values were calculated with the consideration of the bivalency of the Siglec-10-Fc chimera protein. The avidity values of Siglec-10 toward **102** and **106** were 0.11  $\mu$ M and 7.34  $\mu$ M, respectively, indicating a 67-fold higher avidity of **102** than **106**. The result further confirmed the terminal epitope-dependent branch preference and revealed a high avidity glycan-binding partner (**102**) of human Siglec-10.

## DISCUSSION AND CONCLUSION


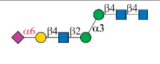
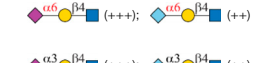
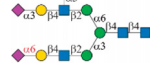

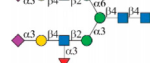
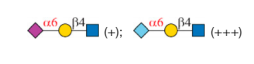

Siglecs are attractive therapeutic targets and several related antibody-based therapies had been developed for the treatment of immune-related diseases. In certain applications, glycan ligands have an advantage over antibodies, such as their ability to dissociate from their target once endocytosed. However, glycan-based therapeutic strategies for cargo delivery and immunomodulation are underinvestigated due to the lack of suitable ligands (Angata et al., 2015). A comprehensive understanding of glycan recognition details by Siglecs is essential toward the discovery and designing of efficient ligands. In fact, recent advances in glycobiology have prompted such applications. For example, high specific efficient *N*-glycan ligands with chemical modifications toward Siglec-2 were reported (Peng and Paulson, 2017). Conjugates of toxins with this novel ligand could be efficiently internalized *via* Siglec-2, resulting in the killing of B-cell lymphoma cells.







In this study, we screened binding profiles of Siglec-3, -9, -10, and -F against a comprehensive *N*-glycan microarray to reveal glycan recognition details of Siglecs (Table 1). The results showed a surprising terminal epitope-dependent branch preference toward *N*-glycans by Siglec-3, -9, and -10. These Siglecs prefer the  $\alpha$ 1-3 branch of *N*-glycans when  $\alpha$ 2-6sialylated epitopes serve as binding ligands, while they have an opposite preference to the  $\alpha$ 1-6 branch when  $\alpha$ 2-3sialylated epitopes serve as ligands. Such a feature could assist in designing high-affinity binding partners of Siglecs. For example, we designed and synthesized an asymmetric *N*-glycan (**102**) with much higher avidity than its positional isomer toward Siglec-10. Note that recombinant Siglec-Fc chimera proteins in the form of disulfide-linked homodimer were used in this study instead of native Siglecs. Even though such chimera proteins were widely used to reveal the glycan recognition of Siglecs and other human GBPs (Blixt et al., 2003; Bochner et al., 2005; Campanero-Rhodes et al., 2006; Rillahan et al., 2012; Rillahan et al., 2013; Gao et al., 2019b; Rodrigues et al., 2020), the nonnatural bivalent form could possibly influence their fine specificity toward glycan-binding partners.

High-avidity binding partners of Siglecs could lead to extensive academic and clinical implementations. For example, tumor cells can escape the surveillance of the immune system *via* inhibition of immune cells through immune checkpoints and their ligands. A promising therapeutic approach for cancer is to block these immune checkpoints, for example, the programmed cell death ligand 1 (PD-L1) and cytotoxic T lymphocyte-associated protein 4 (CTLA-4) (Leach et al., 1996; Topalian et al., 2012). A recent report showed that CD24–Siglec-10 interaction is an innate immune checkpoint that is essential for mediating antitumor immunity and can promote tumor immune escape. The modulation of this interaction is expected to become a new target for tumor therapy (Barkal et al., 2019). The



**TABLE 1 |** Binding specificity of Siglecs toward *N*-glycans observed in this study. Recombinant Siglec-Fc chimera proteins were used in this study.

Siglec	Ligand (preference)	Branch Preference	Strongest <i>N</i> -glycan Binder
Siglec-3		$\alpha$ 1-3 branch $\alpha$ 1-6 branch	
Siglec-9		$\alpha$ 1-3 branch $\alpha$ 1-6 branch	
Siglec-10		$\alpha$ 1-3 branch $\alpha$ 1-6 branch	
Siglec-F		$\alpha$ 1-3 branch	

Symbols:  GlcNAc  Man  Gal  L-Fuc  Neu5Ac  Neu5Gc

high sialylated CD24 that overexpressed on tumor cells functions as the main ligand of Siglec-10. It induces the inhibition of the immune system and promotes tumor immune escape. Additionally, CD24–Siglec-10 interaction could suppress the immune response to the danger-associated molecular pattern (DAMP) (Cai et al., 2009; Rillahan et al., 2012). It is thus tempting to speculate that the strong Siglec-10 binding partner **102**, with or without further modification, may serve as an invaluable reagent to block this immune checkpoint.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

LL conceived and designed the project. SW,CC and LL performed glycan synthesis and glycan microarray assays, MG and DL performed

kinetic assay, SW wrote the manuscript, and LL and XW revised the manuscript, which was edited and approved by all authors.

## FUNDING

This work was supported by the National Institutes of Health (Grant No. U54HL142019 and U01GM125288 to LL). XW and MG were partially supported by the National Institutes of Health (Grant No. R21AI144433 to XW).

## ACKNOWLEDGMENTS

We thank Z biotech LLC (Aurora, CO) for microarray fabrication (supported by NIH R43GM123820).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.645999/full#supplementary-material>.

## REFERENCES

- Angata, T., Nycholat, C. M., and Macauley, M. S. (2015). Therapeutic targeting of siglecs using antibody- and glycan-based approaches. *Trends Pharmacol. Sci.* 36 (10), 645–660. doi:10.1016/j.tips.2015.06.008
- Barkal, A. A., Brewer, R. E., Markovic, M., Kowarsky, M., Barkal, S. A., Zaro, B. W., et al. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature* 572 (7769), 392–396. doi:10.1038/s41586-019-1456-0
- Blixt, O., Collins, B. E., van den Nieuwenhof, I. M., Crocker, P. R., and Paulson, J. C. (2003). Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated

glycoprotein. *J. Biol. Chem.* 278 (33), 31007–31019. doi:10.1074/jbc.M304331200

- Bochner, B. S., Alvarez, R. A., Mehta, P., Bovin, N. V., Blixt, O., White, J. R., et al. (2005). Glycan array screening reveals a candidate ligand for siglec-8. *J. Biol. Chem.* 280 (6), 4307–4312. doi:10.1074/jbc.M412378200

- Cai, L., Guan, W., Kitaoka, M., Shen, J., Xia, C., Chen, W., et al. (2009). A chemoenzymatic route to N-acetylglucosamine-1-phosphate analogues: substrate specificity investigations of N-acetylhexosamine 1-kinase. *Chem. Commun. (Camb)* 20, 2944–2946. doi:10.1039/b904853g

- Campanero-Rhodes, M. A., Childs, R. A., Kiso, M., Komba, S., Le Narvor, C., Warren, J., et al. (2006). Carbohydrate microarrays reveal sulphation as a modulator of siglec binding. *Biochem. Biophys. Res. Commun.* 344 (4), 1141–1146. doi:10.1016/j.bbrc.2006.03.223

- Crocker, P. R., Paulson, J. C., and Varki, A. (2007). Siglecs and their roles in the immune system. *Nat. Rev. Immunol.* 7 (4), 255–266. doi:10.1038/nri2056
- Duan, S., and Paulson, J. C. (2020). Siglecs as immune cell checkpoints in disease. *Annu. Rev. Immunol.* 38 (1), 365–395. doi:10.1146/annurev-immunol-102419-035900
- Freeman, S. D., Kelm, S., Barber, E. K., and Crocker, P. R. (1995). Characterization of CD33 as a new member of the sialoadhesin family of cellular interaction molecules. *Blood* 85 (8), 2005–2012. doi:10.1182/blood.v85.8.2005.bloodjournal8582005
- Fukui, S., Feizi, T., Galustian, C., Lawson, A. M., and Chai, W. (2002). Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nat. Biotechnol.* 20 (10), 1011–1017. doi:10.1038/nbt735
- Gao, C., Hanes, M. S., Byrd-Leotis, L. A., Wei, M., Jia, N., Kardish, R. J., et al. (2019a). Unique binding specificities of proteins toward isomeric asparagine-linked glycans. *Cell Chem. Biol.* 26 (4), 535–547. doi:10.1016/j.chembiol.2019.01.002
- Gao, C., Wei, M., McKittrick, T. R., McQuillan, A. M., Heimburg-Molinaro, J., and Cummings, R. D. (2019b). Glycan microarrays as chemical tools for identifying glycan recognition by immune proteins. *Front. Chem.* 7, 833. doi:10.3389/fchem.2019.00833
- Heimburg-Molinaro, J., Song, X., Smith, D. F., and Cummings, R. D. (2011). Preparation and analysis of glycan microarrays. *Curr. Protoc. Protein Sci.* Chapter 12, Unit12.10. doi:10.1002/0471140864.ps1210s64
- Laszlo, G. S., Estey, E. H., and Walter, R. B. (2014). The past and future of CD33 as therapeutic target in acute myeloid leukemia. *Blood Rev.* 28 (4), 143–153. doi:10.1016/j.blre.2014.04.001
- Lau, K., Thon, V., Yu, H., Ding, L., Chen, Y., Muthana, M. M., et al. (2010). Highly efficient chemoenzymatic synthesis of  $\beta$ 1-4-linked galactosides with promiscuous bacterial  $\beta$ 1-4-galactosyltransferases. *Chem. Commun.* 46 (33), 6066–6068. doi:10.1039/C0CC01381A
- Läubli, H., and Varki, A. (2020). Sialic acid-binding immunoglobulin-like lectins (siglecs) detect self-associated molecular patterns to regulate immune responses. *Cell. Mol. Life Sci.* 77 (4), 593–605. doi:10.1007/s00018-019-03288-x
- Leach, D. R., Krummel, M. F., and Allison, J. P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. *Science* 271 (5256), 1734–1736. doi:10.1126/science.271.5256.1734
- Li, L., Liu, Y., Ma, C., Qu, J., Calderon, A. D., Wu, B., et al. (2015). Efficient chemoenzymatic synthesis of an N-glycan isomer library. *Chem. Sci.* 6 (10), 5652–5661. doi:10.1039/C5SC02025E
- Li, L., Guan, W., Zhang, G., Wu, Z., Yu, H., Chen, X., et al. (2019). Microarray analyses of closely related glycoforms reveal different accessibilities of glycan determinants on N-glycan branches. *Glycobiology* 30 (5), 334–345. doi:10.1093/glycob/cwz100
- Macauley, M. S., Crocker, P. R., and Paulson, J. C. (2014). Siglec-mediated regulation of immune cell function in disease. *Nat. Rev. Immunol.* 14 (10), 653–666. doi:10.1038/nri3737
- Meng, C., Sasmal, A., Zhang, Y., Gao, T., Liu, C. C., Khan, N., et al. (2018). Chemoenzymatic assembly of mammalian O-mannose glycans. *Angew. Chem. Int. Ed. Engl.* 57 (29), 9003–9007. doi:10.1002/anie.201804373
- Muthana, M. M., Qu, J., Li, Y., Zhang, L., Yu, H., Ding, L., et al. (2012). Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from *Bifidobacterium longum* (BLUSP). *Chem. Commun. (Camb)* 48 (21), 2728–2730. doi:10.1039/c2cc17577k
- Palma, A. S., and Chai, W. (2019). “Chapter 18 glycan microarrays with semi-synthetic neoglycoconjugate probes in understanding glycobiology,” in *Synthetic glycomics*, Editors L. Li, P. G. Wang, and W. Guan (Croydon, London, United Kingdom: The Royal Society of Chemistry), 421–446. doi:10.1039/9781788016575-00421
- Peng, W., and Paulson, J. C. (2017). CD22 ligands on a natural N-glycan scaffold efficiently deliver toxins to B-lymphoma cells. *J. Am. Chem. Soc.* 139 (36), 12450–12458. doi:10.1021/jacs.7b03208
- Rillahan, C. D., Schwartz, E., McBride, R., Fokin, V. V., and Paulson, J. C. (2012). Click and pick: identification of sialoside analogues for siglec-based cell targeting. *Angew. Chem. Int. Ed. Engl.* 51 (44), 11014–11018. doi:10.1002/anie.201205831
- Rillahan, C. D., Schwartz, E., Rademacher, C., McBride, R., Rangarajan, J., Fokin, V. V., et al. (2013). On-chip synthesis and screening of a sialoside library yields a high affinity ligand for siglec-7. *ACS Chem. Biol.* 8 (7), 1417–1422. doi:10.1021/cb400125w
- Rodrigues, E., Jung, J., Park, H., Loo, C., Soukhtehzari, S., Kitova, E. N., et al. (2009). A versatile soluble siglec scaffold for sensitive and quantitative detection of glycan ligands. *Nat. Commun.* 11 (1), 5091. doi:10.1038/s41467-020-18907-6
- Song, X., Xia, B., Stowell, S. R., Lasanajak, Y., Smith, D. F., and Cummings, R. D. (2009). Novel fluorescent glycan microarray strategy reveals ligands for galectins. *Chem. Biol.* 16 (1), 36–47. doi:10.1016/j.chembiol.2008.11.004
- Sugiarto, G., Lau, K., Qu, J., Li, Y., Lim, S., Mu, S., et al. (2012). A sialyltransferase mutant with decreased donor hydrolysis and reduced sialidase activities for directly sialylating lewisx. *ACS Chem. Biol.* 7 (7), 1232–1240. doi:10.1021/cb300125k
- Tateno, H., Crocker, P. R., and Paulson, J. C. (2005). Mouse siglec-F and human Siglec-8 are functionally convergent paralogs that are selectively expressed on eosinophils and recognize 6'-sulfo-sialyl Lewis X as a preferred glycan ligand. *Glycobiology* 15 (11), 1125–1135. doi:10.1093/glycob/cwi097
- Topalian, S. L., Drake, C. G., and Pardoll, D. M. (2012). Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Curr. Opin. Immunol.* 24 (2), 207–212. doi:10.1016/j.coi.2011.12.009
- Yu, H., Huang, S., Chokhawala, H., Sun, M., Zheng, H., and Chen, X. (2006). Highly efficient chemoenzymatic synthesis of naturally occurring and non-natural alpha-2,6-linked sialosides: a P. damsela alpha-2,6-sialyltransferase with extremely flexible donor-substrate specificity. *Angew. Chem. Int. Ed. Engl.* 45 (24), 3938–3944. doi:10.1002/anie.200600572
- Yu, H., Yu, H., Karpel, R., and Chen, X. (2004). Chemoenzymatic synthesis of CMP-sialic acid derivatives by a one-pot two-enzyme system: comparison of substrate flexibility of three microbial CMP-sialic acid synthetases. *Bioorg. Med. Chem.* 12 (24), 6427–6435. doi:10.1016/j.bmc.2004.09.030
- Zhang, J. Q., Nicoll, G., Jones, C., and Crocker, P. R. (2000). Siglec-9, a novel sialic acid binding member of the immunoglobulin superfamily expressed broadly on human blood leukocytes. *J. Biol. Chem.* 275 (29), 22121–22126. doi:10.1074/jbc.M002788200
- Zhao, L. (2019). CD33 in alzheimer's disease—biology, pathogenesis, and therapeutics: a mini-review. *Gerontology* 65 (4), 323–331. doi:10.1159/000492596

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Chen, Guan, Liu, Wan and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Sulfation Code of Tauopathies: Heparan Sulfate Proteoglycans in the Prion Like Spread of Tau Pathology

Dylan Mah<sup>1</sup>, Jing Zhao<sup>1,2</sup>, Xinyue Liu<sup>1,3</sup>, Fuming Zhang<sup>1</sup>, Jian Liu<sup>4</sup>, Lianchun Wang<sup>5</sup>, Robert Linhardt<sup>1</sup> and Chunyu Wang<sup>1\*</sup>

<sup>1</sup> Department of Biological Sciences, Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, United States, <sup>2</sup> College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, China, <sup>3</sup> Harvard Medical School, Harvard University, Boston, MA, United States, <sup>4</sup> Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, <sup>5</sup> Morsani College of Medicine, University of South Florida, Tampa, FL, United States

## OPEN ACCESS

### Edited by:

Thomas Simonson,  
École Polytechnique, France

### Reviewed by:

Feng Ding,  
Clemson University, United States  
Anirban Chakraborty,  
University of Texas Medical Branch  
at Galveston, United States

### \*Correspondence:

Chunyu Wang  
wangc5@rpi.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 23 February 2021

**Accepted:** 12 April 2021

**Published:** 20 May 2021

### Citation:

Mah D, Zhao J, Liu X, Zhang F,  
Liu J, Wang L, Linhardt R and  
Wang C (2021) The Sulfation Code  
of Tauopathies: Heparan Sulfate  
Proteoglycans in the Prion Like  
Spread of Tau Pathology.  
Front. Mol. Biosci. 8:671458.  
doi: 10.3389/fmolb.2021.671458

Tauopathies are a heterogeneous family of progressive neurodegenerative diseases defined by the appearance of proteinaceous lesions within the brain composed of abnormally folded species of Microtubule Associated Protein Tau (tau). Alzheimer's Disease (AD), the most common tauopathy, is the leading cause of cognitive decline among the elderly and is responsible for more than half of all cases of senile dementia worldwide. The characteristic pathology of many tauopathies—AD included—presents as Neurofibrillary Tangles (NFTs), insoluble inclusions found within the neurons of the central nervous system composed primarily of tau protein arranged into Paired Helical Fibrils (PHFs). The spatial extent of this pathology evolves in a remarkably consistent pattern over the course of disease progression. Among the leading hypotheses which seek to explain the stereotypical progression of tauopathies is the *prion model*, which proposes that the spread of tau pathology is mediated by the transmission of self-propagating tau conformers between cells in a fashion analogous to the mechanism of communicable prion diseases. Protein-glycan interactions between tau and Heparan Sulfate Proteoglycans (HSPGs) have been implicated as a key facilitator in each stage of the prion-like propagation of tau pathology, from the initial secretion of intracellular tau protein into the extracellular matrix, to the uptake of pathogenic tau seeds by cells, and the self-assembly of tau into higher order aggregates. In this review we outline the biochemical basis of the tau-HS interaction and discuss our current understanding of the mechanisms by which these interactions contribute to the propagation of tau pathology in tauopathies, with a particular focus on AD.

**Keywords:** Alzheimer's disease, heparan sulfate, glycobiology, 2-O and 6-O sulfated heparins, tauopathies, 3-O sulfation, prions and prion disease, neurodegenerative diseases

## INTRODUCTION

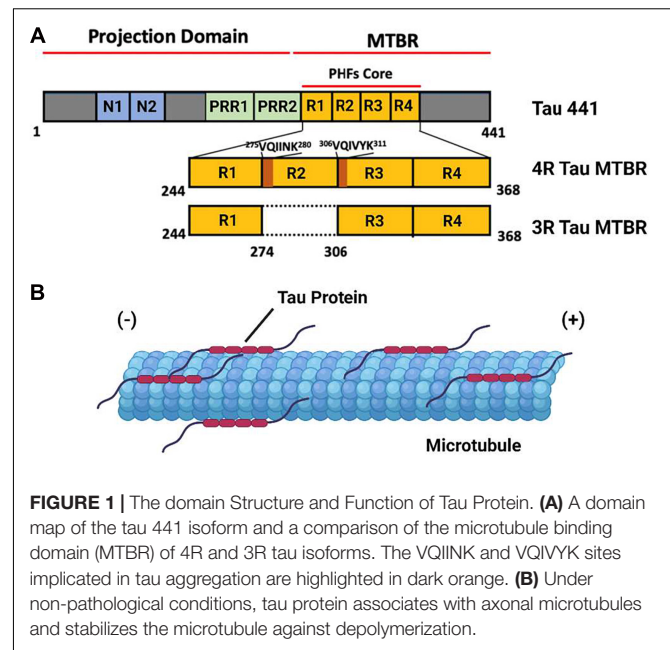
### Tau Protein

Tau protein is the primary constituent of the proteinaceous lesions that characterize tauopathies, a group of debilitating neurodegenerative diseases exemplified by Alzheimer's Disease (AD), the most abundant tauopathy and the leading cause of dementia in the elderly worldwide. Tau is a Microtubule (MT) binding protein encoded by the *MAPT* gene on chromosome 17 (Binder et al., 2005). Six distinct isoforms of this protein are produced in the Central Nervous System (CNS) of adult humans through alternative splicing of *MAPT*, the largest of which consists of a 441 residue polypeptide. Tau 441 consist of an N-terminus projection domain with two inserts (N1 and N2), a Proline Rich Region (PRR) subdivided into PRR1 and PRR2 which contains major tau phosphorylation sites, and an MT Binding Repeat domain (MTBR) composed of four imperfect repeat motifs (R1–R4) that participates in both MT binding and tau aggregation (Figure 1A). The CNS tau isoforms are distinguished by the presence of both (2N), one (1N), or neither (0N) of the N terminus inserts and the presence (4R) or absence (3R) of the second of the four MTBR repeats found in Tau 441, giving rise to a total of six isoforms: 2N4R, 1N4R, 0N4R, 2N3R, 1N3R, and 0N3R. According to this convention, Tau 441 is alternatively referred to as 2N4R tau (Buée et al., 2000; Goedert et al., 1989).

Tau protein is an Intrinsically Disordered Protein (IDP) which lacks a defined secondary or tertiary structure in solution. The binding of tau to a microtubule is mediated by a conformational shift toward a more ordered structure, a feature consistent with the behavior of other IDPs. Under ordinary physiological conditions, tau protein localizes to the axonal segment of the neuronal cytoskeleton, where it interacts with the tubulin heterodimer to stabilize MTs and promote tubulin polymerization (Figure 1B). The function of tau is regulated by a range of post translational modifications, including phosphorylation, acetylation, and methylation, many of which directly modulate its interaction with microtubules (Cleveland et al., 1977; Mukrasch et al., 2007; Barbier et al., 2019).

### Tau Pathology in AD and Other Tauopathies

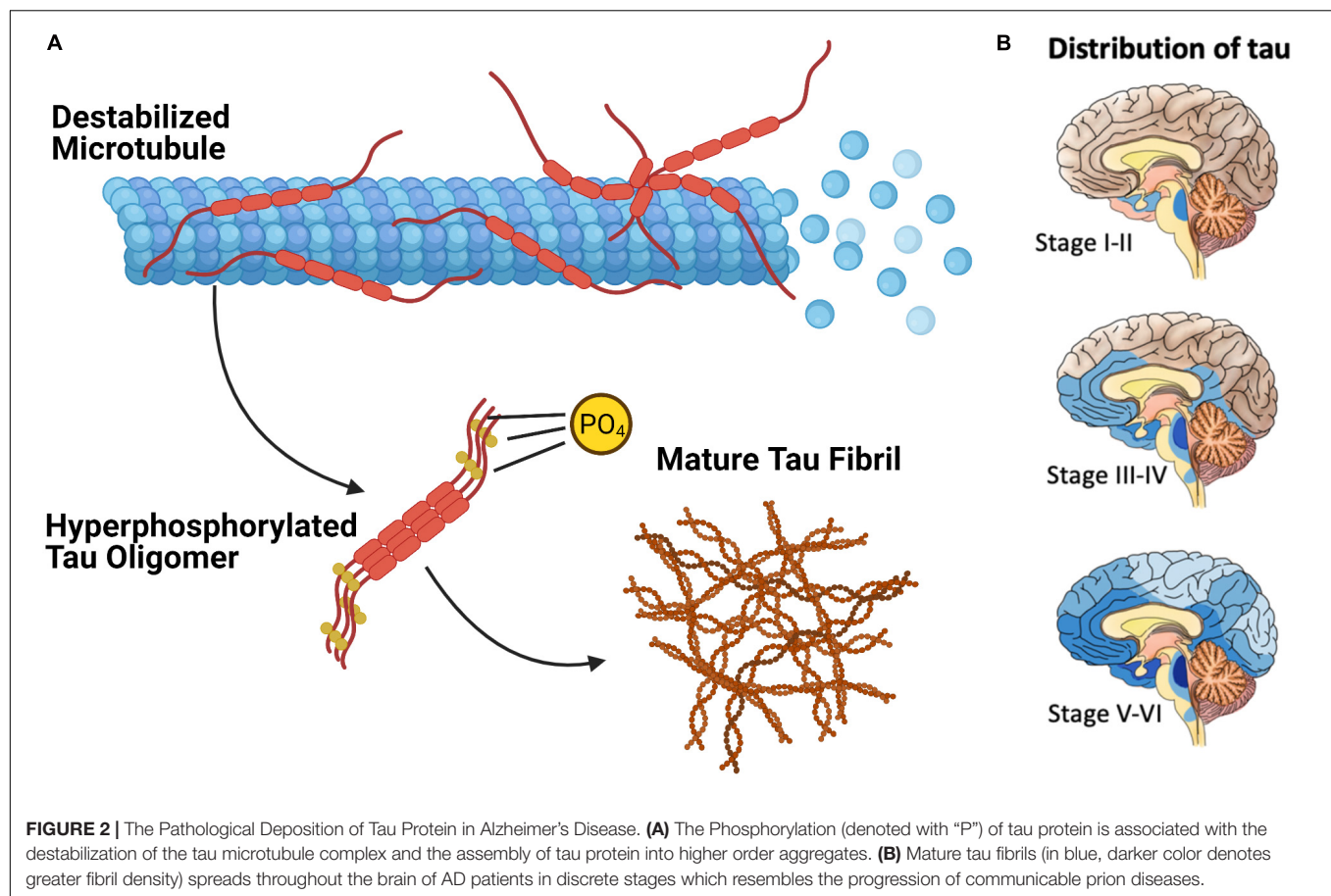
Extensive phosphorylation of tau is associated with the self-assembly of tau monomers into higher order aggregates (Figure 2A; Grundke-Iqbal et al., 1986; Goedert et al., 1996; Amniai et al., 2009; Alonso et al., 2010). Under the pathological conditions associated with tauopathies, these aggregates take the form of misfolded oligomers and filaments that spread throughout the brain in an orderly and stereotypical fashion (Mudher et al., 2017; Goedert et al., 2017b). The characteristic pattern of tau pathology in AD is described by Braak Staging (Braak and Braak, 1991), which begins with the appearance of initial tau lesions in the transentorhinal cortex during stage I. During the subsequent stages of disease progression, the density



of tau lesions increases and NFTs spread to the entorhinal cortex in stage II, then to limbic regions of the brain in stage III, before finally reaching the neocortex in stage IV and beyond (Figure 2B). The progression of tau pathology is accompanied by increases in both phosphorylated and total tau in cerebrospinal fluid (CSF), and correlates remarkably well to the severity of neurodegenerative symptoms (Arriagada et al., 1992; Buerger et al., 2006).

AD is a secondary tauopathy, which can be distinguished from primary tauopathies by the presence of additional species of proteopathic hallmarks beyond the characteristic tau-based inclusions. In the case of AD, this consists of extracellular plaques composed of Amyloid  $\beta$  ( $A\beta$ ), a  $\sim 40$  residue peptide derived from the integral membrane protein APP, while Dementia with Lewy Bodies (DLB) is characterized by extracellular inclusions composed of  $\alpha$ -synuclein in addition to tau tangles. However, this demarcation is not absolute. For example, despite being categorized as primary tauopathies, the various forms of Frontotemporal Dementia (FTD) often present with inclusions composed of TDP-43. Additionally, a number of studies have suggested a high prevalence of tau pathology in cases of so-called mixed dementia in which a patient simultaneously exhibits hallmarks of multiple neurodegenerative pathologies. Another point of distinction that can be drawn between tauopathies are the populations of cells which exhibit tau lesions. Whereas tau pathology in AD primarily affects neurons, other tauopathies are characterized by the presence of additional tau inclusions within glial cells (Table 1). Interestingly, tauopathies also differ in terms of the predominant isoform composing their tau inclusions (Table 1), with AD exhibiting a roughly 2:1 ratio of 4R to 3R tau (Goedert et al., 2012; Wagshal et al., 2015; Irwin, 2016; Custodio et al., 2017; Josephs, 2017; Ferrer, 2018).





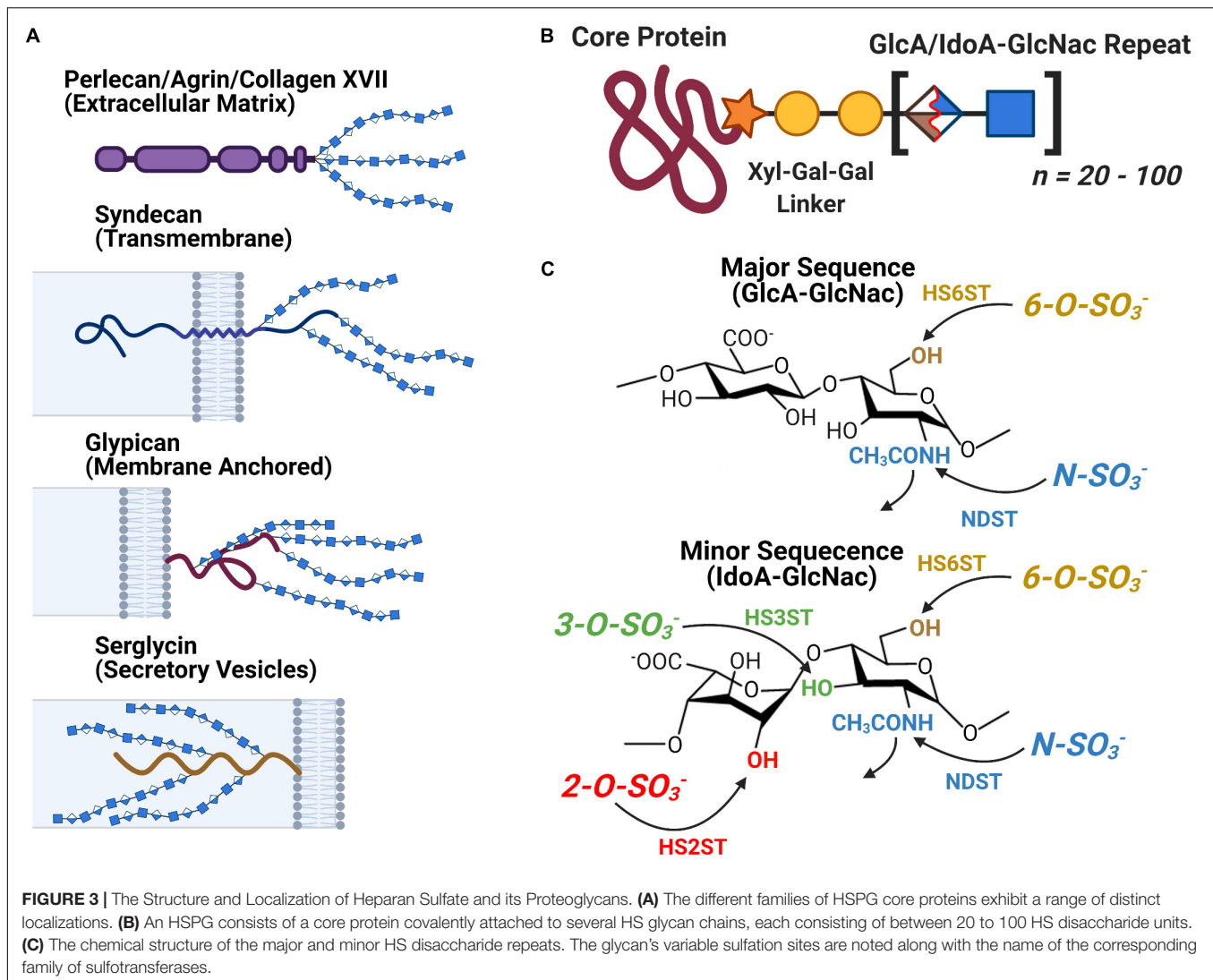
**TABLE 1 |** Features of notable tauopathies.

Clinical diagnosis	Type and localization of tau pathology	Additional proteopathic hallmark(s)	Predominant tau isoform(s)
Alzheimer's disease (AD)	NFT pathology in neurons	Plaques composed of aggregated A $\beta$ peptide	3R + 4R tau
Pick's disease	Pick's bodies within neurons, extensive glial tau inclusions	N/A	3R tau
Corticobasal Degeneration	Coiled inclusions within neurons + glial inclusions	N/A	4R tau
Argyophilic grain disease	Spindle shaped neuronal inclusions	N/A	4R tau
Globular glial Tauopathy	Globular inclusions within astrocytes + oligodendrocytes	N/A	4R tau
Chronic traumatic Encephalopathy	NFT pathology in neurons + astrocytic tau inclusions	High frequency of TDP-43 pathology	3R + 4R tau
Primary age related Tauopathy	NFT pathology in neurons	Absence of A $\beta$ plaques	3R + 4R tau
Dementia with Lewy Bodies (DLB)	NFT pathology in neurons	Lewy bodies of aggregated $\alpha$ -synuclein + tau	3R + 4R tau
Progressive Supranuclear Palsy (PSP)	NFT Pathology + coiled oligodendrocyte inclusions	N/A	4R tau
Frontotemporal Dementia (FTD)	NFT pathology + mixed glial inclusions	~50% exhibit TDP-43 pathology	Varies

## Heparan Sulfate and HSPGs in the Prion Like Spread of Tau Pathology

A substantial body of evidence has established that the spread of tau pathology in the brain occurs through a prion-like mechanism in which seeds of pathological tau are transmitted between cells and nucleate the misfolding of physiological tau in

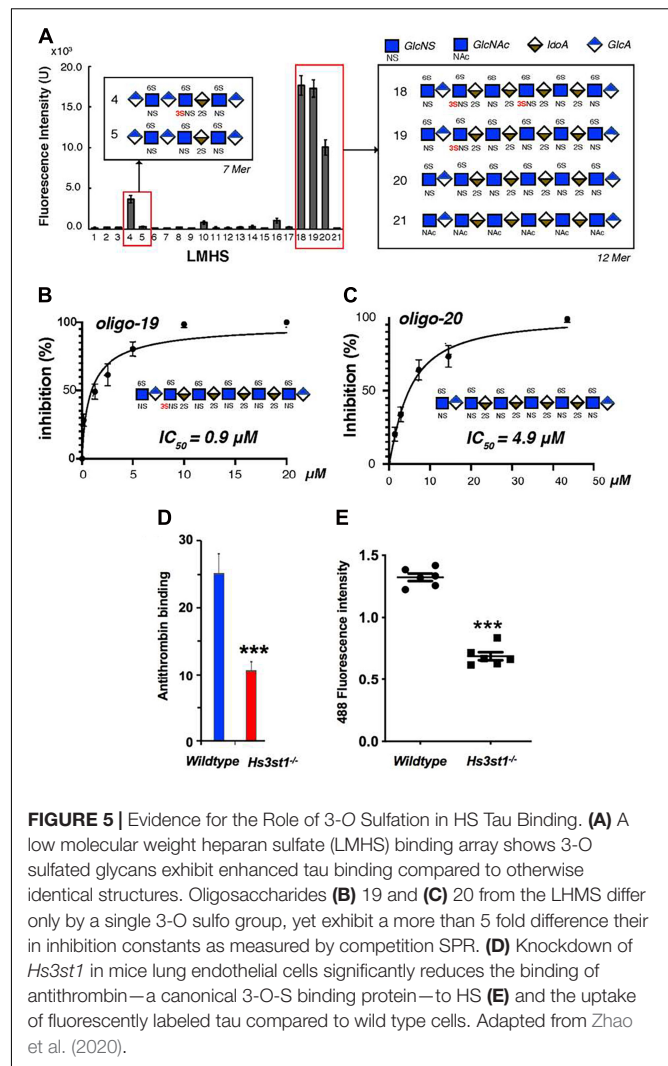
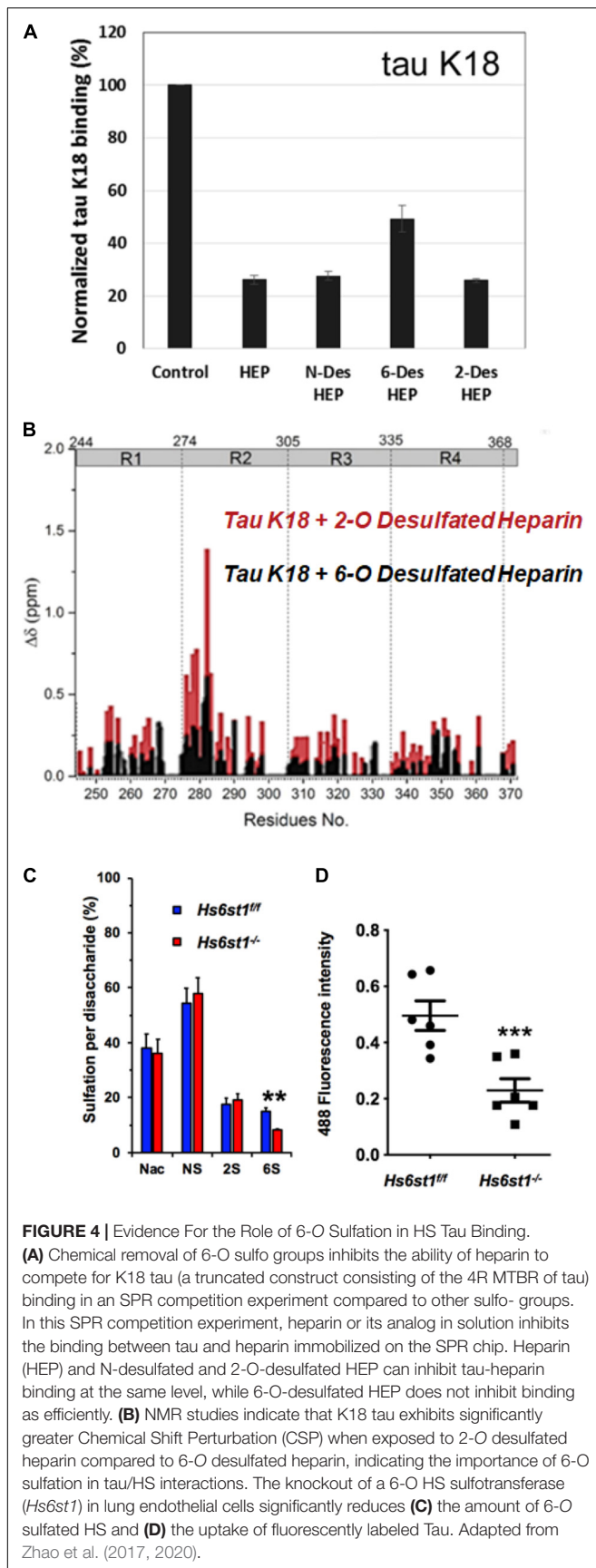
a process known as template misfolding (Brettschneider et al., 2015). The mechanisms of transcellular tau propagation are understood best in the context of AD, where tau pathology is transmitted through the synaptic junction via the secretion of tau seeds by a presynaptic neuron and their subsequent reuptake by post-synaptic neurons (Goedert et al., 2017b;



Mudher et al., 2017; Wang et al., 2017). This pathway has been shown to be mediated by Heparan Sulfate (HS) (Holmes et al., 2013; Christianson and Belting, 2014), a ubiquitous polysaccharide found across virtually all metazoans, from rudimentary invertebrates to humans.

Heparan sulfate is a linear Glycosaminoglycan (GAG) often encountered in the form of Heparan Sulfate Proteoglycans (HSPGs) a protein-glycan conjugate which consist of one of several families of HSPG core protein covalently attached to a series of HS chains ranging in length from approximately 20 to 120 disaccharide subunits (Figure 3). The localization of an HSPG is dictated by the identity of its core protein: perlecan, agrin, and type XVIII collagen based HSPGs are found in the extracellular matrix, syndecan and glypican based HSPGs localize to the exterior of the plasma membrane, and serglycin based HSPGs localize to the interior membrane of secretory vesicles (Figure 3A). The HS polymer is built from repeating disaccharides of either glucuronic (GlcA) or iduronic acid (IdoA) followed

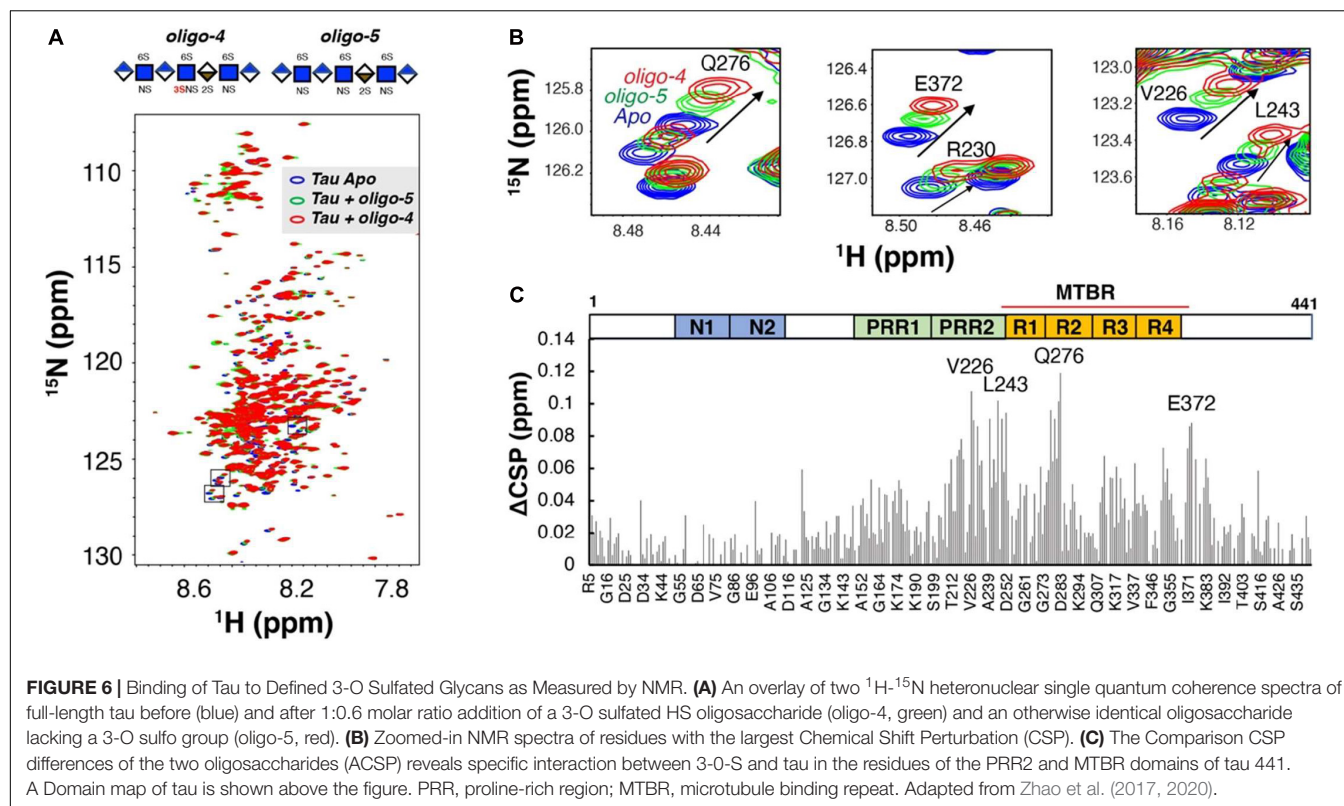
by *N*-acetylglucosamine (GlcNac) with the former (GlcA-GlcNac) being the more prevalent of the two. HS biosynthesis is carried out at the Golgi, during which the growing HS chain can be acted upon by several groups of enzymes to yield a variety of final structures: C5 Epimerase converts GlcA-GlcNac disaccharides to IdoA-GlcNac, while HS sulfotransferases *N*-deacetylase/*N*-sulfotransferase (NDST) and 2-*O*-, 3-*O*-, and 6-*O*-sulfotransferases (HS2ST, HS3ST, and HS6ST) carry out the sulfation of specific sites on the HS disaccharide (Figure 3C). Following synthesis, HS can be further catabolically modified by the endosulfatases Sulf1 and Sulf2, which selectively remove 6-*O*-sulfo groups from cell surface HSPGs. Heparin is a structural isoform of HS secreted by mast cells, notable for its medical use as an anti-coagulant. It is frequently used as an analog for HS to study the molecular interactions of HS *in vitro* due to its widespread availability (Capila and Linhardt, 2002; Esko and Selleck, 2002; Bishop et al., 2007; Xu and Esko, 2014; Li and Kusche-Gullberg, 2016).



Heparan sulfate proteoglycans are known to participate in endocytosis via multiple canonical pathways. In pinocytosis, extracellular molecules bind to HSPGs and are internalized directly via a clathrin and caveolin independent pathway. Depending on the size of the ingestion, this process can be subdivided into micropinocytosis, which involves the internalization of small particles, and macropinocytosis, which facilitates the uptake of larger macromolecular complexes. In receptor mediated endocytosis pathways, HSPGs often act as a cofactor which facilitate the initial rapid capture of a ligand and coordinate its subsequent binding to its receptor (Christianson and Belting, 2014). Current evidence also indicates the existence of novel, non-vesicular secretory pathways which are mediated by HSPGs; however, this process remains poorly understood at this time (Prudovsky et al., 2008; Merezko et al., 2018; Park et al., 2020).

## Tau Determinants of Tau-HS Interaction

HS and heparin are both capable of directly binding to tau monomers, oligomers, and fibrils *in vitro*. This interaction



also occurs *in vivo*, as the GAG sidechains of HSPGs are observed to colocalize with tau-based lesions in the brains of patients suffering from both AD and other tauopathies (Snow et al., 1990; Su et al., 1992). HS/tau binding is driven predominately via electrostatic interactions that occur between positively charged residues in tau and the negatively charged sulfo groups present on the HS GAG. Further work has identified the  $^{275}\text{VQIINK}^{280}$  hexapeptide present at the start of the R2 repeat of the MTBR domain as a major site of contact between tau and heparin. The interactions occurring at this site and the corresponding  $^{306}\text{VQIVYK}^{311}$  hexapeptide at the start of the R3 repeat have been shown to promote the formation of a local extended  $\beta$ -conformation that serves as a nucleation site for tau protein aggregation (Smet et al., 2004; Mukrasch et al., 2005; Sibille et al., 2006).

## Heparan Sulfate in Tau Protein Secretion

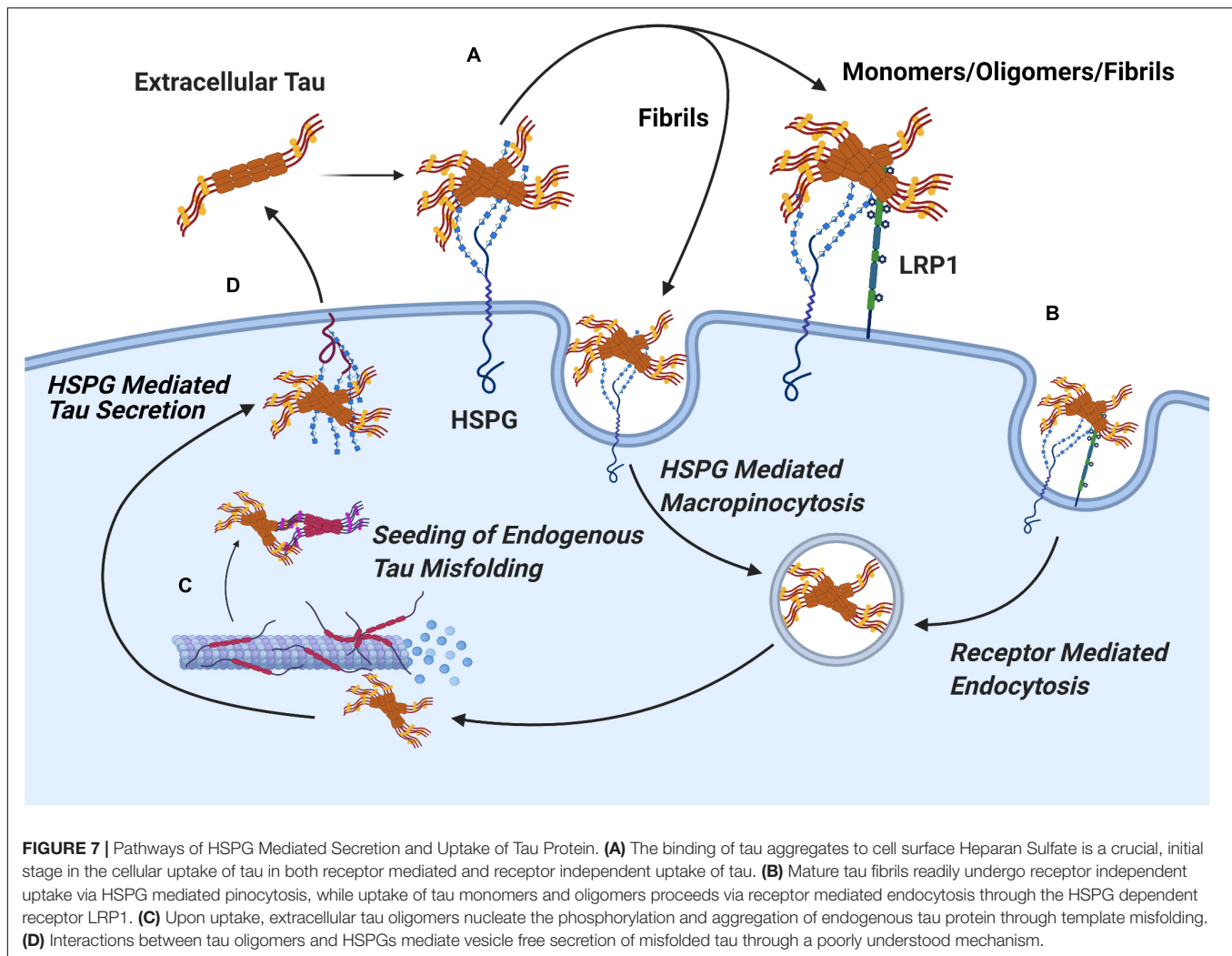
An emerging line of evidence has implicated Tau/HS interactions as a driver of both the secretion of tau into the extracellular space and its subsequent internalization by other cells of the CNS. Although tau protein is predominately an intracellular protein, there exists a small pool of extracellular tau even under ordinary physiological conditions (Goedert et al., 2017a; Yamada, 2017; Pérez et al., 2019). A small portion of this tau protein is present in extracellular vesicles, however, the vast majority of tau has been shown to be membrane free (Wegmann et al., 2016; Yan et al., 2016). Recently Merezkhko et al. reported that the secretion of membrane free tau occurs in an ATP independent fashion, suggesting the use

of a novel vesicle independent secretory pathway. The group also found that this process was abolished by treatment with heparinase—an enzyme which degrades HS—or inhibition of the HS biosynthetic pathway, indicating the participation of HS in facilitating tau protein's entry into the extracellular space (Katsinelos et al., 2018).

## DISTINCT HS SULFATION PATTERNS GOVERN TAU-HS INTERACTION AND UPTAKE

It was initially assumed that the interactions between tau and GAGs—and between GAGs and proteins in general—were non-specific and driven exclusively by the GAG's overall degree of sulfation, making the glycan's individual sulfation sites essentially interchangeable for one another (Capila and Linhardt, 2002). However, contrary to such expectations, our group demonstrated in 2017 substantial differences in the strength of tau/heparin interaction following the chemical removal of different sulfation moieties, indicating that 6-O desulfation significantly reduced tau-heparin binding, while the impact of 2-O desulfation was limited (Figure 4A; Zhao et al., 2017). In the same study, using a truncated tau construct consisting of the 4R MTBR domain known as tau K18, we characterized the binding of tau to heparin, 2-O desulfated heparin, and 6-O desulfated heparin using solution NMR. Our results were indicative of significant 6-O-S mediated contacts in the R2 subdomain of tau

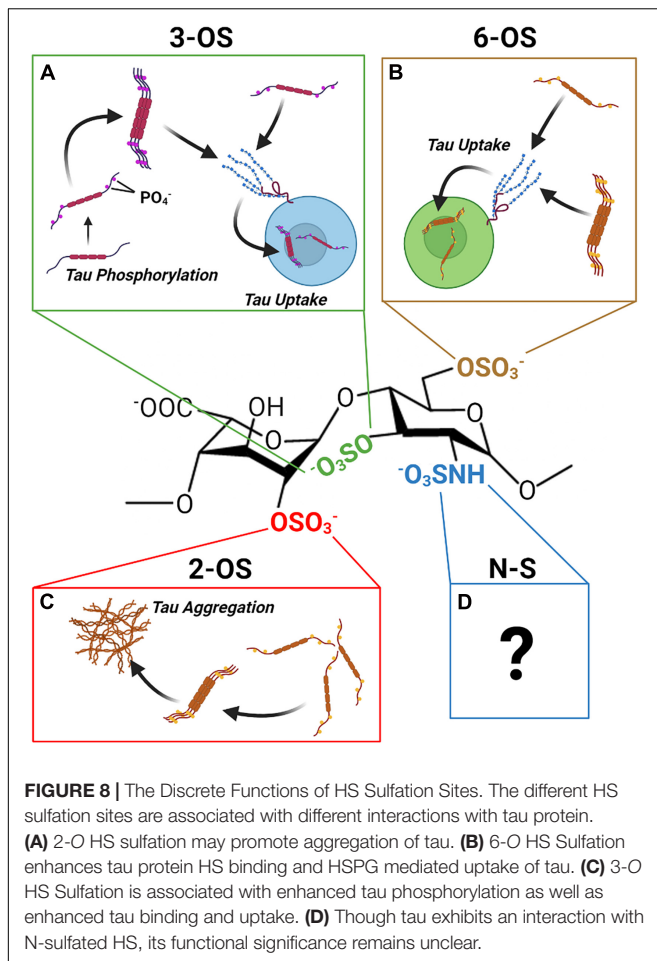




(Figures 4B–D; Zhao et al., 2017). Stopschinski et al. (2018) characterized the HS binding of various proteins involved in neurodegeneration and showed that the preference for specific sulfation moieties is particular to tau protein, with both A $\beta$  and  $\alpha$ -synuclein exhibiting a significantly higher degree of promiscuity for sulfation moieties than tau protein. Rauch et al. (2018) also reported a crucial role for 6-O sulfation of HSPGs and HS 6-O sulfotransferases in this pathway. Work by Sepulveda-Diaz et al. (2015) indicated that HS3ST2, one of the 3-O-S HS Sulfotransferases, acts as a crucial mediator of tau phosphorylation, suggesting a potential link between tau pathology and 3-O sulfated HS. Our group later identified that tau protein exhibits a direct and specific interaction with the 3-O-S moiety, and further implicated 3-O-S as a driver of cellular uptake of tau protein (Figure 5; Zhao et al., 2020), making tau one of only a handful of proteins known to interact with the rare 3-O-S moiety (Thacker et al., 2014). In the same paper, we employed solution NMR to compare the chemical shift perturbation induced by tau's interaction with two chemically defined HS heptamers which differed from one another only

by a 3-O-S moiety. Our results suggest that the major sites of contact between tau and the 3-O-S moiety are localized to the PRR2 domain and the R2 repeat of the MTBR (Zhao et al., 2020; Figure 6).

The current evidence concerning the mechanism of HSPGs in tau protein uptake suggests a role for two distinct pathways. A significant portion of large tau aggregates such as mature fibrils undergo direct internalization by HSPGs via macropinocytosis, while soluble oligomers and tau monomers are dependent on receptor mediated endocytosis to cross the cell membrane (Holmes et al., 2013). Consistent with this, in a recent report by Hudák et al. (2019) syndecan associated HSPGs were linked to the uptake of tau fibrils via a clathrin independent endocytosis mechanism mediated by lipid rafts, as would be expected for macropinocytosis of tau fibrils. Work in CNS cell lines has identified the probable receptor involved in endocytosis of soluble tau monomers and oligomer species as LRP1, a lipoprotein receptor which acts cooperatively with HSPGs (Figure 7; Rauch et al., 2020).



The identification of LRP1 as a receptor for tau protein uptake raises several exciting possibilities. LRP1 is a canonical receptor for lipoproteins, notably Apolipoprotein E (ApoE), the protein product of the *APOE* gene, a major risk factor for AD (Holtzman et al., 2012). ApoE has previously been shown to directly interact with tau protein *in vitro* (Strittmatter et al., 1994; Bachmeier et al., 2014), which suggests the possibility of a ternary interaction between tau, ApoE, and HSPGs which could greatly impact the course of tau pathology.

There exists tentative evidence for alternative pathways of tau protein uptake which do not utilize HS. Perea et al. (2019) reported that the uptake of monomeric tau by primary astrocytes was unaffected by pre-treatment of the cultures with heparin or heparinase, indicating the astrocytes took up tau via an as of yet unidentified HSPG independent pathway. In a comparative study of the uptake of brain derived tau oligomers from patients with AD, PSP, and DLB, Puangmalai et al. (2020) found that the knockout of the HSPG biosynthetic enzyme exostin-2, as well as treatment with HSPG antagonists abolished the uptake of AD and DLB derived tau oligomers as expected. However, uptake of PSP derived oligomers was merely slowed, suggesting the presence

of an HSPG independent mechanisms of uptake specific to PSP derived tau oligomers.

## Role of Heparin and HS in Tau Aggregation

Heparin and other polyanions are capable of inducing the assembly of unphosphorylated tau protein into fibrils (Kampers et al., 1996; Wilson and Binder, 1997), and thus heparin has seen wide-spread use in *in vitro* studies of tau fibrilization. However, a significant body of evidence indicates this does not truly reflect *in vivo* aggregation. Cryo-EM studies have consistently shown that the morphology of heparin induced tau fibrils differ from those found in the brains of patients suffering from tauopathies (Fitzpatrick et al., 2017; Falcon et al., 2018; Zhang et al., 2019). This is further supported by the recent work of Despres et al. (2019) who found that heparin nucleated tau fibrils exhibited a different conformation and activity from fibrils which were seeded by *in vitro* phosphorylated tau and brain derived tau. Using solid state NMR Savastano et al. (2020) characterized the structure of truncated tau and polypeptide constructs from the PRR2 subdomain of the PRR, and found evidence for the incorporation of PRR2 into the rigid core of tau PHFs following heparin induced fibrilization, consistent with our own work indicating extensive contacts between tau and heparin in the PRR domain.

Work by Fichou et al. (2018) found that fibrils derived from both recombinant tau protein and mouse brains could be induced to depolymerize back into oligomers and monomers through the removal of polyanions. Using nanopore based sensors, Giamblanco et al. (2020) monitored tau protein during heparin induced fibrilization, comparing the FTD associated P301L tau isoform to wild type tau, and found that the P301L mutation promoted the assembly of tau monomers into oligomers, and the dissociation of tau fibrils into oligomers. In light of the evidence that tau oligomers exhibit higher cytotoxicity compared to fibrils or monomers (Tian et al., 2013; You et al., 2019), this suggests altered interactions between P301L tau and HS may contribute to familial FTD by destabilizing fibrils which would otherwise sequester tau protein and mitigate tau protein proteotoxicity. A study by Townsend et al. (2020) of the aggregation kinetics of heparin nucleated tau revealed that differential desulfation of heparin dramatically altered the kinetics of heparin induced aggregation of a recombinant tau fragment. 2-O desulfation was found to substantially increase the time required for aggregation and increase the flexibility of the resulting fibrils compared to those induced to aggregate with 6-O desulfated or N-acetylated heparin. This indicates that despite its relative weak contribution to HS-tau binding, 2-O sulfation of HS plays an important role in tau aggregation.

## Recent Evidence for the Importance of Tau-HS Interactions in Tauopathies

The role of HSPGs in the prion-like spread of tau pathology has gained additional support in recent years from the analysis of tissues from the brains of patients with tauopathies and other

forms of *in vivo* evidence. Two meta analyses of genome wide association studies for AD risk factors have implicated enhanced expression of the 3-O HS sulfotransferase gene *Hs3st1* as an AD risk factor, supporting existing observations on the role of 3-O sulfated HS in tau protein uptake and phosphorylation (Witoelar et al., 2018; Schwartzentruber et al., 2021). Consistent with *in vitro* work suggesting enhanced HS-tau binding promotes enhanced spread of tau pathology, Huynh et al. (2019) have reported an increase in HS expression in the brains of patients with AD both in absolute terms and relative to chondroitin sulfate (CS), another member of the GAG family. The group also reported that AD derived HS exhibited a higher binding capacity for tau protein compared to HS derived from healthy brains, indicating AD associated changes in HS enhance the strength of its interactions with tau protein. Lorente-Gea et al. (2020) conducted a systematic analysis of major HSPG core protein expression across the regions of the brains of patients with AD. While changes in the expression of extracellular HSPG core proteins were limited in AD brains, the study revealed upregulation of cell surface syndecan and intracellular serglycin. In particular, there is extensive overexpression of syndecan 4 and serglycin, which are associated with both amyloid and tau pathology in most AD brain samples (Lorente-Gea et al., 2020), indicating a potentially undiscovered role for intracellular HSPGs in tauopathies. Some strides have been taken in recent years toward clinical translation of this line of research into potential drugs for AD and related dementias. Weisová et al. (2019) studied the mechanism of AX004, a therapeutic antibodies known to inhibit tau uptake *in vivo*, and determined the antibody's mechanism was driven by disruption of the tau-HS interaction via binding to the MTBR. Stopschinski et al. (2020) recently reported a synthetic heparin like oligosaccharide capable of disrupting cellular propagation of tau protein at similar activity to full length porcine derived heparin. In the long term, drugs targeting the tau HS interaction could prove to be a novel therapeutic for AD.

## SUMMARY AND OUTLOOK

Recent work by our group and others has revealed, contrary to prior expectations, that the sulfation sites present on the HS disaccharide exhibit specific interactions with tau protein that are not functionally interchangeable. Distinct roles in mediating HS tau interactions relevant to the aggregation, uptake, and phosphorylation of tau have been established for 2-O, 6-O, and 3-O sulfation, respectively (Figure 8). Multiple studies have established that direct interactions with HSPGs drive tau protein uptake in tauopathies, and emerging evidence is also suggestive of a role for HSPGs in tau protein secretion. Research into the

specific HSPG core proteins involved in HSPG tau interactions is limited; but work thus far points to a prominent role for cell surface and potentially intracellular HSPGs rather than ECM localized HSPGs.

A number of questions in this area of research remain outstanding. A definitive role of *N*-sulfation has not been established, despite some evidence for specific interactions between tau and *N*-sulfated HS glycans. In addition, there is no study on the glycan determinants of HSPG mediated tau secretion, and the mechanism by which HSPGs drive non-vesicular tau secretion is still poorly understood.

The extent to which HSPGs play a role in tau protein aggregation *in vivo* remains unclear, despite ample *in vitro* evidence of heparin induced tau aggregation. Nevertheless, the presence of HS within brain derived tau fibrils indicates that interaction between HS and tau must occur at some point during fibril assembly *in vivo*. One possibility is that cell surface HSPGs act as nucleation sites for tau aggregation under pathological conditions and are then internalized. Though evidence exists for HSPG independent tau uptake pathways under certain conditions, the receptors involved have not yet been identified. Further study of non-AD tauopathies such as PSP and by cells where HSPG independent tau uptake is observed may help elucidate the components of these pathways. Finally, looking forward toward translation, the prominent role of HSPGs in tau pathology, and the capacity of glycan-based HS analogs to inhibit tau propagation in cells suggests a potential application of HS based drugs in the treatment of AD and other tauopathies (Alavi Naini and Soussi-Yanicostas, 2018). Glycan-based drug discovery efforts will clearly benefit from a more detailed understanding of the mechanistic roles of HS-sulfation patterns in the pathogenetic mechanisms of tauopathies.

## AUTHOR CONTRIBUTIONS

DM wrote the initial manuscript. CW and DM conceived of the review. XL and JZ contributed to the content of the figure. XL, JZ, RL, FZ, JL, and LW provided feedback and edited the manuscript. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

This research was supported by NIH grant 1RF1AG069039 to CW and R56 AG062344 to LW. DM was supported by NIGMS training grant T32GM067545. Figures 1–3, 7, 8 were created using BioRender (biorender.com).

## REFERENCES

- Alavi Naini, S. M., and Soussi-Yanicostas, N. (2018). Heparan Sulfate as a Therapeutic Target in Tauopathies: Insights From Zebrafish. *Frontiers in Cell and Developmental Biology* 6:163. doi: 10.3389/fcell.2018.0163
- Alonso, A. D., Clerico, J. Di, Li, B., Corbo, C. P., Alaniz, M. E., Grundke-Iqbal, I., et al. (2010). Phosphorylation of Tau at Thr<sup>212</sup>, Thr<sup>231</sup>, and Ser<sup>262</sup> Combined Causes Neurodegeneration \*. *Journal of Biological Chemistry* 285, 30851–30860. doi: 10.1074/jbc.M110.110957
- Amniai, L., Barbier, P., Sillen, A., Wieruszeski, J.-M., Peyrot, V., Lippens, G., et al. (2009). Alzheimer Disease Specific Phosphopeptides of Tau Interfere with



- Assembly of Tubulin but Not Binding to Microtubules. *The FASEB Journal* 23, 1146–1152. doi: 10.1096/fj.08-121590
- Arriagada, P. V., Growdon, J. H., Hedley-Whyte, E. T., and Hyman, B. T. (1992). Neurofibrillary Tangles but Not Senile Plaques Parallel Duration and Severity of Alzheimer's Disease. *Neurology* 42, 631631. doi: 10.1212/WNL.42.3.631
- Bachmeier, C., Shackleton, B., Ojo, J., Paris, D., Mullan, M., and Crawford, F. (2014). Apolipoprotein E Isoform-Specific Effects on Lipoprotein Receptor Processing. *Neuromolecular Medicine* 16, 686–696. doi: 10.1007/s12017-014-8318-6
- Barbier, P., Zejneli, O., Martinho, M., Lasorsa, A., Belle, V., Smet-Nocca, C., et al. (2019). Role of Tau as a Microtubule-Associated Protein: Structural and Functional Aspects. *Frontiers in Aging Neuroscience* 11:204. doi: 10.3389/fnagi.2019.00204
- Binder, L. I., Guillozet-Bongaarts, A. L., Garcia-Sierra, F., and Berry, R. W. (2005). Tau, Tangles, and Alzheimer's Disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1739, 216–223. doi: 10.1016/j.bbdis.2004.08.014
- Bishop, J. R., Schuksz, M., and Esko, J. D. (2007). Heparan Sulphate Proteoglycans Fine-Tune Mammalian Physiology. *Nature* 446, 1030–1037. doi: 10.1038/nature05817
- Braak, H., and Braak, E. (1991). Neuropathological Stageing of Alzheimer-Related Changes. *Acta Neuropathologica* 82, 239–259. doi: 10.1007/BF00308809
- Brettschneider, J., Tredici, K. D., Lee, V. M. Y., and Trojanowski, J. Q. (2015). Spreading of Pathology in Neurodegenerative Diseases: A Focus on Human Studies. *Nature Reviews Neuroscience* 16, 109–120. doi: 10.1038/nrn3887
- Buée, L., Bussière, T., Buée-Scherrer, V., Delacourte, A., and Hof, P. R. (2000). Tau Protein Isoforms, Phosphorylation and Role in Neurodegenerative Disorders. These Authors Contributed Equally to This Work. *Brain Research Reviews* 33, 95–130. doi: 10.1016/S0165-0173(00)00019-9
- Buerger, K., Ewers, M., Pirttilä, T., Zinkowski, R., Alafuzoff, I., Teipel, S. J., et al. (2006). CSF Phosphorylated Tau Protein Correlates with Neocortical Neurofibrillary Pathology in Alzheimer's Disease. *Brain* 129, 3035–3041. doi: 10.1093/brain/awl269
- Capila, I., and Linhardt, R. J. (2002). Heparin-Protein Interactions. *Angewandte Chemie International Edition* 41, 390–412. doi: 10.1002/1521-3773(20020201)41:3<390::AID-ANIE390>3.0.CO;2-B
- Christianson, H. C., and Belting, M. (2014). Heparan Sulfate Proteoglycan as a Cell-Surface Endocytosis Receptor. *Matrix Biology* 35, 51–55. doi: 10.1016/j.matbio.2013.10.004
- Cleveland, D. W., Hwo, S.-Y., and Kirschner, M. W. (1977). Physical and Chemical Properties of Purified Tau Factor and the Role of Tau in Microtubule Assembly. *Journal of Molecular Biology* 116, 227–247. doi: 10.1016/0022-2836(77)90214-5
- Custodio, N., Montesinos, R., Lira, D., Herrera-Pérez, E., Bardales, Y., and Valeriano-Lorenzo, L. (2017). Mixed Dementia: A Review of the Evidence. *Dementia & Neuropsychologia* 11, 364–370. doi: 10.1590/1980-57642016dn11-040005
- Despres, C., Di, J., Cantrelle, F., Li, Z., Huvent, I., Chambraud, B., et al. (2019). Major Differences between the Self-Assembly and Seeding Behavior of Heparin-Induced and in Vitro Phosphorylated Tau and Their Modulation by Potential Inhibitors. *ACS Chemical Biology* 14, 1363–1379. doi: 10.1021/acscchembio.9b00325
- Esko, J. D., and Selleck, S. B. (2002). Order Out of Chaos: Assembly of Ligand Binding Sites in Heparan Sulfate. *Annual Review of Biochemistry* 71, 435–471. doi: 10.1146/annurev.biochem.71.110601.135458
- Falcon, B., Zhang, W., Murzin, A. G., Murshudov, G., Garringer, H. J., Vidal, R., et al. (2018). Structures of Filaments from Pick's Disease Reveal a Novel Tau Protein Fold. *Nature* 561, 137–140. doi: 10.1038/s41586-018-0454-y
- Ferrer, I. (2018). Astroglial Pathology in Tauopathies. *Neuroglia* 1, 126–150. doi: 10.3390/neuroglia1010010
- Fichou, Y., Yanxian, L., Rauch, J. N., Vigers, M., Zeng, Z., Srivastava, M., et al. (2018). Cofactors Are Essential Constituents of Stable and Seeding-Active Tau Fibrils. *Proceedings of the National Academy of Sciences* 115, 13234–13239. doi: 10.1073/pnas.1810058115
- Fitzpatrick, A. W. P., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., et al. (2017). Cryo-EM Structures of Tau Filaments from Alzheimer's Disease. *Nature* 547, 185–190. doi: 10.1038/nature23002
- Giamblanco, N., Fichou, Y., Janot, J.-M., Balanzat, E., Han, S., and Balme, S. (2020). Mechanisms of Heparin-Induced Tau Aggregation Revealed by a Single Nanopore. *ACS Sensors* 5, 1158–1167. doi: 10.1021/acssensors.0c01993
- Goedert, M., Eisenberg, D. S., and Crowther, R. A. (2017a). Propagation of Tau Aggregates and Neurodegeneration. *Annual Review of Neuroscience* 40, 189–210. doi: 10.1146/annurev-neuro-072116-031153
- Goedert, M., Ghetti, B., and Spillantini, M. G. (2012). Frontotemporal Dementia: Implications for Understanding Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine* 2, a006254–a006254. doi: 10.1101/cshperspect.a006254
- Goedert, M., Jakes, R., Spillantini, M. G., Hasegawa, M., Smith, M. J., and Crowther, R. A. (1996). Assembly of Microtubule-Associated Protein Tau into Alzheimer-like Filaments Induced by Sulphated Glycosaminoglycans. *Nature* 383, 550–553. doi: 10.1038/383550a0
- Goedert, M., Masuda-Suzukake, M., and Falcon, B. (2017b). Like Prions: The Propagation of Aggregated Tau and  $\alpha$ -Synuclein in Neurodegeneration. *Brain* 140, 266–278. doi: 10.1093/brain/aww230
- Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J., and Crowther, R. A. (1989). Cloning and Sequencing of the cDNA Encoding an Isoform of Microtubule-Associated Protein Tau Containing Four Tandem Repeats: Differential Expression of Tau Protein MRNAs in Human Brain. *The EMBO Journal* 8, 393–399.
- Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., and Binder, L. I. (1986). Abnormal Phosphorylation of the Microtubule-Associated Protein Tau (Tau) in Alzheimer Cytoskeletal Pathology. *Proceedings of the National Academy of Sciences of the United States of America* 83, 4913–4917. doi: 10.1073/pnas.83.13.4913
- Holmes, B. B., DeVos, S. L., Kfoury, N., Li, M., Jacks, R., Yanamandra, K., et al. (2013). Heparan Sulfate Proteoglycans Mediate Internalization and Propagation of Specific Proteopathic Seeds. *Proceedings of the National Academy of Sciences* 110, E3138–E3147. doi: 10.1073/pnas.1301440110
- Holtzman, D. M., Herz, J., and Bu, G. (2012). Apolipoprotein E and Apolipoprotein E Receptors: Normal Biology and Roles in Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine* 2, a006312–a006312. doi: 10.1101/cshperspect.a006312
- Hudák, A., Kusz, E., Domonkos, I., Jósavay, K., Kodamullil, A. T., Szilák, L., et al. (2019). Contribution of Syndecans to Cellular Uptake and Fibrillation of  $\alpha$ -Synuclein and Tau. *Scientific Reports* 9, 16543. doi: 10.1038/s41598-019-53038-z
- Huynh, M. B., Ouidja, M. O., Chantepie, S., Carpentier, G., Ma, A., Zhang, G., et al. (2019). Glycosaminoglycans from Alzheimer's Disease Hippocampus Have Altered Capacities to Bind and Regulate Growth Factors Activities and to Bind Tau. *PLoS One* 14:e0209573. doi: 10.1371/journal.pone.0209573
- Irwin, D. J. (2016). Tauopathies as Clinicopathological Entities. *Parkinsonism & Related Disorders* 22(Suppl. 1), S29–S33. doi: 10.1016/j.parkreldis.2015.09.020
- Josephs, K. A. (2017). Current Understanding of Neurodegenerative Diseases Associated With the Protein Tau. *Mayo Clinic Proceedings* 92, 1291–1303. doi: 10.1016/j.mayocp.2017.04.016
- Kampers, T., Friedhoff, P., Biernat, J., Mandelkow, E.-M., and Mandelkow, E. (1996). RNA Stimulates Aggregation of Microtubule-Associated Protein Tau into Alzheimer-like Paired Helical Filaments. *FEBS Letters* 399, 344–349. doi: 10.1016/S0014-5793(96)01386-5
- Katsinelos, T., Zeitler, M., Karakatsani, A., and Dimou, E. (2018). Unconventional Secretion Mediates the Trans-Cellular Spreading of Tau. *Cell Rep.* 23, 2039–2055. doi: 10.1016/j.celrep.2018.04.056
- Li, J. P., and Kusche-Gullberg, M. (2016). Heparan Sulfate: Biosynthesis, Structure, and Function. *International Review of Cell and Molecular Biology* 325, 215–273. doi: 10.1016/bs.ircmb.2016.02.009
- Lorente-Gea, L., García, B., Martín, C., Ordiales, H., García-Suárez, O., Piñabasta, K. M., et al. (2020). Heparan Sulfate Proteoglycans Undergo Differential Expression Alterations in Alzheimer Disease Brains. *Journal of*



- Neuropathology & Experimental Neurology* 79, 474–483. doi: 10.1093/jnen/nlaa016
- Merezhko, M., Brunello, C. A., Yan, X., Vihinen, H., Jokitalo, E., Uronen, R.-L., et al. (2018). Secretion of Tau via an Unconventional Non-Vesicular Mechanism. *Cell Reports* 25, 2027.e–2035.e. doi: 10.1016/j.celrep.2018.10.078
- Mudher, A., Colin, M., Dujardin, S., Medina, M., Dewachter, I., Naini, S. M. A., et al. (2017). What Is the Evidence That Tau Pathology Spreads through Prion-like Propagation? *Acta Neuropathologica Communications* 5, 99. doi: 10.1186/s40478-017-0488-7
- Mukrasch, M. D., Biernat, J., Von Bergen, M., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2005). Sites of Tau Important for Aggregation Populate  $\beta$ -Structure and Bind to Microtubules and Polyanions. *Journal of Biological Chemistry* 280, 24978–24986. doi: 10.1074/jbc.M501565200
- Mukrasch, M. D., von Bergen, M., Biernat, J., Fischer, D., Griesinger, C., Mandelkow, E., et al. (2007). The “Jaws” of the Tau-Microtubule Interaction. *Journal of Biological Chemistry* 282, 12230–12239. doi: 10.1074/jbc.M607159200
- Park, G., Kim, B.-S., and Kim, E. (2020). A Novel Function of FAF1, Which Induces Dopaminergic Neuronal Death through Cell-to-Cell Transmission. *Cell Communication and Signaling* 18, 133. doi: 10.1186/s12964-020-00632-8
- Perea, J. R., López, E., Iez-Ballesteros, J. C. D., Ávila, J., Hernández, F., and Bolós, M. (2019). Extracellular Monomeric Tau Is Internalized by Astrocytes. *Frontiers in Neuroscience* 13:442. doi: 10.3389/fnins.2019.00442
- Pérez, M., Avila, J., and Hernández, F. (2019). Propagation of Tau via Extracellular Vesicles. *Frontiers in Neuroscience* 13:698. doi: 10.3389/fnins.2019.00698
- Prudovsky, I., Tarantini, F., Landriscina, M., Neivandt, D., Soldi, R., Kirov, A., et al. (2008). Secretion without Golgi. *Journal of Cellular Biochemistry* 103, 1327–1343. doi: 10.1002/jcb.21513
- Puangmalai, N., Bhatt, N., Montalbano, M., Sengupta, U., Gaikwad, S., Ventura, F., et al. (2020). Internalization Mechanisms of Brain-Derived Tau Oligomers from Patients with Alzheimer's Disease, Progressive Supranuclear Palsy and Dementia with Lewy Bodies. *Cell Death & Disease* 11, 314. doi: 10.1038/s41419-020-2503-3
- Rauch, J. N., Chen, J. J., Sorum, A. W., Miller, G. M., Sharf, T., See, S. K., et al. (2018). Tau Internalization Is Regulated by 6-O Sulfation on Heparan Sulfate Proteoglycans (HSPGs). *Scientific Reports* 8, 6382. doi: 10.1038/s41598-018-24904-z
- Rauch, J. N., Luna, G., Guzman, E., Audouard, M., Challis, C., Sibih, Y. E., et al. (2020). LRP1 Is a Master Regulator of Tau Uptake and Spread. *Nature* 580, 381–385. doi: 10.1038/s41586-020-2156-5
- Savastano, A., Jaipuria, G., Andreas, L., Mandelkow, E., and Zweckstetter, M. (2020). Solid-State NMR Investigation of the Involvement of the P2 Region in Tau Amyloid Fibrils. *Scientific Reports* 10, 21210. doi: 10.1038/s41598-020-78161-0
- Schwartzentruber, J., Cooper, S., Liu, J. Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., et al. (2021). Genome-Wide Meta-Analysis, Fine-Mapping and Integrative Prioritization Implicate New Alzheimer's Disease Risk Genes. *Nature Genetics* 53, 392–402.
- Sepulveda-Diaz, J. E., Naini, S. M. A., Huynh, M. B., Ouidja, M. O., Yanicostas, C., Chantepie, S., et al. (2015). HS3ST2 Expression Is Critical for the Abnormal Phosphorylation of Tau in Alzheimer's Disease-Related Tau Pathology. *Brain: A Journal of Neurology* 138(Pt 5), 1339–1354. doi: 10.1093/brain/awv056
- Sibille, N., Sillen, A., Leroy, A., Wieruszkeski, J.-M., Mulloy, B., Landrieu, I., et al. (2006). Structural Impact of Heparin Binding to Full-Length Tau As Studied by NMR Spectroscopy. *Biochemistry* 45, 12560–12572. doi: 10.1021/bi060964o
- Smet, C., Leroy, A., Sillen, A., Wieruszkeski, J.-M., Landrieu, I., and Lippens, G. (2004). Accepting Its Random Coil Nature Allows a Partial NMR Assignment of the Neuronal Tau Protein. *ChemBioChem* 5, 1639–1646. doi: 10.1002/cbic.200400145
- Snow, A. D., Mar, H., Nochlin, D., Sekiguchi, R. T., Kimata, K., Koike, Y., et al. (1990). Early Accumulation of Heparan Sulfate in Neurons and in the Beta-Amyloid Protein-Containing Lesions of Alzheimer's Disease and Down's Syndrome. *The American Journal of Pathology* 137, 1253–1270.
- Stopschinski, B. E., Holmes, B. B., Miller, G. M., Manon, V. A., Vaquer-Alicea, J., Prueitt, W. L., et al. (2018). Specific Glycosaminoglycan Chain Length and Sulfation Patterns Are Required for Cell Uptake of Tau versus  $\alpha$ -synuclein and  $\beta$ -Amyloid Aggregates. *Journal of Biological Chemistry* 293, 10826–10840. doi: 10.1074/jbc.RA117.000378
- Stopschinski, B. E., Thomas, T. L., Nadj, S., Darvish, E., Fan, L., Holmes, B. B., et al. (2020). A Synthetic Heparinoid Blocks Tau Aggregate Cell Uptake and Amplification. *Journal of Biological Chemistry* 295, 2974–2983. doi: 10.1074/jbc.RA119.010353
- Strittmatter, W. J., Saunders, A. M., Goedert, M., Weisgraber, K. H., Dong, L. M., Jakes, R., et al. (1994). Isoform-Specific Interactions of Apolipoprotein E with Microtubule-Associated Protein Tau: Implications for Alzheimer Disease. *Proceedings of the National Academy of Sciences of the United States of America* 91, 11183–11186. doi: 10.1073/pnas.91.23.11183
- Su, J. H., Cummings, B. J., and Cotman, C. W. (1992). Localization of Heparan Sulfate Glycosaminoglycan and Proteoglycan Core Protein in Aged Brain and Alzheimer's Disease. *Neuroscience* 51, 801–813. doi: 10.1016/0306-4522(92)90521-3
- Thacker, B. E., Xu, D., Lawrence, R., and Esko, J. D. (2014). Heparan Sulfate 3-O-Sulfation: A Rare Modification in Search of a Function. *Matrix Biology* 35, 60–72. doi: 10.1016/j.matbio.2013.12.001
- Tian, H., Davidowitz, E., Lopez, P., Emadi, S., Moe, J., and Sierks, M. (2013). Trimeric Tau Is Toxic to Human Neuronal Cells at Low Nanomolar Concentrations. *International Journal of Cell Biology* 2013, 260787. doi: 10.1155/2013/260787
- Townsend, D., Fullwood, N. J., Yates, E. A., and Middleton, D. A. (2020). Aggregation Kinetics and Filament Structure of a Tau Fragment Are Influenced by the Sulfation Pattern of the Cofactor Heparin. *Biochemistry* 59, 4003–4014. doi: 10.1021/acs.biochem.0c00443
- Wagshal, D., Sankaranarayanan, S., Guss, V., Hall, T., Berisha, F., Lobach, I., et al. (2015). Divergent CSF  $\tau$  Alterations in Two Common Tauopathies: Alzheimer's Disease and Progressive Supranuclear Palsy. *Journal of Neurology, Neurosurgery & Psychiatry* 86, 244250. doi: 10.1136/jnnp-2014-308004
- Wang, Y., Balaji, V., Kaniyappan, S., Krüger, L., Irsen, S., Tepper, K., et al. (2017). The Release and Trans-Synaptic Transmission of Tau via Exosomes. *Molecular Neurodegeneration* 12, 5. doi: 10.1186/s13024-016-0143-y
- Wegmann, S., Nicholls, S., Takeda, S., Fan, Z., and Hyman, B. T. (2016). Formation, Release, and Internalization of Stable Tau Oligomers in Cells. *Journal of Neurochemistry* 139, 1163–1174. doi: 10.1111/jnc.13866
- Weisová, P., Cehlár, O., Škrabana, R., Žilková, M., Filipčík, P., Kováčech, B., et al. (2019). Therapeutic Antibody Targeting Microtubule-Binding Domain Prevents Neuronal Internalization of Extracellular Tau via Masking Neuron Surface Proteoglycans. *Acta Neuropathologica Communications* 7, 129. doi: 10.1186/s40478-019-0770-y
- Wilson, D. M., and Binder, L. I. (1997). Free Fatty Acids Stimulate the Polymerization of Tau and Amyloid Beta Peptides. In Vitro Evidence for a Common Effector of Pathogenesis in Alzheimer's Disease. *The American Journal of Pathology* 150, 2181–2195.
- Witoelar, A., Rongve, A. I., Almdahl, S. I., Ulstein, D., Engvig, A., White, L. R., et al. (2018). Meta-Analysis of Alzheimer's Disease on 9,751 Samples from Norway and IGAP Study Identifies Four Risk Loci. *Scientific Reports* 8, 18088. doi: 10.1038/s41598-018-36429-6
- Xu, D., and Esko, J. D. (2014). Demystifying Heparan Sulfate-Protein Interactions. *Annual Review of Biochemistry* 83, 129–157.
- Yamada, K. (2017). Extracellular Tau and Its Potential Role in the Propagation of Tau Pathology. *Frontiers in Neuroscience* 11:667. doi: 10.3389/fnins.2017.00667
- Yan, X., Nykänen, N.-P., Brunello, C. A., Haapasalo, A., Hiltunen, M., Uronen, R.-L., et al. (2016). FRMD4A-Cytohesin Signaling Modulates the Cellular Release of Tau. *Journal of Cell Science* 129, 20032015. doi: 10.1242/jcs.180745
- You, Y., Perkins, A., Cisternas, P., Muñoz, B., Taylor, X., You, Y., et al. (2019). Tau as a Mediator of Neurotoxicity Associated to Cerebral Amyloid Angiopathy. *Acta Neuropathologica Communications* 7, 26. doi: 10.1186/s40478-019-0680-z
- Zhang, W., Falcon, B., Murzin, A. G., Fan, J., Crowther, R. A., Goedert, M., et al. (2019). Heparin-Induced Tau Filaments Are Polymorphic and Differ from Those in Alzheimer's and Pick's Diseases. Edited by

- Nikolaus Grigorieff, Cynthia Wolberger, Nikolaus Grigorieff, Henning Stahlberg, and Marcus Fändrich. *ELife* 8, e43584. doi: 10.7554/eLife.43584
- Zhao, J., Huvent, I., Lippens, G., Eliezer, D., Zhang, A., Li, Q., et al. (2017). Glycan Determinants of Heparin-Tau Interaction. *Biophysical Journal* 112, 921–932. doi: 10.1016/j.bpj.2017.01.024
- Zhao, J., Zhu, Y., Song, X., Xiao, Y., Su, G., Liu, X., et al. (2020). 3-O-Sulfation of Heparan Sulfate Enhances Tau Interaction and Cellular Uptake. *Angewandte Chemie International Edition* 59, 1818–1827. doi: 10.1002/anie.201913029

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mah, Zhao, Liu, Zhang, Liu, Wang, Linhardt and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer

Jingwen Yue<sup>1,2,3</sup>, Weihua Jin<sup>4,5,6</sup>, Hua Yang<sup>1</sup>, John Faulkner<sup>1</sup>, Xuehong Song<sup>1</sup>, Hong Qiu<sup>7</sup>, Michael Teng<sup>8</sup>, Parastoo Azadi<sup>2</sup>, Fuming Zhang<sup>4,5,6</sup>, Robert J. Linhardt<sup>4,5,6</sup> and Lianchun Wang<sup>1\*</sup>

<sup>1</sup> Department of Molecular Pharmacology and Physiology, Morsani College of Medicine, University of South Florida Health, Tampa, FL, United States, <sup>2</sup> Complex Carbohydrate Research Center, University of Georgia, Athens, GA, United States, <sup>3</sup> Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, United States, <sup>4</sup> Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, United States, <sup>5</sup> Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, United States, <sup>6</sup> Department of Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, United States, <sup>7</sup> Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China, <sup>8</sup> Division of Allergy and Immunology, Department of Internal Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, United States

## OPEN ACCESS

### Edited by:

Wei Yang,  
State College of Florida,  
Manatee-Sarasota, United States

### Reviewed by:

Ding Xu,  
University at Buffalo, United States  
Qing-Xiang Amy Sang,  
Florida State University, United States

### \*Correspondence:

Lianchun Wang  
lianchunw@usf.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 05 January 2021

**Accepted:** 06 April 2021

**Published:** 11 June 2021

### Citation:

Yue J, Jin W, Yang H, Faulkner J, Song X, Qiu H, Teng M, Azadi P, Zhang F, Linhardt RJ and Wang L (2021) Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer. *Front. Mol. Biosci.* 8:649575. doi: 10.3389/fmolb.2021.649575

The severe acute respiratory syndrome (SARS)-like coronavirus disease (COVID-19) is caused by SARS-CoV-2 and has been a serious threat to global public health with limited treatment. Cellular heparan sulfate (HS) has been found to bind SARS-CoV-2 spike protein (SV2-S) and co-operate with cell surface receptor angiotensin-converting enzyme 2 (ACE2) to mediate SARS-CoV-2 infection of host cells. In this study, we determined that host cell surface SV2-S binding depends on and correlates with host cell surface HS expression. This binding is required for SARS-Cov-2 virus to infect host cells and can be blocked by heparin lyase, HS antagonist surfen, heparin, and heparin derivatives. The binding of heparin/HS to SV2-S is mainly determined by its overall sulfation with potential, minor contribution of specific SV2-S binding motifs. The higher binding affinity of SV2-S G614 mutant to heparin and upregulated HS expression may be one of the mechanisms underlying the higher infectivity of the SARS-CoV-2 G614 variant and the high vulnerability of lung cancer patients to SARS-CoV-2 infection, respectively. The higher host cell infection by SARS-CoV-2 G614 variant pseudovirus and the increased infection caused by upregulated HS expression both can be effectively blocked by heparin lyase and heparin, and possibly surfen and heparin derivatives too. Our findings support blocking HS-SV2-S interaction may provide one addition to achieve effective prevention and/treatment of COVID-19.

**Keywords:** heparan sulfate, SARS-Cov-2, spike protein, G614 mutant, lung cancer

## INTRODUCTION

The recent emergence of the novel, pathogenic Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) worldwide poses a global health emergency (Graham and Baric, 2020). To develop specific anti-coronavirus therapeutics and prophylactics, the molecular mechanisms that underlie viral infection need firstly to be defined. The spike protein of coronaviruses facilitates viral entry into target cells including the initial binding on target cell surface and the following fusion of viral and cellular membrane. Pathogenesis studies have illuminated that the spike protein of SARS-CoV-2 (SV2-S) binds to angiotensin-converting enzyme 2 (ACE2) on host cell surface for cell entry, and then is primed by proteinases, including transmembrane serine protease 2 and cathepsin B and L, to mediate viral and host cell membrane fusion (Hoffmann et al., 2020; Walls et al., 2020a). It remains incompletely understood the molecular mechanisms by which SV2-S mediates SARS-CoV-2 infection of host cells (Hoffmann et al., 2020; Walls et al., 2020a).

Heparan sulfate (HS) is a linear polysaccharide that is ubiquitously expressed by most cell types in mammals at the cell surfaces and in the extracellular matrix. HS is used by multiple viruses to attach to the host cell surface. Mechanistically, viruses exploit the HS interaction to increase their concentration at the cell surface and augment their chances of finding a more specific entry receptor (Rusnati et al., 2009; Aquino and Park, 2016; Cagno et al., 2019). Human coronaviruses NL63 and OC43 bind to cell surface HS (de Haan et al., 2008; Milewska et al., 2014). Human coronavirus NL63 utilizes HS for attachment to target cells and employs ACE2 for following host cell entry (Milewska et al., 2014). The recent studies, including our own, reported that SV2-S binds heparin, a highly sulfated form of HS, and chemically synthesized HS (Clausen et al., 2020; Kim et al., 2020; Liu et al., 2020; Tandon et al., 2020; Tiwari et al., 2020a; Zhang et al., 2020c). These studies also showed that heparin inhibits SV2-S pseudotyped virus infection of Vero cells (kidney epithelial cell). Recent studies further reported that HS is essential for SV2-S-mediated SARS-CoV-2 cell entry; however, the correlation of HS expression with cell surface SV2-S binding and the required sulfation structure feature within cell surface HS have not been fully examined (Clausen et al., 2020). Meanwhile, the SARS-CoV-2-S G614 mutant has predominated globally (Walls et al., 2020b; Zhou et al., 2020) and several retrospective cohort studies highlighted that cancer patients are at increased risk for COVID-19 severity and fatality (Liang et al., 2020; Mehta et al., 2020; Poortmans et al., 2020; Rogado et al., 2020; Yu et al., 2020; Zhang et al., 2020a,b), while the underlying molecular mechanisms remain largely unknown.

**Abbreviations:** 2S, 2-O-sulfation; 3S, 3-O-sulfation; 6S, 6-O-sulfation; 2DS-HP, 2-O-desulfated heparin; 6DS-HP, 6-O-desulfated heparin; ACE2, angiotensin-converting enzyme 2; Ext1, exostosin-1; Ext2, exostosin-2; GEPIA2, gene expression profiling interactive analysis 2; HLMVEC, human lung microvascular endothelial cell; HS, Heparan sulfate; Hs2st, heparan sulfate 2-O-sulfotransferase; Hs3st1, heparan sulfate 3-O-sulfotransferase-1; Hs3st4, heparan sulfate 3-O-sulfotransferase-4; Hs6st1, heparan sulfate 6-O-sulfotransferase-1; Hs6st2, heparan sulfate 6-O-sulfotransferase-2; HSase, heparinases I–III; LUSC, lung squamous cell carcinoma; MLEC, mouse lung endothelial cell; mTHP-1, macrophage induced

In this study, we determined that SV2-S binding depends on and correlates with host cell surface HS expression. This binding is required for SARS-CoV-2 virus to infect host cells and can be blocked by heparin lyase, HS antagonist surfen, heparin, and heparin derivatives. The binding of heparin/HS to SV2-S is mainly determined by its overall sulfation with potential, minor contributions of specific SV2-S binding motifs. The higher binding affinity of SV2-S G614 mutant to heparin and upregulated HS expression on the cell surface may be one of the mechanisms underlying the higher infectivity of the SARS-CoV-2 G614 variant and the high vulnerability of lung cancer patients to SARS-CoV-2 infection, respectively. The higher host cell infection by SARS-CoV-2 G614 variant pseudovirus and the increased infection caused by upregulated HS expression both can be effectively blocked by heparin lyase and heparin, and possibly surfen and heparin derivatives too.

## MATERIALS AND METHODS

### Cells

The cell lines A549, SH-SY5Y, 293T, THP-1 were obtained from the American-Type Culture Collection (ATCC, Manassas, VA, United States) (Table 1). The cell lines H441 and Calu-3 were purchased from Thermo Fisher Scientific (Waltham, MA, United States). The primary human lung microvascular endothelial cells (HMVECs) were obtained from Lonza (Basel, Switzerland). The mouse lung endothelial cell (MLEC) lines were generated in our lab (Wang et al., 2005; Wijelath et al., 2010; Qiu et al., 2013, 2018; Zhang B. et al., 2014). The cells were cultured in the conditions recommended by the vendors or as reported (Table 1).

### HS, Heparin, and Chemically Modified Heparins

The porcine intestinal heparin (16 kDa) and HS (11.7kDa) were obtained from Celsus Laboratories, Cincinnati, OH, United States. *N*-desulfated/*N*-acetylated heparin (NDS-HP, 14kDa) and oversulfated heparin (OS-HP, 16 kDa) were obtained from Glycomed Inc. (Alameda, CA, United States). 2-O-desulfated heparin (2DS-HP, 13 kDa) and 6-O-desulfated heparin (6DS-HP, 13 kDa) were prepared as we reported previously (Wang et al., 2002; Bobardt et al., 2004; Zhang et al., 2012, 2013; Zhang F. et al., 2014). All the chemically-modified heparins have no anticoagulant activity as determined by amidolytic anti-factor Xa assay and were negative for endotoxin in the Limulus test (Wang et al., 2002).

### Flow Cytometry

Cells were seeded in collagen I coated cell culture dish. Confluent cells were treated with or without heparinase I, II, and III

from THP-1; NDS-HP, *N*-desulfated/*N*-acetylated heparin; NS, *N*-sulfation; OS-HP, oversulfated heparin; PMA, phorbol 12-myristate-13-acetate; RT, room temperature; SARS, severe acute respiratory syndrome; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SPR, surface plasmon resonance; Sulf1, heparan sulfate 6-O-endosulfatase-1; Sulf2, heparan sulfate 6-O-endosulfatase-2; SV2-S, SARS-CoV-2 spike protein; TCGA, The Cancer Genome Atlas.



**TABLE 1 |** Summary of cells and their culture condition.

Name	Cell type	Source	Culture condition
PNT2	Normal human prostate epithelium immortalized with SV40	Millipore Sigma	RMPI 1640, 2 mM Glutamine, 10% FBS
SH-SY5Y	Human neuroblastoma cell line	ATCC	DMEM/F-12, 10% FBS
H441	Human lung adenocarcinoma epithelial cell	Thermo Fisher Scientific	RMPI 1640, 2 mM Glutamine, 10% FBS
HLMVEC	Human lung microvascular endothelial cells	Lonza	EGM-2 MV bullet kit
MLEC	Mouse lung endothelial cells	Generated in own lab	DMEM, 10% FBS
A549	Human adenocarcinoma alveolar basal epithelial cells	ATCC	RMPI 1640, 2 mM Glutamine, 10% FBS
293T	Human embryonic kidney 293 cells immortalized with SV40	ATCC	DMEM, 10% FBS
mTHP-1	Human macrophage induced from THP-1	ATCC	RPML-1640, 2 mM Glutamine, 10% FBS 100 ng/mL PMA
Calu-3	Human lung adenocarcinoma epithelial cell	AddexBio	AddexBio-formulated EMEM, 10% FBS

(HSase, 5 mU/ml of each) in DPBS at 37°C for 30 min and followed by adding collagenase IV for additional 30 min incubation at 37°C. Then cells were detached and suspended in DPBS-1% BSA for cell surface SV2-S binding or expression of HS or ACE2 analysis. SV2-S binding on cell surface was measured by incubating cells with 250 nM His-tagged SV2-S (Sino Biological, 40591-V08B1). Cell surface-bound SV2-S was measured by flow cytometry (Canto II 488 Laser) after staining with Alexa Fluor488-conjugated goat anti-His antibody. Cell surface HS level was examined after sequential staining with anti-HS antibody 10E4 (Amsbio, 370255-1) at 2.5 µg/ml and Alexa Fluor488 goat anti-mouse IgM secondary antibody. Cell surface ACE2 level was determined after staining with anti-ACE-2 antibody (R&D Systems, AF933) at 1 µg/ml and Alexa Fluor 488 donkey anti-goat IgG secondary antibody. Dead cells were gated based on DAPI staining. The collected data were analyzed using Prism 8 (GraphPad Software). Relative median fluorescence intensity (MFI) was calculated by subtracting MFI of isotype control from antigen staining or BSA from SV2-S.

### Cell-Based Enzyme-Linked Immunosorbent Assay (ELISA)

HS-deficient MLEC were seeded at  $4 \times 10^4$  cells per well in 96 well plates. After overnight culture, the cells were washed and fixed in 4% paraformaldehyde at room temperature (RT) for 15 min. After washing and blocking in 5% BSA overnight at 4°C, 20 nM His-tagged SV2-S was applied and incubated for 1 h at RT. Then cells were washed and incubated with anti-His secondary antibody conjugated with HRP for 1 h at RT. After intensive washing, ELISA substrate was applied. After 25 min, the color development was stopped with 0.5 M HCl. Absorbance at 450 nm was measured. Cells were further stained with Janus Green (Abcam, ab111622) for cell density. Absorbance at 595 nm was measured representing cell density and used to normalize cell surface SV2-S binding.

### Inhibitory Effect of Heparin and Chemically-Modified Heparins on Cell Surface SV2-S Binding

Cell-based ELISA protocol described above was used to determine inhibition. A549 cells were incubated

with 20 nM His-tagged SV2-S in the absence or presence of different concentrations of heparin or chemically modified heparin. The binding was normalized to wells without the test compound to determine inhibitory effect.

### Interaction of SV2-S With Immobilized Heparin in Surface Plasmon Resonance (SPR) Analysis

Surface plasmon resonance analyses were performed using BIAcore 3000 (GE Healthcare, Uppsala, Sweden) to measure the binding kinetics and interaction affinity of SV2-S to heparin as we did previously (Condac et al., 2012; Zhang et al., 2013; Zhang B. et al., 2014). Biotinylated heparin was immobilized on streptavidin-coated Sensor SA chips (Zhang et al., 2013; Zhang B. et al., 2014; Zhang F. et al., 2014). For direct binding analysis, SV2-S D614 (Sino Biological, 40591-V08H) and SV2-S G614 (Sino Biological, 40591-V08H3) were diluted in a running HBS-EP buffer containing 0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% surfactant P20, at pH 7.4. Different concentrations of the proteins were injected over the heparin chip at a flow rate of 30 µl/min. At the end of the injection, a buffer was flowed over the sensor surface to allow dissociation. After 3 min, the sensor surface was regenerated by injecting 30 µl of 2 M NaCl. The response was monitored as a function of time (sensorgram) at 25°C. The sensorgrams of various SV2-S concentrations were fit globally to obtain association rate constant ( $k_a$ ), dissociation rate constant ( $k_d$ ), and equilibrium dissociation constant ( $K_D$ ) using the BiaEvaluation software version 4.0.1 (GE Healthcare, Uppsala, Sweden) and assuming a 1:1 Langmuir model.

### Transient and Stable Expression of ACE2 in 293T Cells

293T-ACE2-GFP expression cell lines were established by transfecting 293T cells with an ACE2-GFP expression vector (OriGene Technologies, RG208442). Two days after transfection, transient 293T-ACE-GFP cells were used for assessing SV2-S binding by flow cytometry. After transfections, GFP positive cells were sorted and seeded at one cell per well of 96 well plates.

The transfected 293T cell clones with stable GFP and ACE2 expression were used for the pseudotyped virus infection assay.

## Production of SARS-CoV-2 Pseudotyped Virus

The following plasmids were obtained from Addgene including pcDNA3.1-SARS2-Spike mammalian expression plasmid (#145032), NL4-3 mCherry Luciferase dual reporter vector (#44965), and the control env expression vector, VSV-G (pMD2.G) (#12259). The SARS-CoV-2 G614 mutant expression vector was generated as recently reported (Zhang et al., 2020a). The SARS-CoV-2 pseudotyped virus was produced by transfecting 293T cells with 10 µg of NL4-3 mCherry Luciferase and 10 µg env expression vector: either VSV-G (pMD2.G) or pcDNA3.1-SARS2 Spike using polyethylenimine (PEI) (25 kDa, 1 µg/µL). The virus supernatant was collected at 48 h after transfection and concentrated at 1:100 ratio using Lenti-X<sup>TM</sup> Concentrator (TaKaRa/Clontech, 631231). Viral titer was determined using Lenti-X<sup>TM</sup> qRT-PCR Titration Kit (TaKaRa/Clontech, 631235).

## Pseudotyped Virus Infection Assay

293T-ACE2 cells were seeded at  $1.25 \times 10^4$  per well in poly-L-Lysine coated white 96 well plate and cultured overnight. Following, the wells were washed with DMEM and incubated with SARS-CoV-2-S pseudotyped virus and positive control VSV pseudotyped virus. At 6 h post-infection, the medium was changed with fresh DMEM supplemented with 10% FBS. At 48 h post-infection, the cells were lysed and measured for luciferase activity using a luciferase assay kit (Promega, E1500).

## Inhibition of SARS-CoV-2 Pseudotyped Virus Infection Assay

Pseudotyped virus infection assay was performed as described above. Before adding the pseudotyped virus into the well, cells were pre-incubated with DMEM supplemented with HSase, surfen (Sigma, S6951), HS, heparin, NDS-HP, 6DS-HP, 2DS-HP, or OS-HP for 30min at 37°C.

## Ext1 Overexpression in 293T-ACE2 Cell Line

We used a human *Ext1* expression lentiviral vector (pLenti-GIII-CMV-GFP-2A-Puro based), which was obtained from abmgood (#195860610395) to overexpress *Ext1*. For lentivirus production, the *Ext1* expression vector (10 µg) and the two packaging viral vectors, pMD2.G (5 µg) and psPax2 (5 µg), both of which were obtained from Addgene, were co-transfected into 293T cells using PEI. The virus supernatant was collected at 48 h after transfection and concentrated at 1:100 ratio using a Lenti-X<sup>TM</sup> Concentrator kit. To overexpress *Ext1*, 293T-ACE2 cells were inoculated with 1:100 diluted *Ext1* overexpression lentivirus with 8 µg/ml polybrene (Santa Cruz Biotech, sc-134220). Two days post lentivirus infection, cells were used for the experiment and *Ext1* overexpression efficiency was assessed by RT-qPCR.

## Reverse Transcription-Quantitative PCR (RT-qPCR)

RT-qPCR was performed to analyze the mRNA expression levels of *Ext1*. TRIzol reagent (Invitrogen) was used to extract the RNA from the 293T-ACE2 cells and then reverse transcribed into cDNA using the Bio-Rad reverse transcription system (#1708840). Bio-Rad SYBR-Green Supermix (#1708880) was used for qPCR analysis. The primer sequences used in this experiment were: *Ext1* forward, 5'-GCTCTTGTCTCGCCCTTTTGT-3' and reverse, 5'-GGTGCAAGCCATTCTACC-3'; *glyceraldehyde 3-phosphate dehydrogenase (GAPDH)* forward, 5'-GTATTGGGCGCCTGGTCACC-3' and reverse, 5'-CGGGAAGATGGTGATGG-3'. The PCR-amplified mRNA was quantified and the results were normalized against *GAPDH* expression. The  $2^{-\Delta\Delta Cq}$  method was used to calculate the relative mRNA expression level.

## Analysis of HS Gene and ACE2 Expression in Lung Squamous Cell Carcinoma Patients and Health Control

Valuable cancer-related RNA sequencing data deposited in The Cancer Genome Atlas (TCGA) has provided ample opportunities for data mining and a deeper understanding of gene functions in cancer (Cancer Genome Atlas Research et al., 2013). For this purpose, the Gene Expression Profiling Interactive Analysis 2 (GEPIA2) software was developed and has aided the investigation of various genes in multiple cancer types leading to identification of potential biomarkers and therapeutic targets (Liang et al., 2018; Cai et al., 2019; Tang et al., 2019). To determine if HS gene and *ACE2* expressions are dysregulated in lung squamous cell carcinoma (LUSC), using the GEPIA2 software, the data deposited in TCGA including 486 LUSC patients with 50 normal controls were analyzed for their differential expression and relative expression abundance. The differential gene expression among LUSC pathological stage I-IV was analyzed also. The HS gene analyzed included *Ext1-2*, *Ndst1-4*, *Hs2st*, *Hs6st1-3*, *Hs3st1-6*, and *Sul1-2* (Qiu et al., 2018).

## Statistical Analysis

Statistical analysis was carried out with Prism 8 for Macintosh. All data are presented as mean  $\pm$  SD or mean  $\pm$  SEM and analyzed using a student's *t*-test for two-group comparison. The two-parameter correlation was determined using the e Pearson's correlation coefficient analysis. In all tests, *P*-value < 0.05 was considered statistically significant.

## RESULTS

### Cell Surface SV2-S Binding Depends on and Correlates With HS Expression

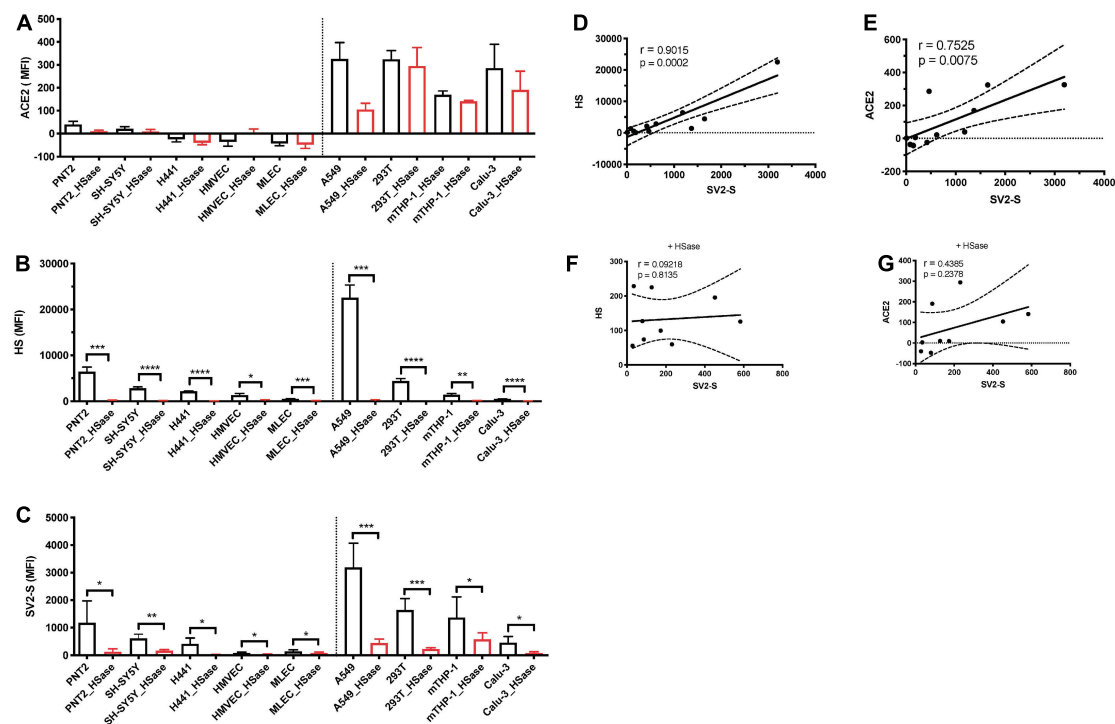
Reported studies have determined that HS is required for cell surface SV2-S binding, but it remains unclear if the binding is cell-context dependent including ACE2 co-expression

(Clausen et al., 2020). To address this issue, we analyzed available cell lines and primary cells available in our lab, including PNT2 (normal human prostate epithelial), SH-SY5Y (human neuroblastoma), H441 and Calu-3 (human lung adenocarcinomas), primary HMVEC, MLEC lines, A549 (human adenocarcinoma alveolar basal epithelial), 293T (human embryonic kidney), and mTHP-1 (macrophage differentiated from THP-1, a human acute monocytic leukemia cell line, by phorbol 12-myristate-13-acetate) (Table 1). Among these cells, A549, 293T, mTHP-1, and Calu-3 also express low levels of ACE2 on cell surface (Figure 1A). These cells express different levels of HS on cell surface assessed by anti-HS antibody 10E4 staining followed by flow cytometry analysis (Figure 1B). Cell surface binding of recombinant SV2-S of these cell lines was assessed by flow cytometry analysis (Figure 1C). The cell surface SV2-S binding positively correlated with the cell surface HS expression with a Pearson's correlation coefficient of 0.9015 (Figure 1D) and with cell surface ACE2 expression of a Pearson's correlation coefficient of 0.7515 (Figure 1E). Treating the cells with HSase, which diminished cell surface HS, did not affect ACE2 expression (Figures 1A,B). The HSase treatment diminished cell surface SV2-S binding and the positive correlation of cell surface ACE2 expression with SV2-S binding

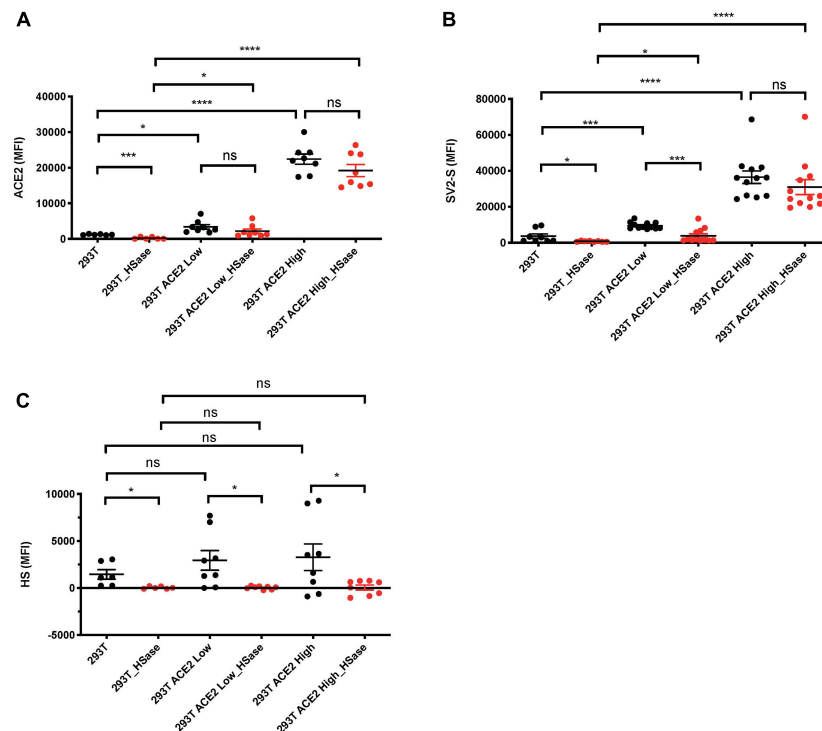
(Figures 1C,E,G), showing that SV2-S binding is determined by cell surface HS even if endogenous ACE2 is expressed on cell surface.

## HS Determines Cell Surface SV2-S Binding Only When ACE2 Is Expressed at Low Levels

We observed that cell surface SV2-S binding was determined by HS, but this occurred in the absence or presence of low levels of endogenous ACE2 expression. Since SV2-S has a high binding affinity to ACE2 (Lu and Sun, 2020), we suspected that HS dependence would be attenuated when ACE2 expression is high. To test this idea, 293T cells were transfected with an ACE2-GFP expression plasmid, and the transient 293T-ACE2-GFP-expressing cells were gated into two populations: the ACE2-GFP low and high expression cells based on GFP expression. Anti-ACE2 antibody staining confirmed that the ACE2-GFP low and -high expression cells express low- and high ACE2 on cell surface, respectively (Figure 2A). The transiently-transfected 293T-ACE2-GFP cells showed increased cell surface SV2-S binding, positively correlating with their cell surface ACE2 expression (Figure 2B). Treating the cells with HSase diminished



**FIGURE 1 |** Cell surface SV2-S binding depends on and correlates with expression levels of HS. **(A,B)** Cell surface expressions of ACE2 and HS were assessed by flow cytometry after incubation with anti-ACE2 and anti-HS antibody 10E4, respectively, and sequential Alexa Fluor 488-tagged secondary antibody. A portion of the cells was treated with heparinases I-III (HSase) before the antibody staining. **(C)** The binding of SV2-S on cell surface was assessed by flow cytometry after sequential incubation with his-tagged SV2-S and Alexa Fluor 488-tagged anti-his antibody. A portion of the cells was treated with HSase before SV2-S binding. Median fluorescence intensity (MFI) was calculated by subtracting MFI of BSA binding or isotype-match naive control antibody binding. The experiments were repeated 3–6 times and an unpaired *t*-test was performed for comparison of the same cells without vs. with HSase treatment. Data are presented as mean  $\pm$  SD.  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ,  $****p < 0.0001$ . **(D–G)** The Pearson's correlation coefficient between SV2-S binding and cell surface expression level of HS or ACE2 was determined by analyzing the data collected in **(A–C)** using Prism 8. The cells analyzed in **(F,G)** were pre-treated with HSase.



**FIGURE 2 |** Heparan sulfate (HS) plays a major role in mediating cell surface SV2-S protein binding only when ACE2 expresses at a low level. 293T cells were transfected with ACE2-GFP or scramble-GFP plasmid. In SV2-S binding analysis, the transient 293T-ACE2-GFP-expressing cells were gated into two populations: the ACE2-GFP-low and -high expression cells based on GFP expression. **(A)** Cell surface ACE2 expression was assessed by flow cytometry after staining with anti-ACE2 antibody and Alexa Fluor 647-conjugated secondary antibody. A portion of the cells was treated with HSase before the specific antibody staining. **(B)** The binding of SV2-S on cell surface was assessed by flow cytometry after sequential incubation with his-tagged SV2-S and Alexa Fluor 647-tagged anti-his antibody. A portion of the cells was treated with HSase before SV2-S binding. **(C)** Cell surface HS expression was assessed by flow cytometry after staining with anti-HS antibody and Alexa Fluor 647-conjugated secondary antibody. A portion of the cells was treated with HSase before the specific antibody staining. MFI was calculated by subtracting MFI of BSA binding or isotype-match naive control antibody staining. The experiments were repeated 6- to 12- times and an unpaired *t*-test was performed for comparison of the same cells without vs. with HSase treatment. Data are presented as mean  $\pm$  SD. ns, not significant; \**p* < 0.05, \*\*\**p* < 0.001, \*\*\*\**p* < 0.0001.

cell surface HS (**Figure 2C**) and SV2-S binding on the low ACE2-expressing 293T cells (**Figure 2B**). The HSase treatment did not significantly reduce cell surface SV2-S binding on the high ACE2-expressing 293T cells (**Figure 2B**). These data show that cell surface HS plays a determining role in mediating cell surface SV2-S binding only when ACE2 expression is low or absent.

### The *N*- and 6-*O*-Sulfation Are the Major Modifications Required for Heparin to Bind SV2-S and the Binding Also Depends on Overall Sulfation

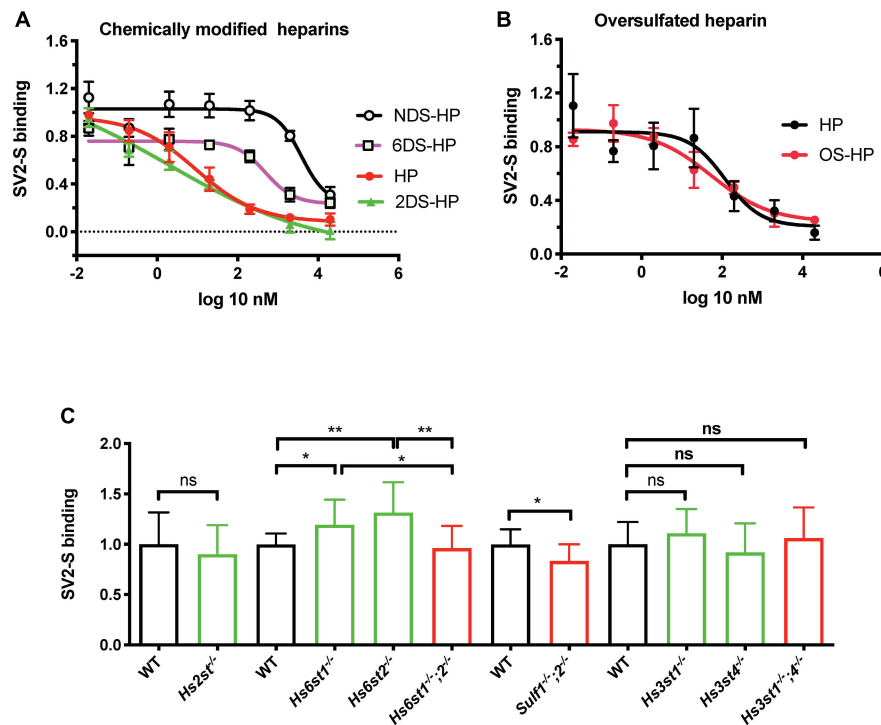
Heparin has been reported to bind SV2-S, but the related structural information remains largely unknown. Heparin possesses *N*-sulfation (NS), 2-*O*-sulfation (2S), 3-*O*-sulfation (3S), and 6-*O*-sulfation (6S) modifications, which form binding-sites for protein ligands (Wang et al., 2002; Zong et al., 2016, 2017). To determine the type of the sulfation modification that is required for heparin to bind SV2-S, we assessed the binding of SV2-S to confluent A549 cells in the presence of heparin, or

chemically modified heparins including NDS-HP, 2DS-HP, 6DS-HP, or OS-HP. Heparin showed a strong inhibitory activity with an IC<sub>50</sub> of  $9.1 \pm 2.1$  nM and 2DS-HP showed a comparable inhibitory activity with an IC<sub>50</sub> of  $2.0 \pm 28.7$  nM (**Figure 3A**). The NDS-HP and 6DS-HP showed much weaker inhibitory activities with IC<sub>50</sub> values at  $3828 \pm 16.3$  nM and  $489.8 \pm 2.1$  nM, respectively (**Figure 3A**). The OS-HP showed higher inhibitory activity than heparin (**Figure 3B**),  $118.9 \pm 2.9$  nM for heparin and  $53.7 \pm 2.7$  nM for OS-HP in new sets of the experiment. These results indicated that heparin requires its NS and 6S, but not 2S, to bind SV2-S, and that binding also depends on the overall sulfation level.

### 2-, 6- and 3-*O*-Sulfation Are Not Required for Cell Surface HS to Bind SV2-S, and the Cell Surface HS to Bind SV2-S Is Mainly Determined by Its Overall Sulfation

We examined serial HS mutant MLEC lines which were generated in our lab (Wijelath et al., 2010; Qiu et al., 2018),





**FIGURE 3 |** The modification of heparin and cell surface HS involved in SV2-S binding. **(A,B)** Inhibition on the binding of SV2-S to cultured A549 cells. After incubation in the absence or presence of different concentrations of heparin (HP), NDS-HP; 2DS-HP, 6DS-HP, or OS-HP, SV2-S that bound to A549 cells was quantified by cell-based ELISA and normalized to the cell density. The SV2-S binding was further normalized to the wells without inhibitors. The experiments were repeated at least three times. Data are presented as mean  $\pm$  SEM. **(C)** SV2-S binding on MLEC surface. The tested HS mutant MLEC lines include wildtype (WT) control, *Hs2st*<sup>-/-</sup> (lack of 2S), *Hs6st1*<sup>-/-</sup> (reduced 6S), *Hs6st2*<sup>-/-</sup> (normal 6S), *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> (no 6S), *Sulf1*<sup>-/-</sup>;2<sup>-/-</sup> (increased 6S), *Hs3st1*<sup>-/-</sup> (reduced 3S), *Hs3st4*<sup>-/-</sup> (reduced 3S), and *Hs3st1*<sup>-/-</sup>;4<sup>-/-</sup> (greatly reduced 3S). Cell surface SV2-S protein binding was assessed by cell-based ELISA after incubation of fixed cells with his-tagged SV2-S and normalization to the cell density. The experiments were repeated 6- to 12- times and the unpaired t-test was performed for comparison between the wildtype and the mutants. Data are presented as mean  $\pm$  SD. Ns, not significant; \**p* < 0.05, \*\**p* < 0.01.

including the cells that are deficient in *HS 2-O-sulfotransferase* (*Hs2st*<sup>-/-</sup>), *HS 6-O-sulfotransferase-1* (*Hs6st1*<sup>-/-</sup>), *Hs6st2*<sup>-/-</sup>, both *Hs6st1* and *Hs6st2* (*Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup>), both *HS 6-O-endosulfatase-1* and *HS 6-O-endosulfatase-2* (*Sulf1*<sup>-/-</sup>;2<sup>-/-</sup>), *3-O-sulfotransferase-1* (*Hs3st1*<sup>-/-</sup>), *Hs3st4*<sup>-/-</sup>, or both *Hs3st1* and *Hs3st4* (*Hs3st1*<sup>-/-</sup>;4<sup>-/-</sup>) to extend the biochemical heparin-SV2-S binding studies and to determine the sulfation modification required for cell surface HS to bind SV2-S (**Figure 3C**). The cell surface SV2-S binding was assessed using a cell-ELISA approach (Wang et al., 2002; Qiu et al., 2018). *Hs2st*<sup>-/-</sup> MLEC HS completely lacks 2S with a slight increase of overall sulfation (Qiu et al., 2018). The *Hs2st*<sup>-/-</sup> MLECs showed a cell surface SV2-S binding comparable to wildtype control, indicating 2S is not required for cell surface HS to bind SV2-S. HS 6S is co-determined by *Hs6sts* and *Sulfs*. The *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> MLEC HS completely lacks 6S with normal overall sulfation (Qiu et al., 2018) and showed cell surface SV2-S binding comparable to wildtype control, indicating 6S is not essential for cell surface HS to bind SV2-S (**Figure 3C**). Intriguingly, the *Hs6st1*<sup>-/-</sup> and *Hs6st2*<sup>-/-</sup> MLECs both exhibited increased cell surface SV2-S binding (**Figure 3C**). The *Hs6st1*<sup>-/-</sup> and *Hs6st2*<sup>-/-</sup> MLEC HS possess a reduced and normal level of 6S, respectively, with normal overall sulfation (Qiu et al., 2018), indicating

the increased cell surface SV2-S binding was due to alteration of HS fine structure, rather than overall sulfation, emerging fine structure contributes to HS binding to SV2-S. To test if increasing 6S would also alter cell surface SV2-S binding, we examined *Sulf1*<sup>-/-</sup>;2<sup>-/-</sup> MLECs which possesses increased 6S with compensatory reductions in NS and 2S and normal overall sulfation (Qiu et al., 2018). The *Sulf1*<sup>-/-</sup>;2<sup>-/-</sup> MLECs exhibited decreased cell surface SV2-S binding (**Figure 3C**) and illustrated alteration of HS structure, rather than overall sulfation, reduced cell surface SV2-S binding, alternatively supporting fine structure contributes to HS binding of SV2-S. 3S is the last sulfation modification in HS biosynthesis and rare in mature HS (Thacker et al., 2014; Zhao et al., 2020). The *Hs3st1*<sup>-/-</sup> and *Hs3st4*<sup>-/-</sup> MLECs both showed reduced binding of antithrombin (AT), a ligand that strictly requires 3S for binding (Qiu et al., 2018). The *Hs3st1*<sup>-/-</sup>;4<sup>-/-</sup> MLECs have further reduced AT binding (Qiu et al., 2018). The reduction of cell surface AT binding reflects that 3S is reduced in *Hs3st1*<sup>-/-</sup>, *Hs3st4*<sup>-/-</sup>, and *Hs3st1*<sup>-/-</sup>;4<sup>-/-</sup> MLEC HS. The HS on *Hs3st4*<sup>-/-</sup> MLECs, but not *Hs3st1*<sup>-/-</sup> and *Hs3st1*<sup>-/-</sup>;4<sup>-/-</sup> MLECs, has slightly reduced overall sulfation (Qiu et al., 2018). The *Hs3st1*<sup>-/-</sup>, *Hs3st4*<sup>-/-</sup> and *Hs3st1*<sup>-/-</sup>;4<sup>-/-</sup> MLECs had cell surface SV2-S binding comparable to wildtype control, showing that 3S is not required

for cell surface HS to bind SV2-S (**Figure 3C**). In summary, these HS mutant cell studies revealed that individual O-sulfation type is not essential for HS to bind SV2-S on cell surface, although fine structure contributes, supporting that overall sulfation is more important for cell surface HS to bind SV2-S.

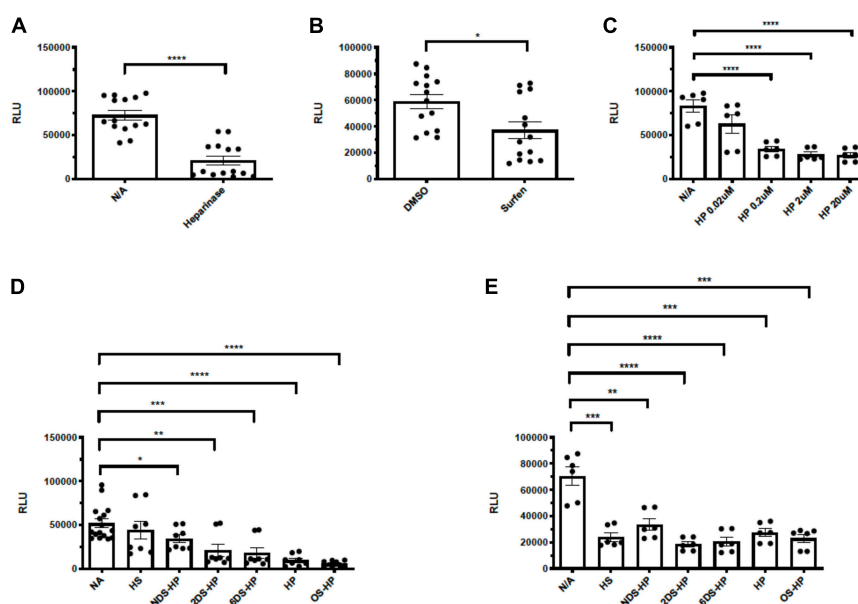
## The Entry of SARS-Cov-2 Pseudotyped VSV Into 293T-ACE2 Cells Depends on Cell Surface HS and Can Be Inhibited by HS Antagonist and Chemically Modified Heparins

Our tested cell lines A549, mTHP-1, 293T, and Calu-3 express low levels of ACE2 on the cell surface and showed little to no infection by SV2-S pseudotyped VSV. Thus, they could not be used to assess the overall role of HS in SARS-CoV-2 infection. Alternatively, we generated a 293T-ACE2 cell line which stably expresses an increased, albeit low, level of ACE2 expression and could be effectively infected by SV2-S pseudotyped virus. We assessed entry of SV2-S pseudotyped virus into the 293T-ACE2 cells after treatment with HSase to degrade cell surface HS. HSase treatment reduced more than 72% SARS-Cov-2 pseudovirus entry into 293T-ACE2 cells, showing that binding of SV2-S to cell surface HS plays a key role in mediating the virus infection of host cells (**Figure 4A**). The 293T-ACE2 cells were inoculated with SV2-S pseudovirus in the presence of surfen [a small molecule antagonist of HS (Schuksz et al., 2008)], heparin, or chemically modified non-anticoagulant heparins including NDS-HP, 2DS-HP, 6DS-HP, and OS-HP to test if blocking the interaction

of cell surface HS with SV2-S inhibited pseudovirus infection, Surfen at 20  $\mu\text{M}$  inhibited SARS-Cov-2 pseudovirus entry by 52% (**Figure 4B**). Heparin at low concentrations (starting at 0.02  $\mu\text{M}$ ) showed a potent, concentration-dependent inhibition with a maximal inhibition (67%) at  $> 2 \mu\text{M}$  (**Figure 4C**). At 2  $\mu\text{M}$ , the de-sulfated heparins showed weaker but significant inhibitory activities with the following order: heparin  $>$  6DS-HP = 2DS-HP  $>$  NDS-HP, and OS-HP showed a higher inhibitory activity than heparin (**Figure 4D**). At 20  $\mu\text{M}$  concentration, heparin, HS, the desulfated heparins, and OS-HP showed comparable inhibition of SARS-Cov-2 pseudovirus infection (**Figure 4E**).

## G614 Mutation Increases SV2-S Binding Affinity to Heparin, and the Entry of SARS-CoV-2 G614 Mutant Into Host Cells Depends on HS and Can Be Inhibited by Heparin

The SARS-CoV-2-S G614 variant mutated D614 into G614 and has predominated globally (Korber et al., 2020; Martin et al., 2020). A recent study reported by Zhang et al. (2020a) uncovered that the G614 mutation increased 9-fold infection of the SARS-CoV-2. Currently, the molecular mechanism underlying the increased infectivity of the mutant remains incompletely understood. We tested if the mutation alters the interaction of SV2-S with heparin by determining the kinetic binding of SV2-S G614 mutant to immobilized heparin in SPR analysis. The sensorgrams of various concentrations of SV2-S were fit globally to obtain association rate constant ( $K_a$ ), dissociation



**FIGURE 4 |** Heparinases, surfen, heparin, and chemically modified non-anticoagulant heparins block SV2-S pseudotyped virus infection of 293T-ACE2 cells. The 293T-ACE2 cells were infected with SV2-S pseudotyped virus expressing luciferase after the cells were treated with HSase (5 mU/ml) (**A**) or in the presence of surfen (20  $\mu\text{M}$ ) (**B**), heparin (HP, **C**), HS or chemically-modified heparins at 2- and 20  $\mu\text{M}$  (**D,E**). The luciferase activity was measured 48 h after the virus infection. VSV pseudotyped virus was used as a control to make sure each well has comparable cell numbers (data are not shown). The experiments were repeated at least 3 times and an unpaired *t*-test was performed for the two-group comparison. Data are presented as mean  $\pm$  SD. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . RLU, relative light unit.

rate constant ( $K_d$ ) and equilibrium dissociation constant ( $K_D$ ) (Table 2) using the BiaEvaluation software and assuming a 1:1 Langmuir model. The wildtype D614 SV2-S and mutant G614 had binding affinity  $K_D$  values of 100 and 44 nM, respectively, indicating the G614 mutation strengthened the binding of SV2-S to heparin (Figures 5A,B). In addition, the SV2-S G614-pseudotyped virus showed ~15-fold higher infection of the 293T-ACE2 cells than the SV2-S D614-pseudotyped virus (Figure 5C), similar to the recent report in literature (Zhang et al., 2020a). Pre-treatment of 293T-ACE2 cells with HSase or infection in the presence of 2  $\mu$ M heparin both inhibited the SV2-S G614 pseudovirus entry into the 293T-ACE2 cells, showing that the infection by the SV2-S G614 mutant depends on host cell surface HS and can be inhibited by heparin (Figure 5D).

## Dysregulated HS Expression May Contribute to the Increased Vulnerability of Lung Cancer Patients to SARS-Cov-2 Infection

Several retrospective cohort studies revealed that cancer patients are at increased risk from COVID-19 severity and fatality (Liang et al., 2020; Mehta et al., 2020; Poortmans et al., 2020; Rogado et al., 2020; Yu et al., 2020; Zhang et al., 2020b). Among the major cancer types analyzed, lung cancer patients with COVID-19 have the highest fatality (Liang et al., 2020; Mehta et al., 2020; Poortmans et al., 2020; Rogado et al., 2020; Yu et al., 2020; Zhang et al., 2020b). The aggravating factor that contributes to the higher fatality of lung cancer is unknown. Using GEPIA2 (Tang et al., 2017), we analyzed the lung cancer mRNA transcript data deposited in TCGA including 486 LUSC patients with 50 normal controls. The expression of HS biosynthetic genes, including *Ext1*, *Ext2*, *Hs6st1*, *Hs6st2*, *Sulf1*, and *Sulf2* are upregulated whereas *Ndst1* is downregulated (Figure 6A), and are not different among pathological stages of LUSC (Stage I-IV) (Figure 6B). In comparison, the expression of major HS genes, including *Ext1-2*, *Ndst1-2*, *Hs2st1*, *Hs6st1-2*, and *Sulf1-2*, are much more abundant than ACE2 (Figure 6C), suggesting that ACE2 expression is low in LUSCs compared to HS. Therefore, HS may serve as the major binding site for SV2-S to mediate SARS-CoV-2-S infection of LUSCs. In HS biosynthesis, *Ext1* and *Ext2* form functional heterodimers to co-polymerize HS disaccharide backbone and determine the HS chain length (expression level) in general (Kraushaar et al., 2010, 2012; Rai et al., 2020). We overexpressed *Ext1* in 293T-ACE2 cells to assess the functional consequence of upregulated *Ext1* expression in the vulnerability of lung cancer

to SARS-CoV-2 (Figure 6D). The *Ext1* overexpressing 293T-ACE2 cells showed higher infection of SV-2S pseudovirus than the scrambled control (Figure 6E), suggesting the *Ext1* upregulation may be one of the mechanisms underlying the higher susceptibility of lung cancer patients to SARS-CoV-2 infection.

## DISCUSSION

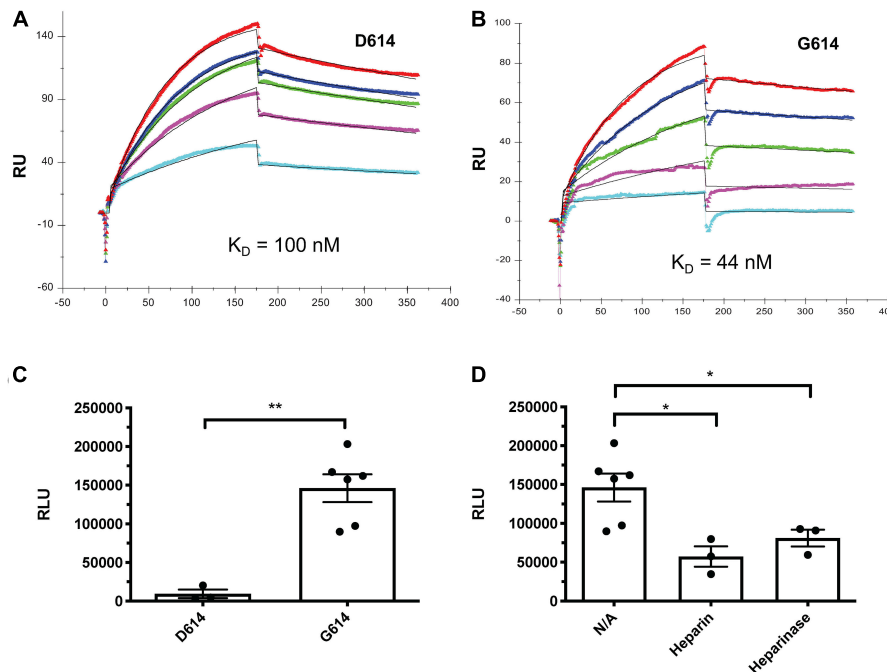
In this study, we assessed 9 cell types derived from different tissues, including epithelial, neuronal, macrophage, kidney and endothelial cells to determine the relative importance of HS and ACE2 in cell surface SV2-S binding. Four of the cell lines (A549, 293T, mTHP-1, and Calu-3) express low levels of ACE2. We observed that cell surface SV2-S binding depends on and highly correlates with HS expression. The cell surface SV2-S binding on ACE2-expressing cell lines also correlates with ACE2 level, but with a lower correlation coefficient than HS. More importantly, the correlation between cell surface SV2-S binding and ACE2 expression was diminished when the cells were pretreated with HSase to remove cell surface HS. These observations indicate that cell surface HS, not ACE2, mediates in the initial phase of cell surface binding of SV2-S. We also observed that HSase treatment greatly inhibited SV2-S pseudotyped virus infection of 293T-ACE2 cells, indicating that cell surface HS is essential for SARS-CoV-2 infection. Intriguingly, when we tested with high ACE2-expressing 293T cells, the dependence of cell surface HS for SV2-S binding was diminished. One explanation for this discrepancy is HS acts as a co-receptor in SV2-S-ACE2 signaling, as suggested recently (Clausen et al., 2020). Low expression of ACE2 on the cell surface represents the limiting factor for SV2-S-ACE2 signaling and requires the abundantly-expressed HS to act as a co-receptor to enrich and facilitate SV2-S binding to ACE2. However, when ACE2 expression is high, the HS co-receptor function is dispensable. Some studies have reported that interferon-mediated induction of ACE2 expression occurs upon SARS-CoV-2 infection (Fu et al., 2020; Su and Jiang, 2020; Tan et al., 2020), suggesting that blocking HS-SV2-S binding by heparin may not be sufficient (Efendizade et al., 2020; Hippensteel et al., 2020; Menezes-Rodrigues et al., 2020) and would need to be combined with other mechanism-based agents to most effectively to treat and/or prevent COVID-19.

In previous studies, we determined that monomeric SV2-S-Fc exhibited an exceptionally high-affinity binding to heparin in SPR analysis with a  $K_D = 40$  pM (Kim et al., 2020). In current SPR analysis, we tested monomeric His-tagged SV2-S from the same vendor and determined that the His-tagged SV2-S remained high-affinity binding to heparin too, but the  $K_D$  was reduced to 100 nM. The difference in binding affinity may be due to the Fc-tagged SV2-S is a disulfide-bonded dimer and contains two SV2-S monomer whereas His-tagged SV2-S is a monomer. Meanwhile, SV2-S contains extensive N- and O-glycosylation (Shajahan et al., 2020; Wang et al., 2020; Watanabe et al., 2020; Xu et al., 2020) and the glycosylation may play a role in protein folding thereby affects the interaction of SV2-S with heparin. It is possible the difference in binding affinity might also be contributed by their

**TABLE 2 |** Summary of kinetic data of wild type spike and mutant spike protein-heparin interactions\*.

Interaction	$k_a$ (1/MS)	$k_d$ (1/S)	$K_D$ (M)
SV2-S D614	$1.3 \times 10^4 \pm 132$	$1.3 \times 10^{-3} \pm 1.6 \times 10^{-5}$	$1.0 \times 10^{-7}$
SV2-S G614	$1.2 \times 10^4 \pm 374$	$5.3 \times 10^{-4} \pm 3.2 \times 10^{-5}$	$4.4 \times 10^{-8}$

\*The data with ( $\pm$ ) in parentheses are the standard deviations (SD) from global fitting of five injections.



**FIGURE 5 |** The SV2-S variant G614 protein has a higher binding affinity to heparin than the wildtype SV2-S D614 protein, and the entry of SARS-CoV-2 G614 pseudotyped virus into host cells depends on cell surface HS and can be inhibited by heparin. **(A,B)** SPR binding kinetics sensorgrams between heparin and SV2-S D614 or SV2-S G614. Concentrations of D614 were (from top to bottom): 1,000, 800, 600, 400, and 200 nM, respectively **(A)**, and G614 were (from top to bottom): 1,000, 600, 400, 200, and 100 nM, respectively **(B)**. The black curves are the fitting curves using a 1:1 Langmuir model from BIAevaluate 4.0.1. **(C)** Infection of 293T-ACE2 cells by the same titer of SARS-CoV2-D614 and SARS-CoV2-G614 pseudovirus. **(D)** SARS-CoV2-G614 pseudovirus depends on HS to infect host cells. The 293T-ACE2 cells were infected with SARS-CoV2-G614 pseudovirus expressing luciferase after the cells were treated with heparinases I-III (5 mU/ml) or in the presence of heparin (2  $\mu$ M). The luciferase activity was measured 48 h after the virus infection. VSV pseudotyped virus was used as a control to make sure each well has comparable cell numbers (data are not shown). The experiments were repeated at least 3 times and an unpaired t-test was performed for two-group comparison. Data are presented as mean  $\pm$  SD. \* $p < 0.05$ , \*\* $p < 0.01$ .

glycosylation difference between SV2-S-Fc and His-tagged SV2-S; however, this needs to be experimentally determined.

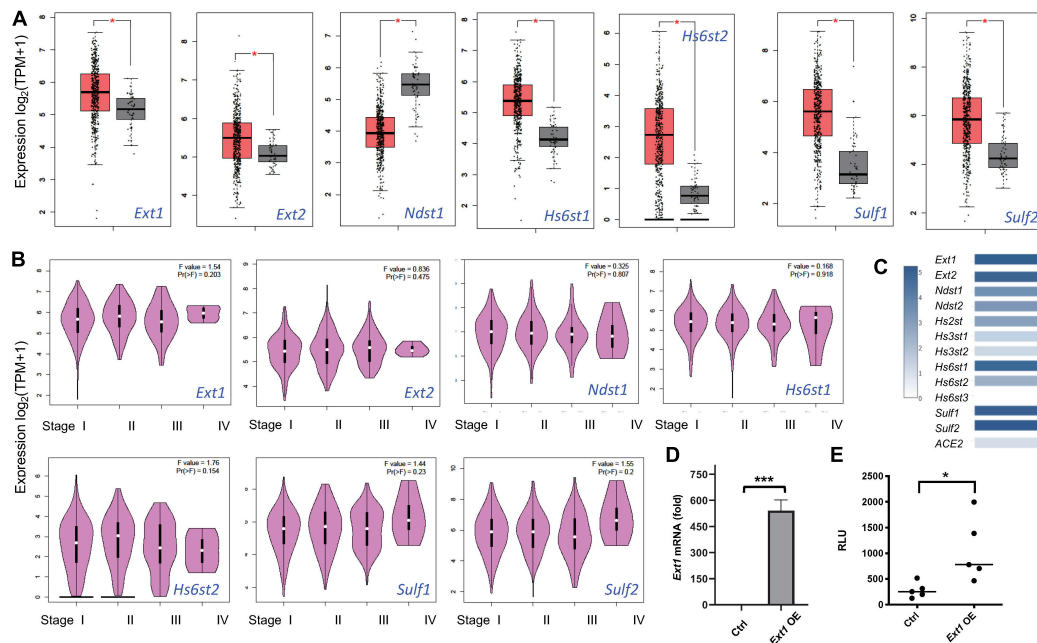
We examined SV2-S binding to serial chemically modified heparins and *O*-sulfation deficient HS mutant MLEC lines to determine if specific modifications are required for HS to mediate cell surface SV2-S binding. We observed both NS and 6S, compared to 2S, are more important for heparin to bind His-tagged SV2-S and the overall sulfation, instead of the individual sulfation modification type, plays the major role in determining the binding affinity of heparin to SV2-S. In agreement with 2DS-heparin study observation, *Hs2st*<sup>-/-</sup> MLECs showed normal cell surface SV2-S binding although *Hs2st*<sup>-/-</sup> MLEC HS completely lacks 2S and has a slight increase of overall sulfation, indicating that 2S is not required for HS to bind SV2-S on cell surface.

The consequences of manipulating 6S in HS-mediated SV2-S binding on MLEC surface appears complex. Although 6DS-heparin displayed a reduced binding affinity to SV2-S, the *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> MLECs showed a normal cell surface SV2-S binding. The *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> MLEC HS does not have 6S with compensatory increases of NS and 2S which result in an unchanged overall sulfation (Qiu et al., 2018), the normal SV2-S binding on *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> MLEC cell surface led to our conclusion that 6S is not essential for HS to bind SV2-S and overall sulfation is more important. A recent study

reported by Clausen et al. (2020) showing the binding of SV2-S on *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> Hep3B cells is significantly reduced. This reduced cell surface SV2-S binding is most likely due to decreased overall sulfation modification since the *Hs6st1*<sup>-/-</sup>;2<sup>-/-</sup> Hep3B HS has reduced 6S and a significantly reduced overall sulfation (Anower et al., 2019).

the *Hs6st1*<sup>-/-</sup> and *Hs6st2*<sup>-/-</sup> MLECs showed increased cell surface SV2-S binding. Since the *Hs6st1* deletion reduces 6S but does not alter NS, 2S and overall sulfation, and the *Hs6st2* deletion does not change NS, 2S, 6S and overall sulfation (Qiu et al., 2018), we therefore concluded that the increased cell surface SV2-S binding is due to the *Hs6st* deletion generates more SV2-S binding site and/or higher binding affinity, and fine structure contributes to HS to bind SV2-S. This notion is alternatively supported by study of *Sulf1*<sup>-/-</sup>;2<sup>-/-</sup> MLECs which showed reduced cell surface SV2-S binding. The *Sulf1*<sup>-/-</sup>;2<sup>-/-</sup> MLEC HS possesses increased 6S with compensatory decreases of NS and 2S and no change of overall sulfation (Qiu et al., 2018), showing the decreased SV2-S binding on the *Sulf1*<sup>-/-</sup>;2<sup>-/-</sup> MLECs is not related to overall sulfation but due to decreasing SV2-S binding site and/or lowering binding affinity. Our study of *Hs6st*- and *Sulf*-deficient MLECs concluded that fine structure contributes also to HS binding of SV2-S.





**FIGURE 6 |** Heparan sulfate gene expression is dysregulated in lung squamous cell carcinoma and replicating the upregulated *Ext1* expression enhances SARS-CoV2 pseudovirus infection. **(A)** Dysregulated HS gene expression in lung squamous cell carcinoma (LUSC). Using Gene Expression Profiling Interactive Analysis 2 (GEPIA2) software, the data deposited in The Cancer Genome Atlas (TCGA) including 486 LUSC patients with 50 normal controls were analyzed. The orange box indicates the tumor samples while the gray one represents the normal controls. The major HS biosynthetic genes, such as *Ext1* and *Ext2* are upregulated in LUSC. \* $p < 0.01$ . **(B)** HS gene expression did not differ among different pathological LUSC stages. **(C)** The relative expression of *ACE2* and HS genes in the LUSC. **(D,E)** Overexpression (OE) of *Ext1* increased SARS-CoV2 pseudovirus infection. Forty-eight hours after infection with a human *Ext1* expression lentiviral vector, the 293T-ACE2 showed increased *Ext1* mRNA expression **(D)** and were inoculated with luciferase-expressing SARS-CoV2 pseudovirus and examined for luciferase activity 48 h after the virus inoculation **(E)**. VSV pseudotyped virus was used as a control to make sure each well has comparable cell numbers (data are not shown). The experiments were repeated at least 3 times and an unpaired *t*-test was performed for two-group comparison. Data are presented as mean  $\pm$  SD. \* $p < 0.05$ , \*\*\* $p < 0.0001$ .

We examined *Hs3st1*<sup>-/-</sup>, *Hs3st4*<sup>-/-</sup> and *Hs3st1*<sup>-/-</sup>;*4*<sup>-/-</sup> MLEC cell lines, which have normal NS, 2S, and 6S with reduced AT binding, indicating 3S is reduced or diminished to determine if 3S contributes to HS-mediated SV2-S binding on cell surface (Qiu et al., 2018). The *Hs3st4*<sup>-/-</sup> MLEC HS, not *Hs3st1*<sup>-/-</sup> and *Hs3st1*<sup>-/-</sup>;*4*<sup>-/-</sup> MLEC HS, has slightly reduced overall sulfation, supporting the notion that 3S is a type of rare modification within HS (Thacker et al., 2014). The *Hs3st4*<sup>-/-</sup> MLECs, *Hs3st1*<sup>-/-</sup> and *Hs3st1*<sup>-/-</sup>;*4*<sup>-/-</sup> MLECs all showed normal levels of cell surface SV2-S binding, indicating 3S is not essential for HS to bind SV2-S and supporting that overall sulfation plays a major role in mediating this interaction.

Recently, Tiwari et al. (2020b) reported that overexpression of *Hs3st3b* preferentially increased SV2-S-mediated cell-to-cell fusion. In mammals, the seven isozymes of *Hs3st* are divided into two subgroups based on the homology of the sulfotransferase domain: the “AT-type” which includes *Hs3st1* and *Hs3st5*, and the “gD-type” that contains *Hs3st2*, *Hs3st3a*, *Hs3st3b*, *Hs3st4*, and *Hs3st6* (Lawrence et al., 2007; Liu and Pedersen, 2007). The “AT-type” and “gD-type” *Hs3sts* generate differently structured 3S-modified ligand binding sites for antithrombin and glycoprotein gD of Herpes simplex virus-1, respectively (Lawrence et al., 2007). We recently reported that deletion of “AT-type” *Hs3st1*, “gD -type” *Hs3st4*, or both reduced MLEC surface AT binding

(Qiu et al., 2018); therefore, we postulate the increased SV2-S-mediated cell-to-cell fusion by *Hs3st3b* overexpression is most likely due to an increase of HS overall sulfation.

As shown in this study, and reports from our own and other groups, heparin potentially inhibited SV2-S-mediated pseudovirus entry into host cells (Clausen et al., 2020; Tandon et al., 2020). We and others also reported that low molecular weight heparin and chemically-split heparin have strong inhibitory effects too (Clausen et al., 2020; Tandon et al., 2020). Extending these reports, we examined surfen, a small molecule HS antagonist (Schuksz et al., 2008), and several chemically modified, non-anticoagulant heparins, which showed anti-inflammatory effects in our previous studies (Wang et al., 2002; Zhang et al., 2012). We observed that surfen and chemically-modified non-anticoagulant heparins including NDS-HP, 2DS-HP, 6DS-HP, and OS-HP all inhibited SV2-S-mediated pseudovirus entry into host cells. Surfen has been used safely in patients for a long time (Schuksz et al., 2008) and heparin is a commonly used anticoagulant in patients; therefore, surfen and the derivatives of heparin, the chemically-modified heparins, may be applied safely to prevent or treat COVID19. Importantly, the chemically-modified heparins do not have anticoagulant activity, avoiding the potential bleeding side effect of heparin treatment.

SARS-CoV-2 variants with SV2-S G614 mutation now predominate globally (Walls et al., 2020b; Zhou et al., 2020). A recent study observed that the SV2-S G614 pseudotyped virus enters ACE2-expression cells more efficiently than the SV2-S D614 wildtype pseudotyped virus (Zhang et al., 2020a). The increased infectivity of the SV2-S G614 pseudotyped virus correlates with less S1-domain shedding and higher SV2-S incorporation into the virion. The G614 mutation does not affect SV2-S binding to ACE2 or neutralization sensitivity of the pseudovirus, suggesting that G614 mutation may assemble more functional SV2-S into the virion to increase infectivity (Zhang et al., 2020a). In the current study, we found that the G614 mutation strengthens SV2-S binding to heparin, suggesting that the G614 mutation may enhance SV2-S interaction with host cell surface HS thereby increase infectivity of the mutant virus. This observation reveals a potential novel mechanism underlying the higher infectivity of the SV2-S G614 variant. More importantly, the entry of the SV2-S G614 pseudotyped virus into host cells depends on host cell surface HS and can be inhibited by heparin, suggesting heparin, surfen, and chemically-modified heparins may be effective to prevent or treat SV2-S G614 variant infection too.

The severity of COVID-19 and the course of the infection is heterogeneous and appears to be more severe in the elderly and individuals with underlying comorbidities. COVID-19 does not appear to significantly impact children, a pattern atypical for most viral respiratory diseases. Several retrospective cohort studies have emerged that cancer patients are at increased risk of COVID-19 severity and fatality due to underlying malignancy, treatment-related immunosuppression, or increased comorbidities (Liang et al., 2020; Mehta et al., 2020; Poortmans et al., 2020; Rogado et al., 2020; Yu et al., 2020; Zhang et al., 2020b). Among the major cancer types analyzed, lung cancer patients show the highest fatality rates for COVID-19. A recent comprehensive clinical study identified a total of 218 COVID-19 positive patients with a malignant diagnosis (Mehta et al., 2020). In this report, a total of 61 (28%) cancer patients died from COVID-19 emerging lung cancer patients being the most vulnerable population, with a fatality rate of 55%. Currently, the mechanisms underlying the higher vulnerability of cancer patients to SARS-CoV-2 infection remain largely unknown (Liang et al., 2020; Mehta et al., 2020; Poortmans et al., 2020; Rogado et al., 2020; Yu et al., 2020; Zhang et al., 2020b). In this study, we uncovered that the expression of several HS biosynthetic genes, including *Ext1*, is upregulated in lung cancer. Replicating the upregulated *Ext1* expression increased infection of the host cells by SARS-CoV-2 pseudovirus, and the increased infection can be blocked by HSase and heparin, suggesting heparin and its derivatives, as well as surfen, may be effective to prevent or block SARS-CoV-2 infection in cancer patients.

## CONCLUSION

Our studies demonstrated that host cell surface SV2-S binding and SV2-S-mediated SARS-CoV-2 pseudovirus infection

depends on host cell surface HS and can be blocked by heparin lyase, HS antagonist surfen, heparin, and non-anticoagulant heparin derivatives. HS binding to SV2-S and the inhibitory effects of heparin and heparin derivatives are mainly determined by their overall sulfation instead of a specific type of sulfation. This conclusion is supported by mutant cell studies, which determined that 2S, 6S, and 3S are not essential for HS to bind SV2-S on cell surface. Our HS mutant cell study also emerged that the binding of HS to SV2-S is contributed by its fine structure too, although this appears minor comparing to its overall sulfation. Our studies also determined that the G614 mutation strengthens binding of SV2-S to heparin, suggesting the increased SV2-S-HS interaction may be one of the mechanisms underlying the higher infectivity of the SARS-CoV-2 G614 variant. Furthermore, we showed HS expression is upregulated in lung cancer and higher HS expression increased host cell infection of SARS-CoV-2, delineating a mechanism accounting for the high vulnerability of lung cancer to SARS-CoV-2 infection. Lastly, we demonstrated that heparin lyase and heparin effectively blocked host cell infection by SARS-CoV-2 G614 variant pseudovirus and increased infection caused by upregulated HS expression, suggesting that heparin, heparin derivatives, heparin lyase, and surfen may be applied to prevent or treat SARS-CoV-2 G614 variant infection and infection of cancer patients.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA-LUSC data was downloaded from NIH CDC Data Portal: <https://portal.gdc.cancer.gov/projects/TCGA-LUSC>, accession number: dbGaP Study Accession phs000178.

## AUTHOR CONTRIBUTIONS

JY, WJ, HY, JF, XS, and HQ conducted the experiments, acquisition, and analysis of the results. JY, FZ, MT, PA, RL, and LW designed the experiments and wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.

## FUNDING

This work was supported by NIH R01 HL093339 (LW), U01 CA225784 (LW), and R56 AG062344 (LW and RL).

## ACKNOWLEDGMENTS

The authors acknowledge Dr. Hyeryun Choe for providing SV2-S G614 mutant plasmid, Dr. David Kang for the SH-SY5Y cells, Dr. Narasaiah Kolliputi for HMVEC and H441 cells, and Dr. Xingmin Sun for THP-1 cells.

## REFERENCES

- Anower, E. K. F., Singh, G., Deng, Y., Gordts, P., and Esko, J. D. (2019). Triglyceride-rich lipoprotein binding and uptake by heparan sulfate proteoglycan receptors in a CRISPR/Cas9 library of Hep3B mutants. *Glycobiology* 29, 582–592. doi: 10.1093/glycob/cwz037
- Aquino, R. S., and Park, P. W. (2016). Glycosaminoglycans and infection. *Front. Biosci.* 21:1260–1277. doi: 10.2741/4455
- Bobardt, M. D., Salmon, P., Wang, L., Esko, J. D., Gabuzda, D., Fiala, M., et al. (2004). Contribution of proteoglycans to human immunodeficiency virus type 1 brain invasion. *J. Virol.* 78, 6567–6584. doi: 10.1128/jvi.78.12.6567-6584.2004
- Cagno, V., Tseligka, E. D., Jones, S. T., and Tapparel, C. (2019). Heparan sulfate proteoglycans and viral attachment: true receptors or adaptation bias? *Viruses* 11:596. doi: 10.3390/v11070596
- Cai, B. H., Wu, P. H., Chou, C. K., Huang, H. C., Chao, C. C., Chung, H. Y., et al. (2019). Synergistic activation of the NEU4 promoter by p73 and AP2 in colon cancer cells. *Sci. Rep.* 9:950.
- Cancer Genome Atlas Research, N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764
- Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* 183, 1043–1057.e15.
- Condac, E., Strachan, H., Gutierrez-Sanchez, G., Brainard, B., Giese, C., Heiss, C., et al. (2012). The C-terminal fragment of axon guidance molecule Slit3 binds heparin and neutralizes heparin's anticoagulant activity. *Glycobiology* 22, 1183–1192. doi: 10.1093/glycob/cws087
- de Haan, C. A., Haijema, B. J., Schellen, P., Wichgers Schreur, P., Te Lintelo, E., Vennema, H., et al. (2008). Cleavage of group 1 coronavirus spike proteins: how furin cleavage is traded off against heparan sulfate binding upon cell culture adaptation. *J. Virol.* 82, 6078–6083. doi: 10.1128/jvi.00074-08
- Efendizade, A., Dmytriw, A. A., Hewitt, K., and Davies, G. A. (2020). Unfractionated heparin in SARS-CoV-2 pneumonia: ischemic stroke case report. *Front. Neurol.* 11:573356. doi: 10.3389/fneur.2020.573356
- Fu, J., Zhou, B., Zhang, L., Balaji, K. S., Wei, C., Liu, X., et al. (2020). Expressions and significances of the angiotensin-converting enzyme 2 gene, the receptor of SARS-CoV-2 for COVID-19. *Mol. Biol. Rep.* 47, 4383–4392. doi: 10.1007/s11033-020-05478-4
- Graham, R. L., and Baric, R. S. (2020). SARS-CoV-2: combating coronavirus emergence. *Immunity* 52, 734–736. doi: 10.1016/j.immuni.2020.04.016
- Hippensteel, J. A., LaRiviere, W. B., Colbert, J. F., Langouet-Astrie, C. J., and Schmidt, E. P. (2020). Heparin as a therapy for COVID-19: current evidence and future possibilities. *Am. J. Physiol. Lung Cell Mol. Physiol.* 319, L211–L217.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8.
- Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., et al. (2020). Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. *Antiviral Res.* 181:104873. doi: 10.1016/j.antiviral.2020.104873
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827.e19.
- Kraushaar, D. C., Rai, S., Condac, E., Nairn, A., Zhang, S., Yamaguchi, Y., et al. (2012). Heparan sulfate facilitates FGF and BMP signaling to drive mesoderm differentiation of mouse embryonic stem cells. *J. Biol. Chem.* 287, 22691–22700. doi: 10.1074/jbc.m112.368241
- Kraushaar, D. C., Yamaguchi, Y., and Wang, L. (2010). Heparan sulfate is required for embryonic stem cells to exit from self-renewal. *J. Biol. Chem.* 285, 5907–5916. doi: 10.1074/jbc.m109.066837
- Lawrence, R., Yabe, T., Hajmohammadi, S., Rhodes, J., McNeely, M., Liu, J., et al. (2007). The principal neuronal gD-type 3-O-sulfotransferases and their products in central and peripheral nervous system tissues. *Matrix Biol.* 26, 442–455. doi: 10.1016/j.matbio.2007.03.002
- Liang, J., Liu, Z., Zou, Z., Tang, Y., Zhou, C., Yang, J., et al. (2018). The correlation between the immune and epithelial-mesenchymal transition signatures suggests potential therapeutic targets and prognosis prediction approaches in kidney cancer. *Sci. Rep.* 8:6570.
- Liang, W., Guan, W., Chen, R., Wang, W., Li, J., Xu, K., et al. (2020). Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *Lancet Oncol.* 21, 335–337. doi: 10.1016/s1470-2045(20)30096-6
- Liu, J., and Pedersen, L. C. (2007). Anticoagulant heparan sulfate: structural specificity and biosynthesis. *Appl. Microbiol. Biotechnol.* 74, 263–272. doi: 10.1007/s00253-006-0722-x
- Liu, L., Chopra, P., Li, X., Wolfert, M. A., Tompkins, S. M., and Boons, G. J. (2020). SARS-CoV-2 spike protein binds heparan sulfate in a length- and sequence-dependent manner. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.10.087288
- Lu, J., and Sun, P. D. (2020). High affinity binding of SARS-CoV-2 spike protein enhances ACE2 carboxypeptidase activity. *J. Biol. Chem.* 295, 18579–18588. doi: 10.1074/jbc.ra120.015303
- Martin, J., Klapsa, D., Wilton, T., Zambon, M., Bentley, E., Bujaki, E., et al. (2020). Tracking SARS-CoV-2 in Sewage: evidence of changes in virus variant predominance during COVID-19 pandemic. *Viruses* 12:1144. doi: 10.3390/v12101144
- Mehta, V., Goel, S., Kabarriti, R., Cole, D., Goldfinger, M., Acuna-Villaorduna, A., et al. (2020). Case fatality rate of cancer patients with COVID-19 in a New York hospital system. *Cancer Discov.* 10, 935–941.
- Menezes-Rodrigues, F. S., Padrao Tavares, J. G., Pires de Oliveira, M., Guzella de Carvalho, R., Ruggero Errante, P., Omar Taha, M., et al. (2020). Anticoagulant and antiarrhythmic effects of heparin in the treatment of COVID-19 patients. *J. Thromb. Haemost.* 18, 2073–2075. doi: 10.1111/jth.14902
- Milewska, A., Zarebski, M., Nowak, P., Stozek, K., Potempa, J., and Pyrc, K. (2014). Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. *J. Virol.* 88, 13221–13230. doi: 10.1128/jvi.02078-14
- Poortmans, P. M., Guarneri, V., and Cardoso, M. J. (2020). Cancer and COVID-19: what do we really know? *Lancet* 395, 1884–1885. doi: 10.1016/s0140-6736(20)31240-x
- Qiu, H., Jiang, J. L., Liu, M., Huang, X., Ding, S. J., and Wang, L. (2013). Quantitative phosphoproteomics analysis reveals broad regulatory role of heparan sulfate on endothelial signaling. *Mol. Cell Proteomics* 12, 2160–2173. doi: 10.1074/mcp.m112.026609
- Qiu, H., Shi, S., Yue, J., Xin, M., Nairn, A. V., Lin, L., et al. (2018). A mutant-cell library for systematic analysis of heparan sulfate structure-function relationships. *Nat. Methods* 15, 889–899. doi: 10.1038/s41592-018-0189-6
- Rai, S., Alsaidan, O. A., Yang, H., Cai, H., and Wang, L. (2020). Heparan sulfate inhibits transforming growth factor beta signaling and functions in cis and in trans to regulate prostate stem/progenitor cell activities. *Glycobiology* 30, 381–395. doi: 10.1093/glycob/cwz103
- Rogado, J., Obispo, B., Pangua, C., Serrano-Montero, G., Martin Marino, A., Perez-Perez, M., et al. (2020). Covid-19 transmission, outcome and associated risk factors in cancer patients at the first month of the pandemic in a Spanish hospital in Madrid. *Clin. Transl. Oncol.* 22, 2364–2368. doi: 10.1007/s12094-020-02381-z
- Rusnati, M., Vicenzi, E., Donalisio, M., Oreste, P., Landolfo, S., and Lembo, D. (2009). Sulfated K5 *Escherichia coli* polysaccharide derivatives: a novel class of candidate antiviral microbicides. *Pharmacol. Ther.* 123, 310–322. doi: 10.1016/j.pharmthera.2009.05.001
- Schukasz, M., Fuster, M. M., Brown, J. R., Crawford, B. E., Ditto, D. P., Lawrence, R., et al. (2008). Surfen, a small molecule antagonist of heparan sulfate. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13075–13080. doi: 10.1073/pnas.0805862105
- Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* 30, 981–988. doi: 10.1093/glycob/cwaa042
- Su, S., and Jiang, S. (2020). A suspicious role of interferon in the pathogenesis of SARS-CoV-2 by enhancing expression of ACE2. *Signal Transduct. Target Ther.* 5:71.
- Tan, H. W., Xu, Y. M., and Lau, A. T. Y. (2020). Angiotensin-converting enzyme 2: the old door for new severe acute respiratory syndrome coronavirus 2 infection. *Rev. Med. Virol.* 30:e2122.
- Tandon, R., Sharp, J. S., Zhang, F., Pomin, V. H., Ashpole, N. M., Mitra, D., et al. (2020). Effective inhibition of SARS-CoV-2 entry by heparin and enoxaparin derivatives. *J. Virol.* 95, e1987–20.
- Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 47, W556–W560.

- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102.
- Thacker, B. E., Xu, D., Lawrence, R., and Esko, J. D. (2014). Heparan sulfate 3-O-sulfation: a rare modification in search of a function. *Matrix Biol.* 35, 60–72. doi: 10.1016/j.matbio.2013.12.001
- Tiwari, V., Beer, J. C., Sankaranarayanan, N. V., Swanson-Mungerson, M., and Desai, U. R. (2020a). Discovering small-molecule therapeutics against SARS-CoV-2. *Drug Discov. Today* 25, 1535–1544. doi: 10.1016/j.drudis.2020.06.017
- Tiwari, V., Tandon, R., Sankaranarayanan, N. V., Beer, J. C., Kohlmeier, E. K., Swanson-Mungerson, M., et al. (2020b). Preferential recognition and antagonism of SARS-CoV-2 spike glycoprotein binding to 3- O -sulfated heparan sulfate. *bioRxiv* [Preprint]. Available online at: <https://www.biorxiv.org/content/10.1101/2020.10.08.331751v1> (accessed October 08, 2020).
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veelsler, D. (2020a). Structure, function, and antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181, 281–292.e6.
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veelsler, D. (2020b). Structure, function, and antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 183:1735. doi: 10.1016/j.cell.2020.11.032
- Wang, D., Baudys, J., Bundy, J. L., Solano, M., Keppel, T., and Barr, J. R. (2020). Comprehensive analysis of the glycan complement of SARS-CoV-2 Spike proteins using signature ions-triggered electron-transfer/higher-energy collisional dissociation (ETHD) mass spectrometry. *Anal. Chem.* 92, 14730–14739. doi: 10.1021/acs.analchem.0c03301
- Wang, L., Brown, J. R., Varki, A., and Esko, J. D. (2002). Heparin's anti-inflammatory effects require glucosamine 6-O-sulfation and are mediated by blockade of L- and P-selectins. *J. Clin. Invest.* 110, 127–136. doi: 10.1172/jci0214996
- Wang, L., Fuster, M., Sriramara, P., and Esko, J. D. (2005). Endothelial heparan sulfate deficiency impairs L-selectin- and chemokine-mediated neutrophil trafficking during inflammatory responses. *Nat. Immunol.* 6, 902–910. doi: 10.1038/ni1233
- Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 369, 330–333. doi: 10.1126/science.abb9983
- Wijelath, E., Namekata, M., Murray, J., Furuyashiki, M., Zhang, S., Coan, D., et al. (2010). Multiple mechanisms for exogenous heparin modulation of vascular endothelial growth factor activity. *J. Cell Biochem.* 111, 461–468. doi: 10.1002/jcb.22727
- Xu, W., Wang, M., Yu, D., and Zhang, X. (2020). Variations in SARS-CoV-2 spike protein cell epitopes and glycosylation profiles during global transmission course of COVID-19. *Front. Immunol.* 11:565278. doi: 10.3389/fimmu.2020.565278
- Yu, J., Ouyang, W., Chua, M. L. K., and Xie, C. (2020). SARS-CoV-2 transmission in patients with cancer at a tertiary care hospital in Wuhan, China. *JAMA Oncol.* 6, 1108–1110. doi: 10.1001/jamaoncol.2020.0980
- Zhang, B., Xiao, W., Qiu, H., Zhang, F., Moniz, H. A., Jaworski, A., et al. (2014). Heparan sulfate deficiency disrupts developmental angiogenesis and causes congenital diaphragmatic hernia. *J. Clin. Invest.* 124, 209–221. doi: 10.1172/jci71090
- Zhang, F., Moniz, H. A., Walcott, B., Moremen, K. W., Linhardt, R. J., and Wang, L. (2013). Characterization of the interaction between Robo1 and heparin and other glycosaminoglycans. *Biochimie* 95, 2345–2353. doi: 10.1016/j.biochi.2013.08.018
- Zhang, F., Moniz, H. A., Walcott, B., Moremen, K. W., Wang, L., and Linhardt, R. J. (2014). Probing the impact of GFP tagging on Robo1-heparin interaction. *Glycoconj J.* 31, 299–307. doi: 10.1007/s10719-014-9522-1
- Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., et al. (2020a). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* 11:6013.
- Zhang, L., Zhu, F., Xie, L., Wang, C., Wang, J., Chen, R., et al. (2020b). Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. *Ann. Oncol.* 31, 894–901. doi: 10.1016/j.annonc.2020.03.296
- Zhang, Q., Chen, C. Z., Swaroop, M., Xu, M., Wang, L., Lee, J., et al. (2020c). Heparan sulfate assists SARS-CoV-2 in cell entry and can be targeted by approved drugs in vitro. *Cell Discov.* 6:80.
- Zhang, S., Condac, E., Qiu, H., Jiang, J., Gutierrez-Sanchez, G., Bergmann, C., et al. (2012). Heparin-induced leukocytosis requires 6-O-sulfation and is caused by blockade of selectin- and CXCL12 protein-mediated leukocyte trafficking in mice. *J. Biol. Chem.* 287, 5542–5553. doi: 10.1074/jbc.m111.314716
- Zhao, J., Zhu, Y., Song, X., Xiao, Y., Su, G., Liu, X., et al. (2020). 3-O-Sulfation of heparan sulfate enhances tau interaction and cellular uptake. *Angew. Chem. Int. Ed. Engl.* 59, 1818–1827. doi: 10.1002/anie.201913029
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zong, C., Huang, R., Condac, E., Chiu, Y., Xiao, W., Li, X., et al. (2016). Integrated approach to identify heparan sulfate ligand requirements of Robo1. *J. Am. Chem. Soc.* 138, 13059–13067. doi: 10.1021/jacs.6b08161
- Zong, C., Venot, A., Li, X., Lu, W., Xiao, W., Wilkes, J. L., et al. (2017). Heparan sulfate microarray reveals that heparan sulfate-protein binding exhibits different ligand requirements. *J. Am. Chem. Soc.* 139, 9534–9543. doi: 10.1021/jacs.7b01399

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yue, Jin, Yang, Faulkner, Song, Qiu, Teng, Azadi, Zhang, Linhardt and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership