UNIVERSITY OF CALIFORNIA SAN DIEGO

Proteomic-Based Screening of the Endothelial Heparan-Sulfate Interactome

A dissertation submitted in partial satisfaction of the

requirements for the degree in Doctor of Philosophy

in

Biomedical Sciences

by

Daniel Rogelio Sandoval

Committee in charge:

Professor Jeffrey Esko, Chair
Professor David Cheresh
Professor Richard Gallo
Professor Tracy Handel
Professor Nathan Lewis

2019

The Dissertation of Daniel Rogelio Sandoval is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

These studies are dedicated to my parents Margot and Tony Sandoval who have sacrificed much and provided continuous support throughout my life and studies.

To my brothers, Adrian and Marco Sandoval, who have constantly encouraged and challenged me.

EPIGRAPH

"The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can,
Pursuing it with eager feet,
Until it joins some larger way
Where many paths and errands meet.
And whither then? I cannot say"

J.R.R Tolkien, The Fellowship of the Ring

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Jeffrey D. Esko, Ph.D. for his support as chair of my committee. In the course of my studies, he has provided me great academic and cultural opportunities with the patience to not only support me as a scientist but also as a person.

I would like to acknowledge Philip L.S.M Gordts, Ph.D., Chrissa Dwyer, Ph.D. and Anne Pham Ph.D. whose care and guidance greatly streamlined my dissertation

I would like to acknowledge members of the Esko lab both past and present for the myriad of scientific discussions.

Chapter 1, in full is being prepared for submission as a publication. Toledo A.G., Sandoval, D.R., Sorrentino, J.T., Kellman B.P., Lewis, N.E., Esko, J.D. I was the secondary investigator for this paper.

Chapter 2, in full has been submitted for publication in the Journal for the American Chemical Socitety. Sandoval, D.R., Toledo, A.G., Painter C.D., Tota, E., Sheikh, O., West, A.M.V., Wells, L., Bicknell, R., Corbett K.D., Xu, D., Esko, J.D. I am the primary investigator and author of this paper.

Chapter 4, in part is being prepared for submission. Toledo, A.G., Painter, C.D., Esko J.D. I am the primary investigator and author of this paper.

VITA

2012 Bachelor of Science, University of California Irvine

2019 Doctor of Philosophy, University of California San Diego

Fields of Study

Major Field: Biological Sciences

Honors and Awards

2014-2018 Program in excellence in Glycoscience (PEG) Training Grant
2016 Carl Storm Underrepresented Minority Fellowship (CSURM)
2016 Graduate Student Association (GSA) Travel Grant
2015 9th and 10th Pan-Pacific Connective Tissue Societies Symposium Travel Award
2013 HHMI Med-Into Grad
2013 San Diego Fellowship
2013 Initiative for Maximizing Student Development (IMSD)
2010-2012 Minority Access to Research Careers (MARC) NIH Grant
2011 Leadership Alliance Summer Research - Early Identification Program (SR-EIP) Summer Research opportunities at Harvard (SROH)
2011 Excellence in Research in the School of Biological sciences, UC Irvine
2010 Minority Health and Health Disparities International Research Training (MHIRT)
2010 Undergraduate Research and Mentoring in the Biological Sciences (URM) NSF grant
2009 Minority Biomedical Research Support Program (MBRS)

ABSTRACT OF THE DISSERTATION

Proteomic-Based Screening of the Endothelial Heparan-Sulfate Interactome

by

Daniel Rogelio Sandoval

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2019

Professor Jeffrey D. Esko, Chair

All cell types express heparan sulfate proteoglycans (HSPGs), either embedded into the cell membrane or released into the extracellular matrix. To identify novel membrane bound heparan sulfate binding proteins (HSBPs), we utilized limited proteolysis to liberate ectodomains of cell surface proteins expressed on human umbilical vein endothelial cells. Chromatography over heparin-Sepharose and mass spectrometry yielded several known HSBPs and identified the heparin binding domain. Novel endothelial HSPBs were identified including PTPRB, CLEC14A, CLEC2B, CD93, and GDF15. We mapped the heparin-binding domain of CLEC14A by mutagenesis and showed that the binding site resides in the C-type lectin domain. Recombinant CLEC14A ectodomain bound with high affinity to heparin oligosaccharides of ≥dp6. Binding occurred in 1:1 stoichiometry and led to increased thermal stability of CLEC14A. Overexpression of membrane bound wild-type CLEC14A or the ectodomain had no effect on *in vitro* angiogenic sprouting, whereas the CLEC14A heparin binding deficient mutant protein inhibited sprouting. To determine how heparin binding mediates CLEC14A function, we then developed a proteomic workflow utilizing chemical cross-linking to identify CLEC14A endothelial protein binding partners and to test if heparin can alter CLEC14A protein-protein interactions. Mass spectrometry identified over 40 CLEC14A binding partners including MMRN2, ITIH5, NRP1, TXLNA, and EFEMP. A model is proposed in which heparan sulfate modulates CLEC14A-protein interactions to control angiogenesis.

CHAPTER 1: A SYSTEMS VIEW OF THE HEPARAN SULFATE INTERACTOME

Abstract

Heparan sulfate proteoglycans are integral components of the glycocalyx, a carbohydrate-rich layer surrounding the plasma membrane of all metazoan cells. The heparan sulfate proteoglycans consist of a small family of proteins decorated with one or more covalently attached heparan sulfate glycosaminoglycan chains. These chains have intricate patterns of sulfated sugars and uronic acid epimers, which dictate their ability to engage a large repertoire of proteins, including extracellular matrix proteins, growth factors and morphogens, cytokines and chemokines, apolipoproteins and lipases, adhesion and growth factor receptors, and components of the complement and coagulation system. This review highlights recent progress in the characterization of the so-called “heparan sulfate-interactome”, with a major focus on systems-wide biochemical strategies as tools for discovery and characterization of the interactome.

1.Introduction

Proteoglycans (PGs) are glycoproteins ubiquitously present at the cell surface and in the extracellular matrix of all vertebrate and invertebrate cells and tissues. The PGs encompass a broad family of proteins decorated with structurally related polysaccharides, collectively denoted as glycosaminoglycans **(Figure 1)**. Notably, only a small number of PGs display covalently linked heparan sulfate (HS) glycosaminoglycan chains. HS is a linear polysaccharide generated by the copolymerization of alternating N-acetyl-D-glucosamine (GlcNAc) and D-glucuronic acid (GlcA). As the chains assemble, a subset of GlcNAc residues undergo N-deacetylation and N-sulfation, followed by epimerization of GlcA units into L-iduronic acid (IdoA), as well as O-sulfation at various positions. Discrete patterns of N-acetylation, N- and O-sulfation, and uronic acid epimerization arise from a complex assembly process, orchestrated by multiple

glycosyltransferases, sulfotransferases and an epimerase. Details of this process can be found elsewhere (1), but of relevance is the observation that patterning of the chains occurs in a cell type specific manner can varies during development and in response to various physiological challenges. Interestingly, the assembly process is not template-driven, but occurs in a heritable manner in different cell types, giving rise to tissue-specific signatures.

This review focuses on the proteins that interact with the HS chains. We argue that the HS-interactome should be considered a distinct subproteome that can be targeted using high-throughput screening methodologies. This subproteome is unique in the sense of covering a broad spectrum of structurally unrelated proteins, with diverse molecular and cellular functions, but with HS-binding ability as a common trait. We also took a systematic approach to look at all the HS-binding proteins reported to date by applying a network biology approach. Our analysis shows that most HS-interacting proteins can be grouped into a limited number of protein-protein interaction networks with remarkable functional commonalities. This is suggestive of HS having a role in the orchestration of coordinated molecular events such as cellular signal transduction, immune responses, etc. In general, this review is an attempt to unify a series of concepts and observations regarding the nature of the HS-interactome, that are currently scattered in the literature. Together, these studies articulate a coherent subproteome that has not been fully explored and that has all the potential to benefit from methodological and theoretical advances in related proteomics areas. In exchange, the opening of a molecular window into the HS-interactome may provide us with novel targets to address complex diseases. Many of the principles described here apply to other classes of proteoglycans, such as those that contain chondroitin/dermatan sulfate or keratan sulfate chains, but due to space limitations these will not be further discussed. Excellent articles covering these subjects are available elsewhere (2,3).

1.1.Heparan Sulfate Binding Proteins: Basic Definitions and Principles of Interaction

The arrangement of the sulfated residues and uronic acids in HS create binding sites for proteins, which collectively are called “heparan sulfate-binding proteins” (HSBPs). HSBPs

encompass a heterogenous group of proteins with the common feature of affinity for HS. Because of the high charge density imparted by the sulfate groups and uronic acids, many intracellular proteins such as histones and transcription factors can also interact with HS. However, throughout this review we reserve the term HSBP for proteins at the plasma membrane or in the extracellular space, where HSPGs typically reside and effect their biological activities. There is evidence of HS in the nucleus but its biological significance remains to be established (4,5).

HSBPs are involved in many fundamental processes, including native immunity, coagulation, lipid metabolism, extracellular matrix assembly, growth factor signaling and cell adhesion. Thus, the HSBP repertoire includes structural proteins, cellular receptors, growth factors, enzymes and enzyme inhibitors. In contrast to lectins, which can be sorted into evolutionarily related groups by their folds and carbohydrate binding domains, most HSBPs are structurally unrelated. Their capacity to bind HS is thus believed to have arisen through convergent evolution. Importantly, they differ in their affinity and specificity; some are highly specific for HS whereas others appear to be promiscuous and can equally engage related glycosaminoglycan classes such as chondroitin/dermatan, and possibly keratan sulfates.

A number of general principles have emerged through structural and biochemical studies of multiple HSBPs, illustrating how HS-protein interactions can lead to a variety of molecular outcomes (**Figure 1.2**). For example, binding to HS can tether proteins, encouraging their presentation near their site of synthesis. Tethering is here used as a generic term to describe the ability of HS chains to localize a protein, to restrict its diffusion, and/or increase its concentration and availability to other receptors. In addition to ligand confinement, tethering may result in significant increase in the protein half-life by offering protection from protease cleavage. These mechanisms are essential to facilitate the formation of morphogen and growth factor gradients, to regulate leukocyte homing and to direct the tissue tropism of several pathogens (6,7).

HS binding can trigger protein oligomerization. Examples include, chemokine oligomerization, dimerization of growth factors, and dimerization (and eventual oligomerization) of cellular receptors (8-11). Evidence for the profound consequences of these mechanisms was recently reported in the context of FGF1/FGFR2 signal transduction, where the coupling of mitogenic or metabolic signaling was found to be balanced by the ability to maintain FGF1/FGFR2 dimer stability in an HS-dependent manner (12). Interestingly, some ligands such as vascular endothelial growth factor A (VEGFA) can undergo alternative splicing, giving rise to isoforms with or without the heparin-binding domains, thus altering its capacity to dimerize and localize in tissues (13).

Finally, HS can also enhance protein-protein interactions. Due to the length, flexibility and variable composition of HS chains, they constitute molecular scaffolds to bring specific HSBPs in proximity to each other. One classic example is the interaction of heparin with antithrombin and thrombin. Binding of heparin to antithrombin is mediated by a pentasaccharide motif of specific sulfation/iduronic acid composition, whereas binding to thrombin is more promiscuous. The approximation of the two proteins, and the conformational change in antithrombin induced by binding to this pentasaccharide, results in 1000-fold enhancement in thrombin inactivation.

2. A Systems View of the HS/Heparin-Interactome

2.1.System-Wide Strategies to Map the HS/Heparin-Interactome

Since the advent of the “Omics-revolution”, system-wide approaches have been readily applied to determine the scope of the HS-interactome. One important insight coming out from these studies is that several hundreds of soluble and membrane proteins can interact with HS, and sometimes with remarkable selectivity. Additionally, many HS-binding proteins are coordinated within common biological pathways, suggesting that HS might act as a global regulator of these processes. Therefore, HS-protein interactions constitute promising

therapeutic targets for treating devastating diseases such as cancer, cardiovascular diseases and unwanted inflammatory responses.

Mass spectrometric-based proteomics strategies have been applied to study multiple subproteomes. The current sensitivity of modern MS instrumentation allows for detection and quantitation of molecular species in the low nanomolar range, even in complex samples such as body fluids, cell lysates and tissue homogenates. The resolution of these techniques circumvents the need for labor-intensive purification schemes while providing a global picture of the targeted proteome. The combination of affinity chromatography and MS-detection is also a powerful strategy to reveal the identity and temporal-spatial dynamics of specific subproteomes at molecular resolution.

Currently, the most common approach to identify novel HS-binding proteins entail some kind of affinity chromatography, followed by proteolytic digestion and MS-detection. Until recently, HS has been difficult to isolate in large quantities. Therefore, heparin has typically been used as a proxy to fractionate and identify HSBPs from different fluids and tissues. Heparin is a highly sulfated HS-subtype rich in iduronic acid, and mainly produced by connective tissue (mucosal) mast cells. Heparin chains are initially synthesized while covalently linked to the proteoglycan serglycin. These chains undergo limited cleavage and are stored within secretory granules, where they play a key role in the storage of proteases and biogenic monoamines (14,15). Heparin is also routinely used in the clinics as an anticoagulant agent due to its ability to bind antithrombin, greatly increasing its capacity to inactivate pro-coagulation enzymes such as thrombin and Factor Xa (16). As mentioned before, this interaction is mediated by a discrete pentasaccharide motif within heparin chains, which includes relatively rare 3-O-sulfated glucosamine residues. Heparin-affinity matrices are inexpensive, constituting a simple choice of affinity material for screening. However, it should be kept in mind that the heparin fine structure significantly differs from most cellular HS chains. Indeed, nearly 85% of the glucosamine residues of heparin are commonly N-sulfated, whereas as few as 20-45% of

the residues are modified in HS. Thus, HS consists of tracts of modified so called S or NS domains rich in IdoA interspersed by domains containing GlcNAc and GlcA residues. This means that the use of heparin inevitably introduces bias towards proteins that can interact with highly charged polyanions. In other words, heparin can act as a strong cation exchanger, and thus not all heparin binding proteins display high affinity to HS or depend on HS for activity. Nevertheless, heparin-affinity chromatography in conjunction with proteomics strategies, is a simple approach to address the identity and scope of the HS-interactome in various cells and tissues.

Historically, heparin-affinity chromatography has been used as a tool to purify and identify many growth factors, including members of the fibroblast growth factor (FGF), bone morphogenic protein (BMP), vascular endothelial cell growth factor (VEGF), and the wingless (WNT) family of growth factors and morphogens, from tumor homogenates, bovine brain, bone, and secretory fluids (17-22). The HS/heparin interactome in human body fluids and cell lines has also been interrogated in a similar fashion (17-28). Plasma, in particular, is easy to collect and is considered a valuable source of information regarding the physiological status of the body. In addition to classical plasma HSBPs such as antithrombin and platelet factor 4 (PF-4), multiple components of the coagulation and the complement system have been found to interact with HS/heparin. The high enrichment of HSBPs within these pathways suggests a crucial role for GAG-protein interactions in the coordination of these systems. Other important pathways relate to the innate immunity, cell adhesion, blood vessel development and regulation of proteolysis.

Many pathogens and virulence factors utilize GAGs in various ways, and several studies have addressed the molecular identity of the HSBPs responsible for these processes. As an example, proteomics screening for heparin-binding merozoite proteins derived from the malaria parasite *Plasmodium Falciparum* managed to identify hundreds of HSBPs (29). These proteins could be functionally categorized into 14 groups involved in transport, pathogenesis and protein

catabolic processes. Interestingly, members of the PfRhopH complex were among the targets displaying the highest affinity to heparin. These proteins are normally secreted by specialized apical organelles named rhoptries, which are essential for parasite invasion and nutrient acquisition (30). A similar study aiming to define the heparin-binding proteome of *Toxoplasma gondii* identified several novel HSBPs involved in parasite development, suggesting that host HS not only regulates invasion but also parasite maturation (31). Other studies have applied similar strategies to identify novel bacterial virulence factors, toxins from viperid snake venoms, and antimicrobial compounds from egg white (32-34). It might be worth to emphasize that most parasites and microbes do not make HS or heparin, thus these organisms exploit host HS to mediate the various activities described above.

One advantage of affinity chromatography is that it provides some additional information on the interaction of GAG-binding proteins with their ligands, by measuring the ionic strength required for their displacement from the affinity matrix. For example, a proteomics screening of human plasma, lung microvascular endothelial cells and human pulmonary fibroblasts, in combination with affinity chromatography onto matrices of immobilized GAGs (heparin, HS, chondroitin sulfate and dermatan sulfate) suggested sharp differences in “the salt-dependent” binding of several proteins to heparin compared to the other ligands (24). In general, electrostatic forces drive heparin-protein interactions, and thus basic amino acids (arginine, lysine and histidine) are typically found in HS/heparin-binding sites. However, hydrogen bonding and van der Waals forces are also important. In fact, HS binding sites often contain hydrophobic residues that can stack along the axial face of sugars. Interestingly, the affinity and binding kinetics of several HSBPs towards HS/heparin correlate with their molecular function and biological activities (35). For example, database queries of a large surface plasmon resonance (SPR) dataset, suggested that matrisome-associated proteins and lipoproteins bind with higher affinity to HS/heparin compared to other HSBPs. Molecular functions associated with high-affinity binding included enzyme-inhibition and protein dimerization.

Although many novel HSBPs have been identified, there tends to be under-representation of some subclasses of proteins in specific subcellular locations. Most molecular events coordinated by GAGs take place near the plasma membrane, but membrane-bound and membrane-associated HSBPs have been underrepresented in most proteomics screening to date. Difficulties in working with membrane proteins include poor solubility, difficulties inherent in detergent extraction, protease resistance, and the fact that intracellular proteins such as histones can also bind to heparin, outcompeting less abundant membrane proteins. To partially circumvent these problems, *Xu et al* developed a cell-surface biotinylation strategy coupled to heparin-affinity chromatography and mass spectrometry (36). The authors identified HMGB1 as a surface associated HSBP in U937 cells and established a crucial role for HS in the oligomerization of and signaling by receptor for advanced glycation end products (RAGE) in endothelial cells. Using a different approach, *Ori et al* isolated membrane-enriched fractions from rat liver homogenate using subcellular fractionation (37). This method resulted in the identification of 147 HSBPs, including several membrane receptors.

A recent study by *Thacker et al*. showed that tailor-made matrices can be generated to identify HSBPs having preferences towards 3-*O*-sulfated HS chains (27). Generally, it is thought that 3-*O*-sulfation occurs late during HS biosynthesis, after the formation of preferential target sequences. The 3-O-sulfotransferases (Hs3sts) constitute the largest family of HS-modifying enzymes, with distinct, albeit overlapping tissue and cellular expression. The Hs3sts fall into two major subfamilies based on whether they act on N-sulfoglucosamine residues on the reducing side of GlcA or IdoA. The former (Hs3st1 and Hs3st5) can generate binding sites for antithrombin, whereas the latter (Hs3st2, 3a, 3b , 4, 5, and 6) generate binding sites for the gD glycoprotein of Herpes simplex virus (38). In the study by *Thacker et al.*, HS affinity matrices were enzymatically engineered with and without 3-O-sulfate groups using recombinant Hs3st1 or Hs3st2. Fractionation of multiple animal sera followed by proteomics analysis identified several HSBPs with a preference towards 3-O-sulfated HS structures. Among them, neuropilin

1, a key regulator of vascular development and axonal guidance, was shown to bind more avidly to HS modified by 3-O-sulfation, with preference for HS modified by Hs3st2. In ex vivo experiments with murine explants of dorsal root ganglia, inactivation of Hs3st2 modulated semaphorin-3a induced axonal growth, whereas inactivation of Hs3st1 did not.

In general, a common caveat in studies based on affinity chromatography relying on naturally occurring HS/heparin is the heterogeneity of the matrices due to the intrinsic polydispersity of the chains. For example, better defined glycan libraries would be required to identify HSBPs that bind to discrete HS-oligosaccharide motifs. To achieve this goal, progress in chemical and chemoenzymatic GAG synthesis is desperately needed (39-42). Challenges and opportunities in this field have been reviewed elsewhere (43). Finally, HS-protein interactions are obviously also dependent on specific molecular features at the protein binding sites. Unfortunately, the identification of protein determinants that facilitate HS-recognition turned out to be less straightforward than initially predicted.

2.Heparan Sulfate Binding Sites: One Story, Many Tales

2.1.On the Nature of HS/Heparin Binding Sequences

In 1989, *Cardin and Weintraub* published a key paper highlighting the importance of the primary structure of HSBPs for GAG recognition (44). By applying sequence analysis and molecular modeling to the heparin-binding domains of four human proteins (apolipoprotein B, apolipoprotein E, vitronectin and platelet factor 4), the authors reported the presence of semi-conserved linear sequences of basic amino acids (lysine, arginine and histidine) interspersed by other, often hydrophobic, residues. Two consensus sequences for GAG binding were then proposed: [-X-B-B-X-B-X-] and [-X-B-B-B-X-X-B-X-], where B denotes a basic amino acid and X a hydrophobic amino acid residue. Modeling of these motifs predicted their presence in alpha-helixes, with the basic residues aligned on the same side towards the solvent, and the hydrophobic residues pointing towards other structural features in the protein. Given that alpha

helices have 3.4 residues per turn, several of the basic residues would align along one face of the helix, facilitating its interaction with the sulfated domains in HS. Although the presence of these motifs suggests that a protein might bind to HS, many HSBPs do not contain Cardin-Weintraub sequences. GAG binding sites can also be located on beta strands and sheets, in which case the positive residues would need to alternate along the strand in order to generate a positively charge surface for docking. In other cases, binding sites are generated by folding of different protein domains generating a common positively charged surface that is favorable for interaction. One such example is found in the antithrombin binding site. Also, the idea of well-conserved motifs within HS-binding sites is difficult to reconciliate with the low degree of sequence conservation across the large number of HSBPs currently known.

Following a similar approach, *Margalit et al* proposed that a 20 Å distance between the basic amino acids (most frequently arginine) is crucial for the interaction of several HSBPs with heparin (45). Following these initial findings, the patterns (and spacing) of basic amino acids within heparin binding sites have been further refined (46).

Recent studies by *Torrens et al* demonstrated that minimal structural motifs denoted "CPC clip motifs" (C: cationic and P: polar residues) are conserved among all heparin-binding domains deposited in the PDB database (47,48). These motifs are not necessarily part of continuous linear structural elements and are thought to act as "staples", to pin the GAG chains onto the protein binding sites. Similarly, an exhaustive collection of experimentally determined HSBPs was recently reported (49). This dataset was subjected to sophisticated network analysis, to elucidate so far unnoticed structural commonalities. In total, 437 non-redundant HSBPs were analyzed using novel sequence similarity metrics and graph analysis. Again, the linear amino acid sequences across all HSBPs were found to be highly variable, ruling out a universal heparin-binding sequence. However, several shorter (mostly tripeptides) and widely spaced motifs were found to be conserved. Based on these findings, the authors proposed a model in which the three-dimensional arrangement of these motifs on the protein surface, and

not the primary sequence per se, is what determines the structural basis for HS/heparin-protein interactions. In conclusion, the data so far is consistent with the idea of HSBPs have different structural elements (and binding modes) to engage GAGs, which probably translates into different affinities, specificities and biological responses.

2.2.New Experimental Approaches to Identify HS/Heparin-Binding Sites

Since the number of reported HSBPs is continuously growing, there is also need for general methods to accurately identify key residues involved in HS recognition. Sequence analysis in conjunction with site-specific mutagenesis and biophysical methods, have been successful in establishing structure-function relationships for specific GAG-protein interactions (50). However, these are labor intensive methods and faster approaches are required to facilitate more informed decisions. In that line, *Vivès RR et al* reported a procedure based on chemical crosslinking of individual HSBPs to EDC/NHS- activated heparin beads (51). The protein-carbohydrate conjugates were subjected to proteolytic digestion using thermolysin, an endopeptidase with broad specificity. After that, the identity of the remaining crosslinked peptide fragments, covering the heparin-binding site, was assessed by Edman degradation. The workflow was applied to three known HSBPs: the pseudorabies virus (PRV) envelope glycoprotein gC, the CC chemokine RANTES and the C-terminal fragment of the Laminin-5 α3 chain. Titration of the protein concentration affected the identification of the binding sequences, leading to the discovery of a novel low affinity heparin-binding site in RANTES.

An alternative MS-based method was also reported by *Ori et al* (52). Their “Protect and Label” strategy is based on the treatment of HSBPs with Sulfo-NHS-acetate to chemically modify exposed primary amines. Since the labeling is performed while the protein is bound to heparin beads, lysine residues involved in binding are generally protected from derivatization. Labeled proteins were then displaced from the beads by salt-elution and subjected to NHS-biotin treatment to specifically tag the protected lysines in the binding site. Proteins were then trypsinized, and biotinylated peptides were isolated using Strep-Tactin-sepharose

chromatography, followed by MS detection. Three known HSBPs: FGF-2, PF-4 and pleiotrophin, were analyzed through this protocol rendering detailed information on their HS/heparin-binding sites. Unfortunately, both approaches heavily rely on NHS-chemistry, which mainly targets exposed lysine residues. Therefore, very little information can be obtained for other amino acids such as asparagine, serine or tyrosine that are also commonly found in HS/heparin-binding sites. Notwithstanding, these protocols have the potential to be easily adapted to medium/high-throughput formats to complement the information derived from HSBPs global discovery screenings.

2.3.Computational Tools to Rationalize Molecular Interactions at the HS/Heparin binding sites

New insights into the nature of GAG-protein interactions have been uncovered through the power of computational algorithms modelled on available structural data. Biomolecular docking techniques followed by molecular dynamics (MD) simulations are useful strategies to interrogate biological systems at atomic resolution. Generally speaking, the computational treatment of GAG-protein interactions is challenging. Some of the difficulties relate to the high conformational flexibility of GAG-oligosaccharides, the indispensability of solvent and electrolytes to understand the interaction under physiological conditions, and the poor complementarity between HSBPs and the ligand in their interfaces. Moreover, only a handful GAG-protein complexes have been solved and deposited in public repositories (~100). On the other hand, there is a growing body of literature addressing some of these problems and demonstrating that computational approaches are still useful tools. When applied to HS/heparin-protein interactions, current “in silico” strategies can facilitate the identification of putative HS-binding sites and might also give clues as to the specificity and selectivity towards different GAG sequences or motifs.

Since basic residues such as lysines and arginines are known to play pivotal roles in GAG-recognition, one popular way of predicting HS/heparin-binding sites is through the

calculation of the protein surface electrostatic potential. The localization of electropositive patches on a protein surface, by tools such as APBS and DeepView, is often a good indicator of the presence of a putative binding site. However, such approaches are built on the assumption that GAG-protein interactions are mainly driven by electrostatic forces, which might not be entirely correct in all cases. In fact, the accumulated data suggests a substantial contribution of non-coulombic forces as well. More sophisticated approaches rely on sulfated probes to identify areas of solvent displacement to initially locate regions that are energetically favorable for GAG-binding (53,54). New online servers for prediction of binding sites have also been developed during the last years, bringing computational approaches closer to the mainstream biological community (55).

Once a binding site has been located, GAG-oligosaccharide libraries can be probed by biomolecular docking techniques (56,57). These approaches aim to achieve a close fit of a ligand into a targeted site, giving insights into the affinity of binding and preferable binding poses. As discussed above, chemical and enzymatic synthesis of GAG-oligosaccharides is very challenging. One advantage of virtual library screenings is that thousands of different structures and conformations can be quickly tested “in silico”. In such a way, promising candidate compounds can be identified, and experimentally verified later on, through other biochemical or biophysical methods. Recently, sophisticated protocols have been reported where combinatorial virtual library screening (CVLS) algorithms were applied to a large set of oligosaccharides, ranging from disaccharide to hexasaccharides (58). All potential hits were subsequently parsed through rigorous logical filters capable of segregating the structures into “high-affinity” and “high-specificity” sequences. Compared to other biomolecules such as polypeptides or nucleic acids, GAG-oligosaccharides contain a high number of rotable bonds, which essentially limits computational docking to di-, tetra- and hexasaccharides. Another caveat is that sometimes is difficult to distinguish between 180° pairs of GAG-oligosaccharide orientations, which might look very similar in either direction (59).

Docking alone doesn't render information on the free-energy of binding. Typically, when the compounds have been fit into the binding sites, the complexes are subjected to MD simulation workflows. MD simulations facilitate the analysis of the trajectories of structural coordinates as a function of time, under the influence of a force field. These force fields entail a collection of equations designed to reproduce molecular geometry and selected properties of the analyzed structures. Several glycan specific force fields have been developed, including GLYCAM and CHARMM (60,61). The total free-energy of binding and the contribution of individual residues can be deduced by averaging the mechanical contributions of the interacting atoms. Application of these approaches to Annexin A2 and PECAM-1 have revealed high- and low-affinity binding sites as well as a dependency on the length of the GAG-chains for proper interaction (62).

A final detail to keep in mind, is that GAG-protein interactions are largely mediated by solvents, but taking this into consideration is often computationally expensive. However, a few studies have shown that docking and MD simulations with or without solvent, give different results from an energetics point of view (63). In fact, the analysis of GAG-protein interfaces in structures deposited in the PDB database, reveals that they are generally more hydrated than protein-protein interfaces. Indeed, half of the interactions at GAG-protein interfaces appeared to be water-mediated. Although many challenges still remain, the development of computational tools have the potential to accelerate the rational design of oligosaccharide mimetics capable of fine-tuning GAG-protein interactions in different biological systems. Progress in this area is of critical importance.

3.A Functional Glance at the HS/Heparin Interactome

3.1. Network Stratification and Functional Exploration of the HS/Heparin-Interactome

The most comprehensive list of HSBPs to date, was compiled by *Ori et al* through database searches and literature data mining (37). This impressive effort resulted in a final

number of 435 non-redundant proteins, which were subjected to network analysis to facilitate the identification of functional and structural patterns associated with HS/heparin binding. One notable result from that study was that HSBPs tend to form molecular networks with higher average clustering coefficient compared to extracellular non-HSBP networks. In other words, they generate densely interconnected functional modules. Since more recent studies have expanded the scope of this subproteome, we decided to revisit the HS-interactome through the lens of network biology, by taking a slightly different approach and emphasizing the characteristic modularity of HSBP interaction networks. In order to do that, we used the 435 proteins previously reported by Ori et al, and supplemented them with new HSBPs identified in later high-throughput studies. All in all, we compiled a final list of 530 non-redundant proteins (**Supplemental table 1.1**), which were subjected to a molecular networking workflow as described below.

The final HSBP list was used to generate a protein network based on physical and functional protein-protein associations (PPA). High confidence PPA (association score > 0.7) were extracted from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database. STRING is an integrated collection of protein-protein interactions based on direct binding data and/or inferred associations from the literature, public databases and other repositories (64). At the time of this study, the whole STRING-DB high confidence PPA-network contained a total of 15,131 proteins and 35,9776 associations, with an average of 47.6 associations per protein. Of the 530 HSBPs in our compiled list, 488 mapped to the high confidence PPA network. This new STRING subnetwork was then selected for further analysis, reflecting 3,136 associations with an average of 12.8 associations per protein. Functional enrichment analysis of the network was performed through the Database for Annotation, Visualization and Integrated Discovery (DAVID) (65) (**Supplemental Table 1.2**). DAVID was run using default settings with thresholds: count >= 2 and EASE <= 0.1. The EASE score is a modified Fisher Exact P-Value which was further adjusted using a Benjamini correction.

As illustrated in **Supplemental Fig. 1,** functional enrichment analysis of this updated HSBP network, retrieved expected ontology terms relating to HS/heparin-binding. However, only 25% of the input proteins were annotated with HS/heparin binding functions, indicating that the current level of annotation in public repositories is less than satisfactory. The largest cluster of functional annotation encompassed terms associated with cellular signaling pathways such as FGF, BMP, RAGE, integrin, and Hedgehog signaling pathways, emphasizing the role of HS (and HSBPs) in signal transduction and as sensors of changing environmental cues. Other biological functions reflected in the ontology tree map included lipid transport and metabolism, chemotaxis, morphogenesis, complement, receptor scavenging activity, and serine-type endopeptidase activity. Some HSBPs were associated with GO terms pointing to a pure structural role, such as conferring elastic and tensile properties to the molecular architecture of the ECM.

Given the previously reported modularity of HSBP networks, we applied the Louvain method for community detection (66) to better identify and group HSBP clusters displaying higher density of interconnected nodes than expected by random chance. This method is based on a heuristic algorithm that aims to detect communities (or cluster of nodes) by optimizing modularity. Modularity is a measure of the tendency of nodes in a large network to display higher local interconnectivity (or edge density) compared with their connections with the rest of the network. This community clustering is an iterative process that is repeated until maximum modularity is achieved, and a hierarchy of communities is generated. By applying this method, we managed to isolate 14 communities (**Supplemental Table 1.3**). The identified clusters were further segregated via force-directed visualization algorithms and subjected to functional enrichment analysis.

Interestingly enough, nearly 75% of all input proteins ended grouped into 6 major clusters. As shown in **Figure 1.3A-C**, three of these clusters displayed a hybrid composition, showing a tendency to get further segregated into at least one more component. Close

inspection of the network revealed that one of these communities (**Fig. 3A**) was mainly composed by members of the complement system on one side, and by chemokines on the other side. Points of crosstalk were defined by 2 focal nodes, factor C3 and factor C5. As expected, functional enrichment of the cluster as a whole, also reflected molecular activities related to these subcomponents.

Although the role of HS in the regulation of chemokine activity has been extensively studied, less is known about its overall impact on complement-mediated functions. One notable exception is complement factor H (CFH), a key fluid-phase regulator of the alternative complement pathway. CFH has been shown to protect from complement-mediated cytotoxicity through its interaction with HS chains at the surface of host cells, where it remains anchored acting as a cofactor for CFI-inactivation of C3b and promoting accelerated decay of the C3 convertase (67-69). Since HS is not expressed by microbes, this is an effective strategy to distinguish between self and non-self-entities during the unfolding of the complement cascade. On the other hand, familial mutations in the HS-binding domain of CFH are associated with impaired HS-binding, resulting in a severe form of thrombotic microangiopathy known as the atypical hemolytic uremic syndrome (70,71). Similarly, allotypic variants of the CFH gene have been associated with development of age-related macular degeneration (AMD), one of the leading causes of adult blindness in the western world (72). Progression into AMD has been found to depend on changes in the HS composition of the human Bruch's membrane/choroid, leading to focal deposits of lipids and proteins (73).

In addition to CFH, many other complement factors are de facto HSBPs, raising the possibility of HS serving as a global regulator of the complement system. Examination of their binding kinetics and dissociation constants (Kd) towards heparin points to a broad range of affinities (from 2 to 320 nM) as determined by surface plasmon resonance (74). As mentioned before, cluster subcomponents reflecting complement and chemokine activities were also interconnected through factor C3 and C5, both central molecules in the complement cascade.

Interestingly, some of their proteolytic fragments are known to display immunomodulatory functions of their own. For example, C3a and C5a can induce smooth muscle contraction, increased vascular leakage, modulation of leukocyte recruitment, and generation of cytotoxic oxygen radicals. The fact that all proteins in the network were selected based on their HS-binding capacity suggests that HS could potentially have a role in mediating biological cross-talks.

A second cluster identified by our method was characterized by the grouping of a large variety of growth factors (**Fig. 3B**), including a central component comprising the FGF-protein family, which has been extensively studied in the context of their HS-binding capacity. Other growth factors included members of the BMP signaling pathway such as BMP2 and BMP4, as well as known BMP antagonists such as Fstl1. BMP signaling is a critical pathway that is required for proper bone and lung development, among other processes. Indeed, genetic alteration of HS-structure has been shown to affect downstream BMP signaling, leading to a wide range of morphogenetic defects (75). Finally, a smaller component was also identified in connection with these growth factor communities, encompassing proteins involved in sodium and calcium sensing and transport. There is some data pointing to a role for the syndecan family of HSPGs in the regulation of stretch activated ion-channels. A recent review on this topic can be found elsewhere (76).

Regulation of the coagulation system is another striking example of how HS structures coordinate complex protein-protein interaction networks. As shown in **Figure 1.3C**, many HSPBs involved in the regulation of the coagulation cascade were efficiently captured by the Louvain community clustering. As expected, functional enrichment reflected the large impact of HS-chains on regulation of protease activity by facilitating interactions both with components displaying serine-protease activity as well as components with serine-protease inhibitor capacity. This fine balance is sometimes mediated by differences in specificity towards various HS-motifs, as mentioned earlier in the context of thrombin and antithrombin. All serine protease

inhibitors (or serpins), exert their function through a common mechanism, which entails dramatic conformational changes and covalent trapping of the targeted protease, followed by clearance of the complexes from circulation through recognition by specific receptors. Interestingly, some of the most studied GAG-binding serpins such as antithrombin (serpin C1) and Heparin cofactor II (serpin D1) can inhibit procoagulant factors such as thrombin and Factor Xa, whereas other such as serpin A5 can target anticoagulant factors such as Protein C. In general, the rate of these interactions is largely accelerated by binding to GAGs.

A fascinating evolutionary aspect of these systems is that in jawless fish such as lampreys, an unexpected connection between coagulation and regulation of blood pressure has recently been uncovered. Several recent studies have shown evidence for the ability of lamprey angiotensinogen to regulate vascular tone as well as behaving like a potent heparin-dependent thrombin inhibitor, totally functional in the lamprey coagulation system (77,78). These two ancestral functions seem to have become uncoupled in gnathostomes during evolution, but it can be speculated that perhaps some of the mammalian coagulation serpins could still display some kind of undiscovered vasomodulatory properties. The evolutionary history of the HS-interactome and its co-evolution in parallel with the GAG-biosynthetic machinery is an area that hasn't been deeply explored and warrants further investigation.

Finally, we also identified other HSBP showing similar patterns as described above (**Figure 1.3d-f**), getting densely grouped into clusters reflecting common functional themes such as lipoprotein metabolism and uptake, cholesterol transport, assembly of the extracellular matrix and growth factor signaling. All the functional enrichment data for each individual cluster is summarized in **Supplemental Table 4**.

Summarizing, it appears that in spite of the large structural and molecular diversity of the HSBPs, they tend to group into protein-protein interaction networks with common functional themes. We suggest that there is a potential link between the evolutionary pressure driving the acquisition of HS-binding capacity, and the concomitant increase in multicellularity and

increased complexity. The fact that most unicellular organisms lack the biosynthetic tools to assemble complex GAG polysaccharides is a strong indication that this might be the case. In a multicellular context, the flux of chemical information is obviously dramatically increased since cells are not only responsive to cell-autonomous and environmental changes, but they are also expected to integrate and respond to communication originating from adjacent cells. These relationships become more imperative as we move from cells to tissues to organs to the enormous complexity of invertebrates and vertebrates. As shown in this section, most HSBPs are either present at the cell membrane or embedded in the extracellular matrix, which constitute the forefront of intercellular communication. Many of the biological processes in which they are involved, such as cellular out-side-in signaling or native immunity, directly adhere to the notion of chemical communication that need to be carefully integrated to facilitate coordinated responses. Perhaps, the evolutionary acquisition of HS-binding ability allowed for an additional global check-point that is advantageous to multicellular biological systems, to quickly adapt and more effectively respond, to changing environmental cues. If that is the case, the large chemical space that can be coded into HS-structures will then constitute a flexible arsenal that can be rapidly deployed to fine-tune the effect and timing of these responses.

4.Future Perspectives

To the best of our knowledge, it appears that most studies addressing the scope of the HS-interactome have mainly focused on qualitative aspects, i.e. the complement of proteins present in a sample at a given time and set of conditions. However, it is reasonable to assume that quantitative changes might be important as well, especially during pathological processes. For example, during systemic inflammatory responses and sepsis, it has been reported that the levels of circulating disease-associated molecular patterns (DAMPs) are often increased. In patients undergoing severe septic shock. Some of these markers, such as histones and high mobility group box 1 (HMGB1), correlate with poor survival, higher APACHE II scores, and

increased levels of pro-inflammatory cytokines (79,80). Some of these markers are also well-known HS-binding proteins, which are completely absent or appear at very low-levels in normal human plasma. Tracking the rise of HS-binding DAMPs over time in a quantitative fashion might have some predictive value to facilitate sepsis diagnosis and monitoring of therapies.

Similarly, dramatic remodeling of HSPGs in the vascular glycocalyx can be triggered by a plethora of inflammatory conditions. This process is partially mediated by upregulation of sheddases such as metalloproteases, that are able to cleave off HSPG ectodomains from the vascular wall and release them into the circulation. Also, the endogenous heparanase, a glycosidase with endoglucuronidase activity can act upon HS chains to release oligosaccharides, which will substantially boost the inflammatory response. In the light of these findings, it is conceivable that alterations of the HS-structures themselves, through genetic or posttranslational mechanisms, will probably impact the HS-interactome resulting in quantitative changes in the HSBP networks and hence, altering their biological activities.

As suggested throughout this review, during coordinated molecular events such as signaling, HS may also engage multiple components within the same pathway. How these protein-carbohydrate interactions will affect the spatial-temporal dynamics of signaling networks at a systems level is difficult to predict. As discussed above, dissection of the structural details that facilitate these interactions requires access to well-defined oligosaccharide libraries as well as novel screening methodologies to measure simultaneous binding of receptors and protein ligands. Notwithstanding, given the large number of HSBPs reported so far and the fact that many of them occur in shared pathways, clearly indicate that the role of HS in regulating biological processes might be more comprehensive than previously expected.

Finally, new bioinformatics approaches are definitely needed to rapidly evaluate experimental data to be able to make predictions on structure-function relationships. Unfortunately, the level of annotation regarding the HS-binding properties of proteins in public repositories is currently poor. Some specialized knowledge databases have tried to solve for

this problem. For example, very recent updates to Matrix-DB explicitly address GAG-related features, including incorporation of 50 GAG-protein interacting sequences, cross-references to the carbohydrate database GlyTouCan and a GAG-3D builder (81,82). However, much more is required. This issue is of critical importance since most pathway analysis and general bioinformatics tools collect their primary information from public databases. That means that enrichment or changes in the HS-interactome in experimental datasets might be obscured by the current lack of appropriate annotation.

Realization of the potential of the HS-interactome as a molecular window to explore basal cellular physiology and to understand complex diseases, might require multidisciplinary efforts. There is now sufficient evidence showing that HS-mediated functions affect a broad range of physiological and pathophysiological processes, making them of general interest to the scientific community and no longer exclusively confined to the realm of glycobiology experts and carbohydrate chemists.

Acknowledgements

Chapter 1, in full is being prepared for submission as a publication. Toledo A.G., Sandoval, D.R., Sorrentino, J.T., Kellman B.P., Lewis, N.E., Esko, J.D. I was the secondary investigator for this paper.

Figures

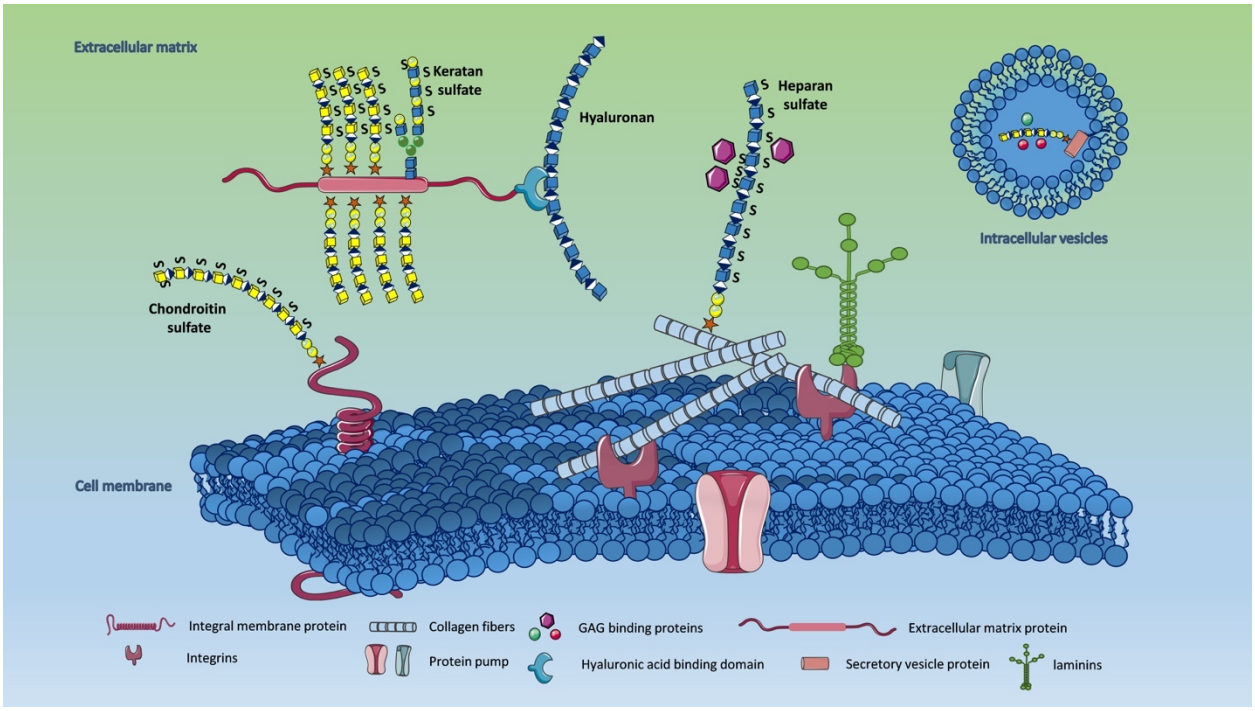


Figure 1.1. General overview of the major classes of proteoglycans and glycosaminoglycans present at cell surfaces and in the extracellular matrix. Proteoglycans can either be secreted into the extracellular space or anchored to the cell surface as transmembrane proteins. Sulfated regions of the oligosaccharide chains form binding sites that facilitate interaction with glycosaminoglycan binding proteins.

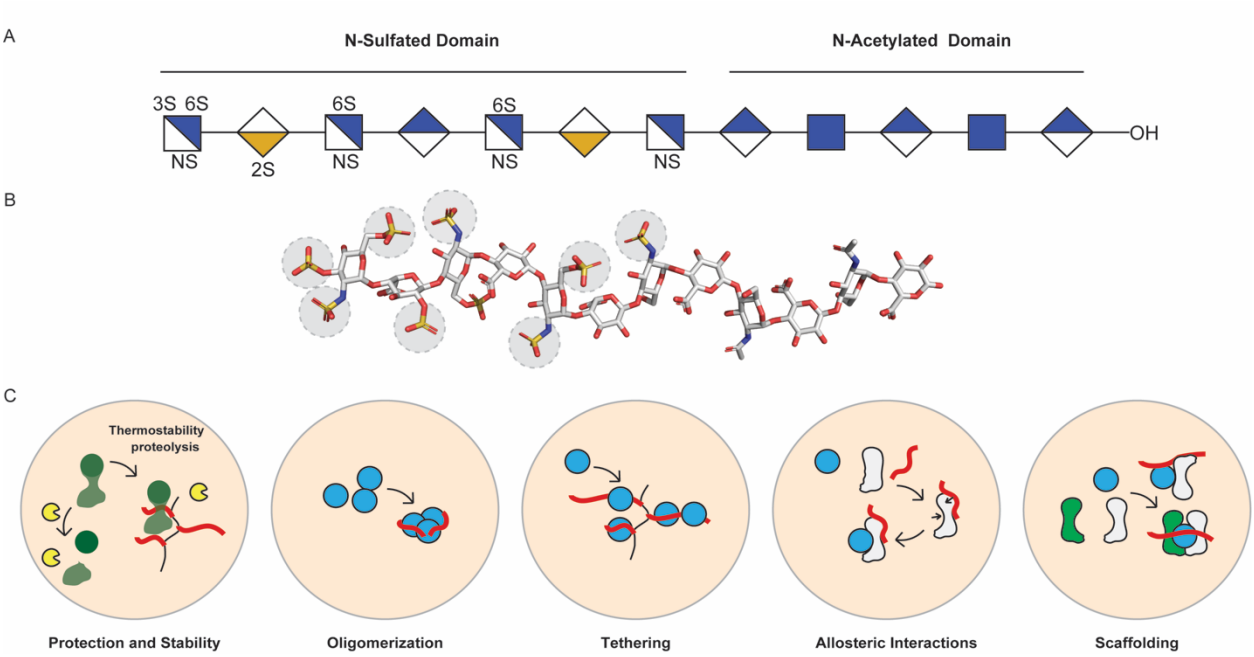


Figure 1.2: Principles of heparan sulfate-protein interactions. (a) Symbolic representation of a dp12 heparan sulfate oligomer composed of alternating disaccharide units of N-acetylglucosamine and glucuronic/iduronic acid. These units can be modified with sulfate groups at the 3-, 6-, and N-position on the N-acetylglucosamine or at the 2 position of the iduronic acid units. Heparan sulfate chains display typical distributions of N-sulfated or N-acetylated domains that can be recognized by multiple ligands. (b) Stick representation of (a). Sulfur (yellow), Oxygen (red), Nitrogen (blue), Carbon (white). (c) Heparan sulfate can block proteolytic sites, promote stability, induce oligomerization, and/or tethering. It can also induce conformational changes, allowing ligands to dock on target proteins as well as acting as a molecular scaffold to bridge protein complexes.

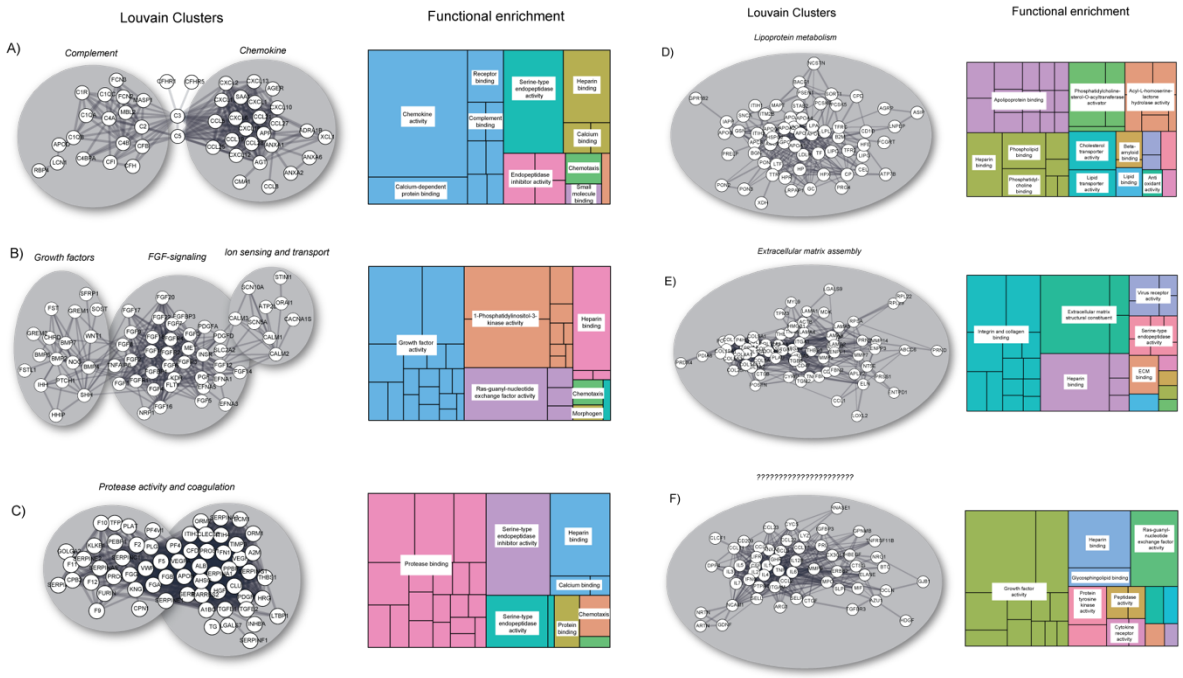


Figure 1.3: Cluster analysis of 530 non-redundant human HS/heparin binding proteins derived from literature searches. Functional enrichment of complement and chemokine (a), FGF signaling (b), coagulation (c), lipoprotein metabolism (d), matrix assembly (e), and growth factor signaling (f).

References

1. Lindahl, U., Couchman, J., Kimata, K., and Esko, J. D. (2015) Proteoglycans and Sulfated Glycosaminoglycans. in *Essentials of Glycobiology* (rd, Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. eds.), Cold Spring Harbor (NY). pp 207-221

2. Mizumoto, S., Yamada, S., and Sugahara, K. (2015) Molecular interactions between chondroitin-dermatan sulfate and growth factors/receptors/matrix proteins. *Current opinion in structural biology* **34**, 35-42

3. Caterson, B., and Melrose, J. (2018) Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* **28**, 182-206

4. Stewart, M. D., and Sanderson, R. D. (2014) Heparan sulfate in the nucleus and its control of cellular functions. *Matrix biology : journal of the International Society for Matrix Biology* **35**, 56-59

5. Kovalszky, I., Hjerpe, A., and Dobra, K. (2014) Nuclear translocation of heparan sulfate proteoglycans and their functional significance. *Biochimica et biophysica acta* **1840**, 2491-2497

6. Wang, L., Fuster, M., Sriramarao, P., and Esko, J. D. (2005) Endothelial heparan sulfate deficiency impairs L-selectin- and chemokine-mediated neutrophil trafficking during inflammatory responses. *Nature immunology* **6**, 902-910

7. Kobayashi, F., Yamada, S., Taguwa, S., Kataoka, C., Naito, S., Hama, Y., Tani, H., Matsuura, Y., and Sugahara, K. (2012) Specific interaction of the envelope glycoproteins E1 and E2 with liver heparan sulfate involved in the tissue tropismatic infection by hepatitis C virus. *Glycoconjugate journal* **29**, 211-220

8. Liang, W. G., Triandafillou, C. G., Huang, T. Y., Zulueta, M. M., Banerjee, S., Dinner, A. R., Hung, S. C., and Tang, W. J. (2016) Structural basis for oligomerization and glycosaminoglycan binding of CCL5 and CCL3. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 5000-5005

9. Crown, S. E., Yu, Y., Sweeney, M. D., Leary, J. A., and Handel, T. M. (2006) Heterodimerization of CCR2 chemokines and regulation by glycosaminoglycan binding. *J Biol Chem* **281**, 25438-25446

10. Zhang, F., Zhang, Z., Lin, X., Beenken, A., Eliseenkova, A. V., Mohammadi, M., and Linhardt, R. J. (2009) Compositional analysis of heparin/heparan sulfate interacting with fibroblast growth factor.fibroblast growth factor receptor complexes. *Biochemistry* **48**, 8379-8386

11. Xu, D., Young, J. H., Krahn, J. M., Song, D., Corbett, K. D., Chazin, W. J., Pedersen, L. C., and Esko, J. D. (2013) Stable RAGE-heparan sulfate complexes are essential for signal transduction. *ACS chemical biology* **8**, 1611-1620

12. Huang, Z., Tan, Y., Gu, J., Liu, Y., Song, L., Niu, J., Zhao, L., Srinivasan, L., Lin, Q., Deng, J., Li, Y., Conklin, D. J., Neubert, T. A., Cai, L., Li, X., and Mohammadi, M. (2017) Uncoupling the Mitogenic and Metabolic Functions of FGF1 by Tuning FGF1-FGF Receptor Dimer Stability. *Cell reports* **20**, 1717-1728

13. Robinson, C. J., and Stringer, S. E. (2001) The splice variants of vascular endothelial growth factor (VEGF) and their receptors. *Journal of cell science* **114**, 853-865

14. Braga, T., Grujic, M., Lukinius, A., Hellman, L., Abrink, M., and Pejler, G. (2007) Serglycin proteoglycan is required for secretory granule integrity in mucosal mast cells. *The Biochemical journal* **403**, 49-57

15. Ringvall, M., Ronnberg, E., Wernersson, S., Duelli, A., Henningsson, F., Abrink, M., Garcia-Faroldi, G., Fajardo, I., and Pejler, G. (2008) Serotonin and histamine storage in mast cell secretory granules is dependent on serglycin proteoglycan. *The Journal of allergy and clinical immunology* **121**, 1020-1026

16. Casu, B., Naggi, A., and Torri, G. (2015) Re-visiting the structure of heparin. *Carbohydrate research* **403**, 60-68

17. Maciag, T., Mehlman, T., Friesel, R., and Schreiber, A. B. (1984) Heparin binds endothelial cell growth factor, the principal endothelial cell mitogen in bovine brain. *Science* **225**, 932-935

18. Shing, Y., Folkman, J., Sullivan, R., Butterfield, C., Murray, J., and Klagsbrun, M. (1984) Heparin affinity: purification of a tumor-derived capillary endothelial cell growth factor. *Science* **223**, 1296-1299

19. Shing, Y., Folkman, J., Haudenschild, C., Lund, D., Crum, R., and Klagsbrun, M. (1985) Angiogenesis is stimulated by a tumor-derived endothelial cell growth factor. *J Cell Biochem* **29**, 275-287

20. Klagsbrun, M., and Shing, Y. (1985) Heparin affinity of anionic and cationic capillary endothelial cell growth factors: analysis of hypothalamus-derived growth factors and fibroblast growth factors. *Proc Natl Acad Sci U S A* **82**, 805-809

21. Hauschka, P. V., Mavrakos, A. E., Iafrati, M. D., Doleman, S. E., and Klagsbrun, M. (1986) Growth factors in bone matrix. Isolation of multiple types by affinity chromatography on heparin-Sepharose. *J Biol Chem* **261**, 12665-12674

22. Brigstock, D. R., Steffen, C. L., Kim, G. Y., Vegunta, R. K., Diehl, J. R., and Harding, P. A. (1997) Purification and characterization of novel heparin-binding growth factors in uterine secretory fluids. Identification as heparin-regulated Mr 10,000 forms of connective tissue growth factor. *J Biol Chem* **272**, 20275-20282

23. Saito, A., and Munakata, H. (2007) Analysis of plasma proteins that bind to glycosaminoglycans. *Bba-Gen Subjects* **1770**, 241-246

24. Gesslbauer, B., Derler, R., Handwerker, C., Seles, E., and Kungl, A. J. (2016) Exploring the glycosaminoglycan-protein interaction network by glycan-mediated pull-down proteomics. *Electrophoresis* **37**, 1437-1447

25. Bjarnadottir, S. G., and Flengsrud, R. (2014) Affinity chromatography, two-dimensional electrophoresis, adapted immunodepletion and mass spectrometry used for detection of porcine and piscine heparin-binding human plasma proteins. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **944**, 107-113

26. Kumar, V., Hassan, M. I., Kashav, T., Singh, T. P., and Yadav, S. (2008) Heparin-binding proteins of human seminal plasma: purification and characterization. *Molecular reproduction and development* **75**, 1767-1774

27. Thacker, B. E., Seamen, E., Lawrence, R., Parker, M. W., Xu, Y., Liu, J., Vander Kooi, C. W., and Esko, J. D. (2016) Expanding the 3-O-Sulfate Proteome--Enhanced Binding of Neuropilin-1 to 3-O-Sulfated Heparan Sulfate Modulates Its Activity. *ACS chemical biology* **11**, 971-980

28. Raboudi, N., Julian, J., Rohde, L. H., and Carson, D. D. (1992) Identification of cell-surface heparin/heparan sulfate-binding proteins of a human uterine epithelial cell line (RL95). *J Biol Chem* **267**, 11930-11939

29. Zhang, Y., Jiang, N., Lu, H., Hou, N., Piao, X., Cai, P., Yin, J., Wahlgren, M., and Chen, Q. (2013) Proteomic analysis of Plasmodium falciparum schizonts reveals heparin-binding merozoite proteins. *Journal of proteome research* **12**, 2185-2193

30. Counihan, N. A., Kalanon, M., Coppel, R. L., and de Koning-Ward, T. F. (2013) Plasmodium rhoptry proteins: why order is important. *Trends in parasitology* **29**, 228-236

31. Zhang, Y., Jiang, N., Jia, B., Chang, Z., Zhang, Y., Wei, X., Zhou, J., Wang, H., Zhao, X., Yu, S., Song, M., Tu, Z., Lu, H., Yin, J., Wahlgren, M., and Chen, Q. (2014) A comparative study on the heparin-binding proteomes of Toxoplasma gondii and Plasmodium falciparum. *Proteomics* **14**, 1737-1745

32. Guyot, N., Labas, V., Harichaux, G., Chesse, M., Poirier, J. C., Nys, Y., and Rehault-Godbert, S. (2016) Proteomic analysis of egg white heparin-binding proteins: towards the identification of natural antibacterial molecules. *Scientific reports* **6**, 27974

33. Paes Leme, A. F., Kitano, E. S., Furtado, M. F., Valente, R. H., Camargo, A. C., Ho, P. L., Fox, J. W., and Serrano, S. M. (2009) Analysis of the subproteomes of proteinases and heparin-binding toxins of eight Bothrops venoms. *Proteomics* **9**, 733-745

34. Hsiao, F. S., Sutandy, F. R., Syu, G. D., Chen, Y. W., Lin, J. M., and Chen, C. S. (2016) Systematic protein interactome analysis of glycosaminoglycans revealed YcbS as a novel bacterial virulence factor. *Sci Rep* **6**, 28425

35. Peysselon, F., and Ricard-Blum, S. (2014) Heparin-protein interactions: from affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis. *Matrix biology : journal of the International Society for Matrix Biology* **35**, 73-81

36. Xu, D., Young, J., Song, D., and Esko, J. D. (2011) Heparan sulfate is essential for high mobility group protein 1 (HMGB1) signaling by the receptor for advanced glycation end products (RAGE). *J Biol Chem* **286**, 41736-41744

37. Ori, A., Wilkinson, M. C., and Fernig, D. G. (2011) A systems biology approach for the investigation of the heparin/heparan sulfate interactome. *J Biol Chem* **286**, 19892-19904

38. 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

39. Zong, C., Venot, A., Li, X., Lu, W., Xiao, W., Wilkes, J. L., Salanga, C. L., Handel, T. M., Wang, L., Wolfert, M. A., and Boons, G. J. (2017) Heparan Sulfate Microarray Reveals That Heparan Sulfate-Protein Binding Exhibits Different Ligand Requirements. *J Am Chem Soc* **139**, 9534-9543

40. de Paz, J. L., Horlacher, T., and Seeberger, P. H. (2006) Oligosaccharide microarrays to map interactions of carbohydrates in biological systems. *Methods in enzymology* **415**, 269-292

41. de Paz, J. L., Spillmann, D., and Seeberger, P. H. (2006) Microarrays of heparin oligosaccharides obtained by nitrous acid depolymerization of isolated heparin. *Chemical communications*, 3116-3118

42. Noti, C., de Paz, J. L., Polito, L., and Seeberger, P. H. (2006) Preparation and use of microarrays containing synthetic heparin oligosaccharides for the rapid analysis of heparin-protein interactions. *Chemistry* **12**, 8664-8686

43. Pomin, V. H., and Wang, X. (2018) Synthetic Oligosaccharide Libraries and Microarray Technology: A Powerful Combination for the Success of Current Glycosaminoglycan Interactomics. *ChemMedChem* **13**, 648-661

44. Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**, 21-32

45. Margalit, H., Fischer, N., and Ben-Sasson, S. A. (1993) Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues. *J Biol Chem* **268**, 19228-19231

46. Fromm, J. R., Hileman, R. E., Caldwell, E. E., Weiler, J. M., and Linhardt, R. J. (1997) Pattern and spacing of basic amino acids in heparin binding sites. *Archives of biochemistry and biophysics* **343**, 92-100

47. Pulido, D., Rebollido-Rios, R., Valle, J., Andreu, D., Boix, E., and Torrent, M. (2017) Structural similarities in the CPC clip motif explain peptide-binding promiscuity between glycosaminoglycans and lipopolysaccharides. *Journal of the Royal Society, Interface* **14**

48. Torrent, M., Nogues, M. V., Andreu, D., and Boix, E. (2012) The "CPC clip motif": a conserved structural signature for heparin-binding proteins. *Plos One* **7**, e42692

49. 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. *Molecular bioSystems* **13**, 852-865

50. Xu, D., and Esko, J. D. (2014) Demystifying heparan sulfate-protein interactions. *Annual review of biochemistry* **83**, 129-157

51. Vives, R. R., Crublet, E., Andrieu, J. P., Gagnon, J., Rousselle, P., and Lortat-Jacob, H. (2004) A novel strategy for defining critical amino acid residues involved in protein/glycosaminoglycan interactions. *J Biol Chem* **279**, 54327-54333

52. Ori, A., Free, P., Courty, J., Wilkinson, M. C., and Fernig, D. G. (2009) Identification of heparin-binding sites in proteins by selective labeling. *Mol Cell Proteomics* **8**, 2256-2265

53. Lortat-Jacob, H., Grosdidier, A., and Imberty, A. (2002) Structural diversity of heparan sulfate binding domains in chemokines. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1229-1234

54. Bitomsky, W., and Wade, R. C. (1999) Docking of glycosaminoglycans to heparin-binding proteins: Validation for aFGF, bFGF, and antithrombin and application to IL-8. *J Am Chem Soc* **121**, 3004-3013

55. Mottarella, S. E., Beglov, D., Beglova, N., Nugent, M. A., Kozakov, D., and Vajda, S. (2014) Docking server for the identification of heparin binding sites on proteins. *Journal of chemical information and modeling* **54**, 2068-2078

56. Takaoka, T., Mori, K., Okimoto, N., Neya, S., and Hoshino, T. (2007) Prediction of the Structure of Complexes Comprised of Proteins and Glycosaminoglycans Using Docking Simulation and Cluster Analysis. *Journal of chemical theory and computation* **3**, 2347-2356

57. Samsonov, S. A., and Pisabarro, M. T. (2016) Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology* **26**, 850-861

58. Sankaranarayanan, N. V., and Desai, U. R. (2014) Toward a robust computational screening strategy for identifying glycosaminoglycan sequences that display high specificity for target proteins. *Glycobiology* **24**, 1323-1333

59. Forster, M., and Mulloy, B. (2006) Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochemical Society transactions* **34**, 431-434

60. Singh, A., Tessier, M. B., Pederson, K., Wang, X., Venot, A. P., Boons, G. J., Prestegard, J. H., and Woods, R. J. (2016) Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Canadian journal of chemistry* **94**, 927-935

61. Sarkar, A., Yu, W., Desai, U. R., MacKerell, A. D., and Mosier, P. D. (2016) Estimating glycosaminoglycan-protein interaction affinity: water dominates the specific antithrombin-heparin interaction. *Glycobiology* **26**, 1041-1047

62. Gandhi, N. S., and Mancera, R. L. (2009) Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology* **19**, 1103-1115

63. Samsonov, S. A., Teyra, J., and Pisabarro, M. T. (2011) Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *Journal of computer-aided molecular design* **25**, 477-489

64. von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., and Bork, P. (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic acids research* **33**, D433-437

65. Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4**, P3

66. Blondel, V. D. G., Jean-Loup; Lambiotte, Renaud; Lefebvre, Etienne. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10008

67. Pangburn, M. K., Atkinson, M. A., and Meri, S. (1991) Localization of the heparin-binding site on complement factor H. *J Biol Chem* **266**, 16847-16853

68. Ormsby, R. J., Jokiranta, T. S., Duthy, T. G., Griggs, K. M., Sadlon, T. A., Giannakis, E., and Gordon, D. L. (2006) Localization of the third heparin-binding site in the human complement regulator factor H1. *Molecular immunology* **43**, 1624-1632

69. Loeven, M. A., Rops, A. L., Berden, J. H., Daha, M. R., Rabelink, T. J., and van der Vlag, J. (2015) The role of heparan sulfate as determining pathogenic factor in complement factor H-associated diseases. *Molecular immunology* **63**, 203-208

70. Loeven, M. A., Rops, A. L., Lehtinen, M. J., van Kuppevelt, T. H., Daha, M. R., Smith, R. J., Bakker, M., Berden, J. H., Rabelink, T. J., Jokiranta, T. S., and van der Vlag, J. (2016) Mutations in Complement Factor H Impair Alternative Pathway Regulation on Mouse Glomerular Endothelial Cells in Vitro. *J Biol Chem* **291**, 4974-4981

71. Lehtinen, M. J., Rops, A. L., Isenman, D. E., van der Vlag, J., and Jokiranta, T. S. (2009) Mutations of factor H impair regulation of surface-bound C3b by three mechanisms in atypical hemolytic uremic syndrome. *J Biol Chem* **284**, 15650-15658

72. Clark, S. J., Higman, V. A., Mulloy, B., Perkins, S. J., Lea, S. M., Sim, R. B., and Day, A. J. (2006) His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-disease-associated form. *J Biol Chem* **281**, 24713-24720

73. Kelly, U., Yu, L., Kumar, P., Ding, J. D., Jiang, H., Hageman, G. S., Arshavsky, V. Y., Frank, M. M., Hauser, M. A., and Rickman, C. B. (2010) Heparan sulfate, including that in Bruch's membrane, inhibits the complement alternative pathway: implications for age-related macular degeneration. *Journal of immunology* **185**, 5486-5494

74. Yu, H., Munoz, E. M., Edens, R. E., and Linhardt, R. J. (2005) Kinetic studies on the interactions of heparin and complement proteins using surface plasmon resonance. *Biochimica et biophysica acta* **1726**, 168-176

75. Hu, Z., Wang, C., Xiao, Y., Sheng, N., Chen, Y., Xu, Y., Zhang, L., Mo, W., Jing, N., and Hu, G. (2009) NDST1-dependent heparan sulfate regulates BMP signaling and internalization in lung development. *Journal of cell science* **122**, 1145-1154

76. Mitsou, I., Multhaupt, H. A. B., and Couchman, J. R. (2017) Proteoglycans, ion channels and cell-matrix adhesion. *The Biochemical journal* **474**, 1965-1979

77. Wei, H., Cai, H., Wu, J., Wei, Z., Zhang, F., Huang, X., Ma, L., Feng, L., Zhang, R., Wang, Y., Ragg, H., Zheng, Y., and Zhou, A. (2016) Heparin Binds Lamprey Angiotensinogen and Promotes Thrombin Inhibition through a Template Mechanism. *J Biol Chem* **291**, 24900-24911

78. Wang, Y., and Ragg, H. (2011) An unexpected link between angiotensinogen and thrombin. *FEBS letters* **585**, 2395-2399

79. Wang, H., Yang, H., and Tracey, K. J. (2004) Extracellular role of HMGB1 in inflammation and sepsis. *Journal of internal medicine* **255**, 320-331

80. Xu, Z., Huang, Y., Mao, P., Zhang, J., and Li, Y. (2015) Sepsis and ARDS: The Dark Side of Histones. *Mediators of inflammation* **2015**, 205054

81. Chautard, E., Ballut, L., Thierry-Mieg, N., and Ricard-Blum, S. (2009) MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics* **25**, 690-691

82. Clerc, O., Deniaud, M., Vallet, S. D., Naba, A., Rivet, A., Perez, S., Thierry-Mieg, N., and Ricard-Blum, S. (2018) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. *Nucleic acids research*

CHAPTER 2: PROTEOMICS-BASED SCREENING OF THE ENDOTHELIAL HEPARAN-SULFATE INTERACTOME

Abstract

We report a novel proteomics workflow to identify and characterize membrane-anchored and extracellular proteins that bind to heparin. The technique is based on Limited Proteolysis of live cells in the absence of denaturation and fixation, Heparin-Affinity chromatography, and high-resolution LC-MS/MS proteomics, which we designate as LPHAMS. Application of LPHAMS to primary murine and human endothelial cells and U937 cells led to the identification of 75 plasma membrane, extracellular matrix proteins, and soluble secreted proteins, which includes many previously unidentified heparin-binding proteins. The method also facilitates the mapping of the heparin-binding domains, making it possible to make predictions on the location of the heparin-binding site. To validate the discovery feature of LPHAMS, we characterized one of the newly discovered heparin-binding proteins, CLEC14A, a member of the C-type lectin family that modulates angiogenesis. The C-type lectin domain of CLEC14A binds to heparin with nanomolar affinity via a monovalent interaction and the heparin-binding site was mapped by molecular modeling and mutagenesis. CLEC14A can physically interact with other glycosaminoglycans including endothelial heparan sulfate and chondroitin sulfate E, but not with neutral or sialylated oligosaccharides. The combination of limited proteolysis and mass spectrometry led to the identification of previously undocumented glycosaminoglycan-binding proteins and mapping of their ligand binding sites. The technique is applicable to other cells and glycans and provides a way to expand the repertoire glycan-binding proteins for further study.

Introduction

All animal cells express heparan sulfate proteoglycans (HSPGs), either as trans-membrane, secreted, and extracellular matrix proteins or as components of storage granules. HSPGs perform many functions in cells mediated to a large extent by the capacity of the heparan sulfate (HS) chains to interact with other proteins. HSPGs can tether and present cytokines, chemokines, growth factors and morphogens (50). They can also act as a template to bring together growth factors and their receptor tyrosine kinases, lowering the concentration of the growth factor required for signaling. Binding of HS to proteases and protease inhibitors can lead to allosteric activation or inhibition of enzyme activity, for example in the activation of antithrombin and inhibition of Factors II and X in the coagulation cascade (83). HS also can induce oligomerization of soluble and membrane proteins.

Notably, most angiogenic factors and vascular growth factor receptors interact with HS, emphasizing the potential role of HS-protein interactions in vascular development and angiogenesis (84,85). Mice displaying undersulfated HS in endothelial tissues show altered tumor angiogenesis due to dysregulated vascular endothelial cell growth factor (VEGF) and fibroblast growth factor (FGF) signaling and altered chemokine and selectin-mediated responses to acute inflammation (86-88). The size of the endothelial “HS-interactome,” i.e. the repertoire of heparan sulfate-binding proteins (HSBPs) on the surface and surrounding extracellular matrix of endothelial cells, is unknown in part due to technical challenges in working with membrane proteins and protein complexes. Moreover, attempts to elucidate the interactome using live cells and tissue extracts often enriched intracellular proteins, such as DNA and RNA binding proteins (37). To circumvent these problems, preparative steps involving purification of plasma membranes has been applied to cells and tissues (37). Cell surface biotinylation strategies coupled with streptavidin enrichment prior to affinity chromatography also have proven useful in studies of cultured cells (36,37). Typically, heparan-sulfate binding

membrane proteins are difficult to identify in this way because of the need for detergents or prior enrichment techniques.

In this report, we describe a new simple strategy to identify plasma membrane and extracellular HSBPs that also permits the determination of the binding domains that interact with heparin/heparan sulfate. The workflow combines Limited Proteolysis in the absence of denaturation, Heparin-Affinity chromatography, and high-resolution LC-MS/MS proteomics (LPHAMS). Application of LPHAMS to endothelial cells led to the identification of known HSBPs including membrane receptors, secreted, and extracellular matrix proteins, along with a set of previously unknown HSBPs. As a validation of LPHAMS, we characterized the heparin binding properties of CLEC14A, a previously undocumented HSBP involved in angiogenesis.

Experimental Section

Limited Proteolysis Proteomics Screening. Confluent HUVEC grown in 100 mm diameter dishes in EGM-2 medium (lonza) were washed twice with 5 ml of M199 medium (Life Technologies). Cells were treated with Proteinase K (250 ng/ml) in M199 for 10 min at room temperature. The solution was collected, centrifuged at 400 x g to remove cellular debris, and then placed on ice. The samples were applied to a 1-ml HiTrap heparin-Sepharose column (GE Healthcare) equilibrated in 150 mM NaCl in 25 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) HEPES (pH 7.1). Columns were washed with 0.3 M NaCl in 25 mM HEPES buffer (pH 7.1) to remove low affinity binding proteins and step eluted with 1 M NaCl in 25 mM HEPES buffer (pH 7.1). An in-solution digestion was performed on proteins in the high salt wash with mass spec grade trypsin gold (Promega) at 37°C. Peptides were desalted using C18 Tips (Pierce), and dried using a speed-vac centrifuge. Samples were analyzed by liquid chromatography/tandem mass spectrometry (LC-MS/MS) on a TripleTOF 5600. Murine brain (mBMEC) and lung (mLEC) microvascular endothelial cells were isolated as described (86)and were cultured on gelatin (Sigma) in Dulbecco's Modified Eagles Medium

(DMEM; Lonza) containing 20% (v/v) fetal bovine serum (Atlanta Biologicals) heparin (100 µg/ml) and endothelial cell growth supplement (ECGS, 50 µg/ml; VWR). Confluent mBMEC and mLEC were washed with serum free DMEM, and then treated with proteinase K (25 ng/ml) for 15 min. U937 cells were cultured Roswell Park Memorial Institute 1640 medium (RPMI) containing 10% FBS (v/v). The cells were centrifuged, washed three times with PBS, and then digested with proteinase K (500 ng/ml) in PBS for 45 min with rotation. The supernatants from digestions of mBMEC, mLEC, and U937 cells underwent the same work flow for heparin purification and mass spectrometry as HUVEC.

Liquid chromatography Mass Spectrometry LC-MS-MS: Trypsin-digested peptides were analyzed by ultra-high pressure liquid chromatography (UPLC) coupled with tandem mass spectroscopy (LC-MS/MS) using nano-spray ionization. The nano-spray ionization experiments were performed using a TripleTof 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-phase UPLC (Waters corporation nano ACQUITY) using a 20 cm-75 micron ID glass capillary packed with 2.5-µm C18 (130) CSHTM beads (Waters corporation). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80%) of ACN (Acetonitrile) at a flow rate of 250 µl/min for 1h. The buffers used to create the ACN gradient were: Buffer A (98% H2O, 2% ACN, 0.1% formic acid, and 0.005% TFA) and Buffer B (100% ACN, 0.1% formic acid, and 0.005% TFA). MS/MS data were acquired in a data-dependent manner in which the MS1 data was acquired for 250 msec at m/z of 400 to 1250 Da and the MS/MS data was acquired from m/z of 50 to 2,000 Da. The Independent data acquisition (IDA) parameters were as follows; MS1-TOF acquisition time of 250 milliseconds, followed by 50 MS2 events of 48 milliseconds acquisition time for each event. The threshold to trigger MS2 event was set to 150 counts when the ion had the charge state +2, +3 and +4. The ion exclusion time was set to 4 seconds. Finally, the collected data were analyzed using Protein Pilot 4.5 (ABSCIEX) for peptide identifications.

Peptide Alignment. Identified peptides from all HUVEC proteinase K limited proteolysis mass spectrometry experiments were mapped and counted to their respective UniProt protein sequence by using the multiple sequence alignment tool Clustal Omega (89,90).

Expression and purification of CLEC14A. We synthesized mammalian codon optimized (Genewiz) human CLEC14A-290 (residues 1-290) and CLEC14A-325 (residues 1-325) DNA and cloned them into pcDNA3.1A (+) (Invitrogen) with a C-terminal His_6 tag. CLEC14A alanine mutants were constructed by site directed mutagenesis using QuickChange DNA Mutagenesis Kits (Agilent) and confirmed by sequencing (Genewiz). To produce recombinant protein, HEK293F cells (1.5-2 $x10^6$ cells/ml) were transfected 2.5 µg/ml of plasmid DNA using polyethyleneimine (PEI, 9 µg/ml) in FreeStyle 293 Expression Medium (Gibco). One day later the cells were treated with valproic acid (2 mM), and 5 days after the initial transfection, the conditioned medium was mixed with cOmplete, EDTA-free Protease Inhibitor (Roche). Recombinant protein was purified by chromatography on a 1ml Ni^{2+} Sepharose 6 Fast Flow column (GE LifeSciences). Samples were loaded with FreeStyle 293 Expression Medium supplemented with 30 mM imidazole, washed with 30 mM imidazole, 0.5 M NaCl, 20 mM Tris buffer (pH 7.4) and recombinant protein was eluted using 0.5 M NaCl, 0.3 M imidazole in 20 mM Tris buffer (pH 7.4). The protein was further purified by size exclusion chromatography (HiLoad 16/60 Superdex 200, prep grade. GE LifeSciences) in 0.3 M NaCl, 5% glycerol in 20 mM Tris buffer (pH 7.4). Mutant CLEC14A was purified in the same manner as wild-type CLEC14A. For studies involving N-linked glycans, CLEC14A was treated with (Neb) for 16 hours at 37° C under non-denaturing conditions.

CLEC14A *N*-glycan site mapping and glycopeptide analysis. Details are provided in Supplementary Information.

Heparin ELISA. Porcine mucosal heparin (SPL Scientific Protein Laboratories) was immobilized (50 µL at 1mg/ml) in 96-well Carbobind plate (Corning) in 0.1M sodium acetate buffer (pH 5.5) for 2 hr at room temperature. Wells were washed with PBST, blocked with 1%

BSA (Sigma) for 2 hr at 37 °C, and washed again with PBST. The wells were incubated with the indicated concentrations of recombinant CLEC14A at room temperature for 1 hr. Bound ligand was quantitated using THE™ His Tag antibody (Genescript) and anti-mouse HRP (Cell Signaling). The Kd value was calculated by fitting the binding data to a single-site binding model in Prism.

Endothelial heparan sulfate purification. HUVEC were grown on gelatin in EGM-2 medium until confluent. The cells were radiolabeled with $^{35}SO_4$ (20 µCi) in 5 ml of F12 medium (Gibco) supplemented with 10% fetal calf serum depleted of glycosaminoglycans (ref). The culture medium was collected after 24 hrs and the cell layer was treated with trypsin for 10 min at 37ºC. The trypsin solution was collected and centrifuged to remove cell debris. Secreted GAGs in the growth medium and cell surface GAGs were pooled and digested with 0.4 mg/ml Pronase (Sigma) overnight at 37 °C. Samples diluted with 2 volumes of wash buffer and purified by anion exchange chromatography. Columns were prepared by washing 1 ml of 50% slurry of DEAE Sepharose beads (GE LifeSciences) with 50 mM sodium acetate, 0.2 M NaCl, 0.1% Triton X, pH 6.0. After applying the sample, the columns were washed with wash buffer, and eluted with 2.5 ml of elution buffer (50 mM sodium acetate, 2 M NaCl, pH 6.0). Samples were then desalted (PD-10 column, GE LifeSciences), the GAG was eluted in 10% ethanol. Samples were lyophilized, resuspended in 50 mM Tris, 50 mM NaCl, 25 mM $MgCl_2$, pH 8. To remove DNA and chondroitin sulfate, samples were treated with 20 kUnit/ml of DNase1 (Sigma) and 20 mUnits of chondroitinase ABC (Amsbio) for 3 hrs at 37°C. To liberate the [^{35}S]heparan sulfate chains from residual peptides, the samples were β-eliminated with 0.4 M NaOH overnight at 4 °C. The [^{35}S]heparan sulfate was then re-purified by anion exchange chromatography and desalted.

Nitrocellulose Filter Binding Assay. Recombinant CLEC14A, BSA, or FGF2 was incubated with 10,000 counts of [^{35}S]heparan sulfate for 30 minutes at room temperature.

Samples were added to prewashed nitrocellulose membranes on a vacuum apparatus and rapidly filtered. The filters were added to 5 ml of Ultima Gold XR (Perkin Elmer) scintillation fluid and counted by liquid scintillation.

Surface Plasmon Resonance. A Nicoya OpenSPR was used to generate binding curves for CLEC14A binding to heparin, porcine intestinal mucosal dermatan sulfate (Celsus Laboratories), chondroitin sulfate E (Sigma), Chinese hamster ovary cell heparan sulfate (rHS01, TEGA Therapeutics, Inc.), and umbilical cord hyaluronan (Sigma). Protein was immobilized on a Nicoya carboxyl sensor using Nicoya amine coupling kit. Carboxyl sensors were functionalized using 0.2 ml of a 1:1 mix of N-hydroxysuccinimide (0.1 M) and 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC, 0.4 M) before coupling to recombinant CLEC14A under flow conditions. Ethanolamine was used to block remaining active sites on the chip. In other experiments, biotinylated heparin (Sigma) was immobilized onto a Nicoya Streptavidin Sensor chip. All surfaces were washed with SPR buffer (20 mM HEPES, 150 mM NaCl, 5mM $CaCl_2$, 17 mM NaN_3, 5 mM $MgCl_2$, 0.1% BSA, and 0.05% Tween20 pH 7.2) and regenerated with 20 mM HEPES buffer (pH 7.2) containing 3 M NaCl. Ligands were allowed to associate with the chip at a flow rate of 20 µl/min in SPR buffer for 4 min, and allowed to dissociate for 5 min. Regeneration buffer was used before each injection of ligand to clean the surface chip.

Analytical Size Exclusion Chromatography (SEC). Purified CLEC14A (150-200ug) and sized-defined heparin-derived oligosaccharides (Iduron) were incubated in 20mM Tris buffer (pH 7.5) containing 0.2 M NaCl. Complexes were resolved on a Superdex200 column (size 10/300, GE Healthcare). The column was calibrated with gel filtration standards (Bio-Rad) consisting of thyroglobulin (669 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), equine myoglobin (17 kDa), and vitamin B 12 (1.35 kDa). For accurate molecular mass determinations, CLEC14A was resolved on a Superdex200 column (size 10/300), and the eluate was passed in-line to a miniDAWN TREOS MALS detector followed by an Opti lab T-rEX refractive index detector (Wyatt Technology).

Analytical Heparin-Sepharose Chromatography. CLEC14A was applied to a 1-ml HiTrap heparin-Sepharose column (GE Healthcare) in PBS. Protein was eluted with a gradient of NaCl from 150 mM to 1 M.

Differential Scanning Fluorimetry. CLEC14A (6 μM) was incubated with 5X SYPRO Orange Protein Gel Stain (Thermo Fisher) in PBS. CLEC14A thermal denaturation was monitored on a CFx96 Real-Time PCR system (BioRad) from 25-98°C at a rate of 1°C/minute. Heparin and heparin derivatives (Iduron) were added to the solution at final concentration of 48 μM. Melting temperatures were calculated using first derivatives of the data fit assuming a Gaussian distribution and plotted in Prism.

Protein Structures and Molecular Modeling. Models of lectin domain (residues 33-173) of human CLEC14A, and the RICIN domain (residues 17-130) of a PTPRβ isoform were generated using full-chain protein structure prediction server Robetta with the RosettaCM protocol (91). CD93 (residues 37-241) models were generated using Phyre2 (92). Electropotential maps were generated from TSP1 (PDB 2OUJ), SDF1 (PDB 2NWG), LOXL2 (PDB 5ZE3), APLP2 (PDB 5TPT) crystal structures. Structures were visualized using Pymol.

Results and Discussion

Analysis of heparin-binding proteins on the surface of endothelial cells

Limited proteolysis is a powerful tool to map conformational features of proteins. By using suboptimal conditions for proteolysis (limiting enzyme, reduced temperature, and omission of reducing agents and denaturants), limited cleavage occurs at exposed hinges or loops resulting in the liberation of intact protein domains (93,94). When applied to cells, limited proteolysis can be used to isolate and purify ectodomains of cell surface transmembrane proteins and subdomains of secretory and extracellular matrix proteins (95). We hypothesized that analysis of these liberated domains by chromatography on heparin-Sepharose would enrich for HS-binding domains and potentially indicate sites of contact between the protein and the

ligand (Figure 2.1a). To establish the feasibility of the approach, we treated confluent monolayers of human umbilical vein endothelial cells (HUVEC) with varying concentrations of proteinase K and trypsin (Methods). The extent of proteolysis was monitored by SDS-PAGE and silver staining of proteolytic fragments released from human umbilical vein endothelial cells. Conditions were adjusted to shift the pattern of bands on the gel from the pattern obtained for samples treated only with buffer, but not to the extent that all of the material migrated as low molecular weight peptides. To enrich for HS-binding domains, we subjected the samples to heparin-affinity chromatography. Heparin is structurally related to HS, although it is more highly sulfated, enriched in iduronic acid, and more highly-negatively charged. Its commercial availability makes it an inexpensive surrogate for HS. Samples obtained after proteolytic digestion or in mock digestions with PBS were bound to heparin-Sepharose, and weakly bound proteins were washed out with low ionic strength buffer (0.3 M NaCl in 20 mM HEPES, pH 7.2). Strongly bound proteins were eluted with buffer containing 1 M NaCl. The eluted material was then analyzed by LC-MS/MS.

Proteomic characterization of the material displaying high-affinity for heparin yielded numerous candidate HSBPs. The protein hits were filtered based on the presence of signal peptides (membrane and secreted proteins), subcellular localization deduced via database searches and manual curation of the literature. In HUVEC, a total of 34 proteins were confidently identified in 6 independent experiments using Proteinase K or chymotrypsin for limited proteolysis (Table 2.1). They included known HSBPs, such as thrombospondin 1 (THBS1) (96-98), hedgehog-interacting protein (HHIP) (99), and vascular endothelial growth factor receptor 1 (VEGFR1) (100). Previously unknown HSBPs were also detected including C-type lectin domain family 14 member A (CLEC14A), tyrosine receptor phosphatase beta (PTPRβ), lysyl oxidase like protein 2 (LOXL2), transmembrane protein 132 (TMEM132A), amyloid precursor like protein 2 (APLP2), growth/differentiation factor 15 (GDF15), hyaluronan synthase 1 (HAS1), ectonucleoside triphosphate diphosphohydrolase 1 (ENTPD1), and the

sodium channel protein type 10 subunit alpha (SCN10A). We extended the analysis to mouse brain microvascular endothelial cells (BMEC) and mouse lung microvascular endothelial cells (MLEC), which yielded both unique HSBPs as well as proteins identified in HUVEC (Table 2.1). The method also can be used with non-adherent cells, as shown for U937 histiocytic lymphoma cells (Table 2.1). In total, 75 HSBPs were identified, including 37 HSBPs not previously known to bind to heparin. Over half of the proteins identified by LPHAMS were secreted soluble proteins or extracellular matrix proteins. Presumably, many of these proteins were present in the extracellular matrix or bound to the cell surface, given that the cells were only gently rinsed with PBS prior to limited proteolysis. Twelve Type I, two Type II , four polytopic, and two GPI-anchored membrane proteins were identified in this way (Table 2.1). Thus, LPHAMS has the capacity to identify a broad range of membrane-associated and extracellular proteins and can be applied to different cell types.

LPHAMS facilitates mapping of heparin-binding domains in HSBPs

Alignment of the peptides identified in the mass spectra to primary protein sequences in the UniProt database often mapped to specific subdomains in the known and putative HSBPs, suggesting the possibility that LPHAMS could help identify the heparin-binding sites in these proteins (Figure 2.2a). For example, peptides derived from thrombospondin-1 (THBS1) were confined to the N-terminal laminin G (LamG)-like domain where the heparin-binding site was previously mapped by heparin-affinity chromatography of proteolytic fragments, molecular docking and X-ray crystallography (Figure 2.2a) (96-98). Several of the proteins (e.g. VEGFR1 (100), HHIP (99), stromal cell derived factor 1 (SDF1, CXCL12) (101,102) showed partial alignment of the recovered peptides with the putative heparin-binding domains, and in other cases none of the recovered peptides corresponded to the location of the heparin binding site, for example in fibronectin (FN1), where peptides mapping to the documented N-terminal heparin-binding domain were not recovered. In contrast, peptides mapping to a heparin-binding

site in the C-terminal domain were detected (103). Peptides mapping to endostatin, the heparin binding domain in collagen XVIII (COL18A1) (104), and in the heparin binding domain in Annexin A2 (ANXA2) also were not recovered (105). A Disintegrin and Metalloproteinase with Thrombospondin Repeats 4 (ADAMTS4) was identified in the screen as well, consistent with the observation that the protein can interact with heparan sulfate (106,107), but the position of the binding site has not been established. Recovered peptides in secreted HSBPs aligned well with domains previously shown to bind heparin, for example in connective tissue growth factor (CTGF) (22,108), hepatic derived growth factor (HDGF) (109,110), perlecan (HSPG2) (111) and laminin alpha 4 (LAMA4) (112,113). Examination of the structure of THSB1 and SDF1, which have been co-crystalized with heparin, showed that peptide sequences retrieved by LPHAMS aligned with the heparin-binding site(97,98,101).

We next inspected peptides derived from previously unidentified HSBPs (Figure 2.2b) and examined their position in available crystal structures or in generated molecular models based on related structures to search for patches of positively charged amino acids fitting the constraints described by Xu and Esko for heparin binding domains (50) (Figure 2.3). For reference, the crystal structure for THSB1 and SDF1 is shown with the electropositive surface previously documented to bind heparin (Figure 2.3a) (97,101). Peptides for APLP2 came mainly from the E2 domain, a heparin-binding domain found in amyloid precursor protein (APP) and APLP1. Examination of the crystal structure for human E2 APLP2 dimer (PDB 5TPT) revealed a large stretch of positive charge on each monomer that aligns with the peptides obtained by LPHAMS (Figure 2.3b). LOXL2 had peptides spanning the second to fourth scavenger receptor cysteine-rich (SRCR) domain and the lysyl oxidase-like domain (Figure 2.2b). Inspection of the crystal structure (PDB 5ZE3) (114) revealed a large electropositive patch (45 x 22 Å) spanning the dimer interface of the SRCR4 domain, once again highly consistent with a putative heparin-binding site (Figure 2.3b). We modeled the R-type lectin domain of PTPR□, and the C-type lectin domains of CLEC14A and cluster of differentiation 93 (CD93) using Phyre2 and/or

Robetta (Figure 2.3c). PTPRβ consists of 1997 amino acids, and a single ricin-like fold followed by 17 fibronectin repeats. All of the peptides recovered by LPHAMS mapped specifically to the N-terminal ricin domain (Figure 2.2b). An area of positive charge spanning 20 X 20 Å was present in the model (Figure 2.3c). In CLEC14A, peptides predominantly localized to the C-Type lectin and EGF domains (Figure 2.2b). Interestingly, modeling of CLEC14A suggested a patch of positive charge stretching 10 X 30 Å, indicative of a putative heparin-binding site. Another C-type lectin, CD93 (a C-type lectin 14 family member), also had peptides in its lectin domain that aligned with a putative heparin-binding site (Figure 2.2b).

Some of the identified proteins lacked three dimensional structures or were not of sufficient homology to previously crystalized proteins to allow molecular modeling. These proteins include Transmembrane Protein 132A (T132A), ADAMTS1, and the □1 and □2 chains of Type V collagen (COL5A1 and COL5A2). Interestingly, the recovered peptide sequences in Type V collagen corresponded with thrombospondin 1 (TSPN1)/LamG domains and collagen EMF1a (COLF1). TSPN1 and LamG are protein modules known to interact with heparin (37), whereas COLF1 has not been previously associated with heparin binding.

CLEC14A binds heparan sulfate

To validate LPHAMS as a discovery tool, we analyzed the glycosaminoglycan-binding properties of CLEC14A, a member of the C-type lectin family 14. CLEC14A plays a role in physiological and pathological angiogenesis, but its identification as a heparin-binding protein and the structure and function of the heparin-binding domain has not been characterized (115,116). CLEC14A is Type I transmembrane protein containing a C-type lectin domain (CTLD), an EGF-module and an endomucin domain rich in serine and threonine residues (Figure 2.1b). A 21-amino acid transmembrane peptide connects the ectodomain to a 71-amino acid cytoplasmic tail. C-type lectins in general bind calcium, and many bind glycans through their carbohydrate recognition domain in the CTLD. CLEC14A belongs to a subgroup of C-type

lectins that include CD93, thrombomodulin, and CD248 (endosialin) (117). As shown in Table 2.1, both CLEC14A and CD93 were identified by LPHAMS.

Recombinant cDNA constructs spanning residues 1-325 (the entire ectodomain, CLEC14A-325) and residues 1-290 (the ectodomain lacking the endomucin domain, CLEC14A-290) were expressed in HEK293F cells (Figure 2.1b). Primary sequence analysis indicated the presence of a single consensus sequon (Asn-Leu-Ser) for *N*-linked glycosylation at Asn189. SDS-PAGE before and after PNGase F digestion demonstrated that CLEC14A-325 contained an asparagine-linked glycan chain (Supplemental Figure S2.1a). Enzymatic deglycosylation of CLEC-290 using PNGase F in the presence of ^{18}O-labeled water ($H_2{}^{18}O$) labeled the sites of N-glycosylation (118,119). The extent of glycosylation was estimated at ~60% based on the recovery of ^{18}O-labeled peptides and peptides containing an asparagine residue in position 189 (Materials and Supplemental Table 2.1). Glycan analysis showed that Asn189 was occupied predominantly by either a complex-type biantennary, disialylated and core fucosylated glycan ($HexNAc_4Hex_5Fuc_1NeuAc$) or a hybrid-type, core fucosylated structure ($HexNAc_3Hex_6Fuc_1NeuAc_1$) (Supplemental Figure S2.1b). Removal of the *N*-glycans did not affect the ability of recombinant CLEC14A-325 to engage heparin (Supplemental Figure S2.1c). CLEC14A contains a C-type lectin fold related to the E/L/P-Selectins, which bind to glycans containing sialyl Lewis X (120). We tested whether CLEC14A could bind to sialyl Lewis X and other classes of acidic and neutral glycans through the Consortium for Functional Glycomics Protein-Glycan Interaction Core using a glycan array covering 609 different glycan structures unrelated to GAGs. CLEC14A-325 did not bind significantly to any of the glycans in the presence or absence of calcium (Supplemental Figure S2.2). In contrast, surface plasmon resonance (SPR) confirmed that CLEC14A-325 bound to immobilized heparin (Figure 4a). Binding of CLEC14A-290, lacking the endomucin domain was reduced in comparison (Figure 4b). As shown below, CLEC14A-325 behaves as a dimer by size exclusion chromatography, whereas CLEC14A-290 migrates as a monomer, most likely explaining the difference in binding

of the two recombinant proteins to immobilized heparin (Figure 2.4a,b). To assess binding of different classes of GAGs, we immobilized recombinant CLEC14A-325 onto a SPR chip and tested different GAGs as the analyte. Under these conditions, heparin exhibited similar binding characteristics with high on and low off rates, like the results obtained when heparin was immobilized on the chip (Figure 2.4a,c). Chondroitin sulfate E, dermatan sulfate, and HS derived from Chinese hamster ovary cells (rHS01) also bound, but not as avidly as heparin (Figure 2.4d-f). No binding was detected with hyaluronan (data not shown). The lack of dissociation in the SPR experiments, presumably due to valency effects of extended polysaccharide chains, prevented calculation of values of K_D.

Binding of CLEC14A to endothelial HS was tested using ^{35}S-labeled HS as isolated from HUVEC. Samples were mixed with recombinant CLEC14A-325 and the solution was rapidly filtered through nitrocellulose. Free, uncomplexed GAGs do not bind to nitrocellulose, as demonstrated by the lack of counts bound to the filter when [^{35}S]HS was incubated with bovine serum albumin (Figure 2.5) (121). In contrast, [^{35}S]HS bound to CLEC14A-325 in a dose dependent manner (3 µg of protein bound 10 ± 4% of the input [^{35}S]HS counts) (Figure 2.5A). For comparison, 0.4 µg of FGF2, which binds to HS with high affinity (17,18,122), sequestered 35 ± 6% of input counts (Figure 2.5).

Heparin oligosaccharides bind to a single binding site in CLEC14A

Size exclusion chromatography using globular protein standards showed that CLEC14A-290 ran as a ~31 kDa monomer consistent with its predicted molecular mass of 31.6 kDa. In contrast, CLEC14A-325 migrated with an effective mass of ~85 kDa but the predicted molecular mass was only 35.6 kDa, suggesting that CLEC14-325 behaves as a dimer or trimer (Figure 2.6b). When the experiment was repeated using multiangle light scattering (MALS) to more accurately estimate molecular mass, CLEC14A-290 and CLEC14A-325 eluted with molecular masses of 35.4 +/- 2.4 kDa and 87.1 +/- 18.1 kDa, respectively. Many C-type lectins behave as

trimers to increase their avidity for multivalent ligands (123), suggesting that one function of the endomucin domain is to facilitate oligomerization of CLEC14A.

As indicated above, modeling studies predicted a 10 x 32 Å electropositive surface embedded in the CTLD of CLEC14A, which could accommodate theoretically a dp10-12 heparin oligosaccharide (Figure 2.3c). Incubation of CLEC14A-290 with dp10 heparin oligosaccharides shifted the elution pattern of the protein, increasing the apparent mass to ~38 kDa (Figure 2.6a). Incubation with heparin yielded a large complex of average relative mass of 195 kDa, suggesting a stoichiometry of 5:1 CLEC14A-290:heparin. The data also supports the idea that CLEC14A accommodates oligosaccharides of ~dp10 because heparin consists of a variety of chains of average molecular mass ~14 kDa (~dp50). Incubation of CLEC14A-325 oligomer with dp10 heparin oligosaccharides shifted its mass from 90 kDa to 114 kDa; the difference of 24 kDa is close to the predicted value if CLEC14A-325 behaves like a trimer and binds three dp10 oligosaccharides (Figure 2.6b). Together, these findings suggest that CLEC14A contains a single binding site for heparin, and that these sites act independently in the oligomers.

Binding of glycans to proteins can stabilize them against denaturation. To test the impact of heparin on CLEC14A stability, we examined the response of CLEC14A-325 to thermal denaturation using Differential Scanning Fluorimetry (DSF). In this technique, denaturation is measured by binding of a hydrophobic dye to hydrophobic residues exposed by denaturation. An increase in melting temperature induced by ligand-protein binding reflects enhanced protein stability. CLEC14A-325 showed a typical biphasic melting curve, melting at 55°C based on the first derivative spectrum. The addition of heparin increased thermostability by 10°C in a concentration dependent manner (Figure 2.7a). Individual oligosaccharides dp6-18 added at a 9:1 ratio also increased the stability of CLEC14A-325 against thermal denaturation (Figure 2.7b), with the effect reaching saturation with a dp12 oligosaccharide. Based on NMR structures for heparin oligosaccharides, a dp12 oligosaccharide has extended length of ~38-48Å (124). This length corresponds well with the size of the modeled heparin binding site shown in Figure

2.3c. Chemically desulfating heparin at *N*- or C6 positions of glucosamine units or the C2 position of uronic acids reduced its ability to stabilize CLEC14A, indicating that binding involves contacts with multiple sulfate groups (Figure 2.7c).

Genetic mapping of the heparin binding site of CLEC14A

The CLEC14A CTLD contains a large positively charged surface patch decorated with arginine and lysine residues R141, K158, R161, K165 (Figure 2.8a). To determine if these residues are involved in heparin binding, they were converted one-by-one to alanine residues i, and recombinant protein was produced in HEK293F cells. Chromatography of the recombinant proteins on heparin-Sepharose showed that R141A, K158A, and R165A mutations had little effect on the salt concentration required for elution of the mutated proteins compared to the wildtype protein (488-528 mM NaCl for mutants R141A, K158A and K165A vs. 515 mM for the wildtype) (Figure 2.8b). However, the R161A variant eluted at much lower concentration of NaCl (373 mM) indicating an impairment in its heparin-binding capacity (Figure 2.8c). We validated this finding using an ELISA in which heparin was immobilized on a plate. Wildtype CLEC14A-325 bound to immobilized heparin with an apparent affinity of ~25 nM. In contrast, the R161A variant essentially lost its capacity to bind immobilized heparin under these conditions (Figure 2.8d). Interestingly, the R161A mutant was more thermally stable compared to the wildtype protein, indicating that the decreased binding to heparin was not caused by unfolding of the mutant protein. As expected, heparin stabilized the wildtype protein to thermal denaturation, but had a much-reduced effect in the mutant (Figure 2.8e).

Conclusions

In this report we describe the development and application of LPHAMS, a proteomic workflow integrating limited proteolysis, heparin-affinity chromatography and high-resolution LC-MS/MS. Application of LPHAMS to human and murine endothelial cells led to the identification

of large number of HSBPs and in many examples the method revealed the subdomains that facilitate binding to HS. We identified an endothelial transmembrane protein CLEC14A as a novel glycosaminoglycan-binding protein. CLEC14A, a C-type lectin, most likely exists as a trimer and does not bind typical *N*-linked and *O*-linked glycans, but instead binds to glycosaminoglycans. In practice, the method is simple, does not require pre-fractionation methods or detergents, and can be applied to a variety of cell types. Type I, Type II and polytopic membrane proteins as well as extracellular matrix and secreted proteins were discovered using this method.

Many investigators have used affinity chromatography coupled with mass spectrometry to identify heparin-binding proteins, but typically the source material consisted of a body fluid such as blood, serum, or cerebral spinal fluid. This approach led to the purification of soluble growth factors, plasma proteins of the coagulation cascade and complement systems, and RNA and DNA binding proteins, but few membrane proteins were identified (17,18,20-22,96,125). To enrich for membrane proteins, a technique was devised that involved isolation of a plasma membrane fraction, for example from liver (37). Solubilization of the membranes and affinity purification of the material over heparin-Sepharose led to the identification of 148 HSBPs, including 79 membrane proteins. Although effective, this strategy involves homogenization of tissues, purification of plasma membranes, and detergents for solubilization of otherwise insoluble membrane proteins. Another approach employed cell surface biotinylation of cultured cells, enrichment by streptavidin-affinity chromatography and fractionation by heparin-Sepharose chromatography (36,37). Many HSBPs were discovered in this way, but few membrane proteins were enriched possibly because of the low abundance of membrane proteins or limited access to the tagging reagents.

LPHAMS takes advantage of suboptimal proteolysis to cleave selectively proteins at exposed protein hinges or loops, and can easily modified depending on the ultimate targets. When applied to cells, specific enrichment for protein ectodomains of cell surface proteins

occurred leading to the identification of previously unidentified heparin-binding proteins. One limitation of this method is the requirement that target proteins have accessible protease cleavage sites, but the use of a broad-spectrum serine protease (Proteinase K) minimizes this potential problem. On the other hand, the use of a broad-spectrum protease could easily lead to the under-representation of certain proteins due to undesirable extensive proteolysis, but this limitation might be avoided by varying the duration of proteolytic treatment, concentration of the protease, the type of protease and temperature. Other GAGs or glycans could be used as the affinity matrix as well.

LPHAMS also has the advantage of providing information about putative glycosaminoglycan-binding sites in the heparin-binding proteins. These predictions were often consistent with prior mapping studies in which the site of interaction was deduced by crystallography, NMR, or modeling of heparin-binding proteins. Several of the binding sites in the HSBPs identified in this study mapped to larger domains that presumably depend on folding and approximation of subdomains to generate the positively charged surfaces with affinity for heparin. Of the 75 identified proteins, 63 had at least one so-called Cardin-Weintraub sequence, [-X-B-B-X-B-X-] or [-X-B-B-B-X-X-B-X] where X is a hydrophilic amino acid and B is a basic amino acid. However, these Cardin-Weintraub sequences were not necessarily predictive, because they mapped outside of the established heparin-binding site or domains not exposed at the cell surface. In some known heparin-binding proteins, we did not find the relevant domains expected from prior studies of the heparin-binding sites. For example, peptides from the heparin binding domain in COL18A1 and ANXA2 were not recovered, possibly due to proteolytic fragmentation. Conceivably, some of the proteins identified by LPHAMS might not actually bind to heparin, but instead form a complex with a bona fide heparin-binding protein which then led to its copurification. However, these interactions are of interest as well because they define possible complexes that warrant further studies to elucidate their biological function.

In this study, we show for the first time that CLEC14A binds to GAGs, and not carbohydrates that typically associate with C-type lectins. CLEC14A is most likely a trimeric protein, typical of many C-type lectins. Although oligomerization can increase the avidity of lectins for multivalent ligands, binding to heparin oligosaccharides appeared to occur independently in each monomer. Interestingly, the heparin binding site in CLEC14A does not map to the site typically associated with the carbohydrate recognition domain of this lectin subfamily (117), and binding does not depend on calcium. These observations suggest that the ability to bind heparin and other GAGs evolved independently of the C-type lectin fold. *CLEC14A* is an endothelial-specific gene up-regulated during tumor angiogenesis and regulates endothelial cell migration and adhesion in vitro and angiogenesis in vivo (115,116,126-132). The C-type lectin domain has been shown to engage other matrix proteins, such as multimerin 2 (MMRN2) and heat shock protein 70-1A (126,127). Antibodies blocking these interactions or targeting the C-type lectin domain decreased cell migration and tumor angiogenesis in a MMRN2 (116,127) or a VEGFA dependent fashion (133). Whether these antibodies block the GAG binding site is not known.

CLEC14A bound to specific glycosaminoglycans with fast on-rates and undetectable off-rates. Heparin and chondroitin sulfate E have high charge density and bound more avidly compared to heparan sulfate, chondroitin sulfate A and dermatan sulfate which have lower charge density (Fig. 4). There was no detectable binding to hyaluronan. CLEC14A also binds to heparan sulfate from HUVEC (Figure 2.5). These findings suggest that CLEC14A prefers highly charged polysaccharides, consistent with dramatically reduced binding resulting from mutation of specific arginine residues in the putative GAG binding site (Fig. 8) or desulfation of heparin (Fig. 7). Although it is possible that CLEC14A prefers a specific arrangement of sulfated residues in the ligand, we think its most likely that the overall charge determines the affinity of the interaction, which would suggest that the GAG binding site is somewhat promiscuous.

In summary, we describe a facile method for discovery of glycosaminoglycan-binding proteins using limited proteolysis. The technique can be readily adapted to other cell types by altering proteolytic conditions and in theory can be used to identify proteins that interact with other carbohydrate ligands by variation of the affinity matrix. The technique also aids in mapping the ligand binding site. Finally, the characterization of CLEC14A as a heparin-binding protein validates the technology and suggests further studies of CLEC14A function in vivo.

Acknowledgements

We thank Dr. David Smith (Emory University) for analysis of CLEC14A binding to the CFG glycan array. This work was supported in part by the Program of Excellence in Glycoscience P01 HL107150 and grant P01 HL131474 from the NIH (to J.D.E.), a supplement to P01 HL131474 (to D.R.S.), R01 AR070179 (to D.X.), R01 GM104141 (to K.D.C.), and P41 GM103490 (to L.W.)

Chapter 2, in full has been submitted for publication in the Journal for the American Chemical Socitety. Sandoval, D.R., Toledo, A.G., Painter C.D., Tota, E., Sheikh, O., West, A.M.V., Wells, L., Bicknell, R., Corbett K.D., Xu, D., Esko, J.D. I am the primary investigator and author of this paper.

Table 2.1: Heparin-Binding Proteins Identified by LPHAMS

Gene Name	Protein Identification	Coverage (percent)	Spectral Count	Cell Type	Location	Heparin Binding
CLEC14A	C-type lectin domain family 14 member A	30.6	8	HUVEC	Type I Transmembrane	?
TMEM132A	Transmembrane 132A	10.9	4	HUVEC	Type I Transmembrane	?
APLP2	Amyloid-like protein 2	7.5	6	HUVEC	Type I Transmembrane	?
PTPRB	Receptor-type tyrosine-protein phosphatase beta	5.1	4	HUVEC	Type I Transmembrane	?
Flt1	Vascular endothelial growth factor receptor 1	9.8	5	HUVEC/BMEC	Type I Transmembrane	Yes (100)
STIM	Stromal interaction molecule 1	1	1.6	MLEC	Type I Transmembrane	?
NRG1	Neuregulin-1	5.2	1	MLEC	Type I Transmembrane	Yes (134-136)
CLEC2B	C-type lectin 2	19.5	3	U937	Type I transmembrane	?
GPNMB	Transmembrane glycoprotein NMB	19.4	9	U937	Type I transmembrane	Yes (137)
BSG	Basigin	10.8	2	U937	Type I transmembrane	?
CD93	Complement component C1q receptor	10.1	2	U937	Type I transmembrane	?
HLA-DPB1	Major histocompatibility complex, class I, B	3.9	1	U937	Type I Transmembrane	?
MTDH	Protein LYRIC	11.2	3	MLEC	Type II Transmembrane	?
SLC3A2	4F2 cell-surface antigen heavy chain	6.4	1	U937	Type II transmembrane	?
HAS1	hyaluronan synthase 1	4.3	1	HUVEC	Polytopic Membrane	?
ENTPD1	Ectonucleoside triphosphate diphosphohydrolase 1	1.3	1	HUVEC	Polytopic Membrane	?
SCN10A	Sodium channel protein type 10 subunit alpha	1.1	2	HUVEC	Polytopic Membrane	?
KCNAB2	Voltage-gated potassium channel subunit beta-2	10.8	2	U937	Polytopic Membrane	?

Table 2.1: Heparin-Binding Proteins Identified by LPHAMS, continued.

Gene Name	Protein Identification	Coverage (percent)	Spectral Count	Cell Type	Location	Heparin Binding
FN1	Fibronectin	49.6	122	HUVEC	Extracellular matrix	Yes (138)
MMRN2	Multimerin-2	15.2	10	HUVEC	Extracellular matrix	?
COL5A2	Collagen alpha-2(V) chain	9.5	5	HUVEC	Extracellular matrix	?
COL8A1	Collagen alpha-1(VIII) chain	7	2	HUVEC	Extracellular matrix	?
MMRN1	Multimerin 1	6.8	11	HUVEC	Extracellular matrix	?
LAMB2	Laminin subunit beta-2	4.7	1	HUVEC	Extracellular matrix	?
LAMA4	Laminin subunit alpha-4	10	8	HUVEC/BMEC	Extracellular matrix	Yes (112,139)
COL18A1	Collagen alpha-1(XVIII) chain	9.3	7	HUVEC/BMEC	Extracellular matrix	Yes (104)
CYR61	Protein CYR61	33	14	HUVEC/BMEC/MLEC	Extracellular matrix	Yes (140,141)
COL5A1	Collagen alpha-1(V) chain	6.9	3	HUVEC/MLEC	Extracellular matrix	?
HSPG2	Perlecan	23.9	58	HUVEC/MLEC/BMEC	Extracellular matrix	Yes (111)
COL14A1	Collagen alpha-1(XIV) chain	1.2	1	MLEC	Extracellular matrix	Yes (142)
VTN	Vitronectin	5.9	1	U937	Extracellular matrix	Yes (143)
HHIP	Hedgehog-interacting protein	51.5	48	HUVEC/MLEC	GPI Anchored	Yes (99)
GPC4	Glypican 4	8.4	3	MLEC/BMEC	GPI Anchored	?
VWA1PDGFD	Platelet-derived growth factor, D polypeptide	8.4	2	BMEC	Secreted	?
F2	Prothrombin	6.3	1	BMEC	Secreted	Yes
IGFBP7	Insulin-like growth factor binding protein 7	5.7	1	BMEC	Secreted	Yes (144)
Serpinc1	Antithrombin-III	5.2	2	BMEC	Secreted	Yes
IGFBP3	Insulin-like growth factor binding protein 3	3.4	7	BMEC	Secreted	Yes (145-147)

Table 2.1: Heparin-Binding Proteins Identified by LPHAMS

Gene Name	Protein Identification	Coverage (percent)	Spectral Count	Cell Type	Location	Heparin Binding
ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2	3	2	BMEC	Secreted	?
Prl7c1	Prolactin-7C1	2.8	1	BMEC	Secreted	?
C3	Complement C3	1.9	1	BMEC	Secreted	Yes
SERPINE1	Plasminogen activator inhibitor 1	52.5	24	HUVEC	Secreted	Yes (148)
LOXL2	Lysyl oxidase homolog 2	20.9	10	HUVEC	Secreted	?
GDF15	Growth/differentiation factor 15	16.6	4	HUVEC	Secreted	?
ADAMTS1	A disintegrin and metalloproteinase with thrombospondin motifs 1	9.1	5	HUVEC	Secreted	Yes (149)
SULF2	Sulfatase 2	5.6	2	HUVEC	Secreted	Yes (150)
THSD4	Thrombospondin type-1 domain-containing protein 4	2	2.5	HUVEC	Secreted	?
ANXA2	Annexin A2	16.5	6	HUVEC/BMEC	Secreted	Yes (105)
ADAMTS4	A disintegrin and metalloproteinase with thrombospondin motifs 4	6.3	1	HUVEC/BMEC	Secreted	Yes (106,107)
APP	Amyloid beta	2.9	1	HUVEC/BMEC	Secreted	Yes (151)
CTGF	Connective tissue growth factor	45.9	62	HUVEC/BMEC/MLEC	Secreted	Yes (108)
THBS1	Thrombospondin 1	35.7	60	HUVEC/BMEC/MLEC	Secreted	Yes (152)
ALB	Albumin	14	9	HUVEC/BMEC/MLEC	Secreted	No
HDGF	Hepatoma-derived growth factor	32.5	15	HUVEC/MLEC/U937	Secreted	Yes (109)
MANF	Mesencephalic astrocyte-derived neurotrophic factor	45.4	10	HUVEC/U937/MLEC	Secreted	?
SDF1	Stromal cell-derived factor 1	34.4	1	MLEC	Secreted	Yes (102)

Table 2.1: Heparin-Binding Proteins Identified by LPHAMS

Gene Name	Protein Identification	Coverage (percent)	Spectral Count	Cell Type	Location	Heparin Binding
Serpinh1	Serpin H1	6	1	MLEC	Secreted	?
CLCF1	Cardiotrophin-like cytokine factor 1	3.6	1	MLEC	Secreted	?
ELANE	Neutrophil elastase	68.9	81	U937	Secreted	Yes (153)
CTSG	Cathepsin G	62.8	56	U937	Secreted	Yes (154,155)
AZU1	Azurocidin	58.2	14	U937	Secreted	Yes (156)
LGALS7	Galectin 7	44.9	4	U937	Secreted	?
FABP5	Fatty acid-binding protein 5	40.7	3	U937	Secreted	?
C1QBP	Complement component 1 Q subcomponent-binding protein	29.1	8	U937	Secreted	?
B2M	Beta-2-microglobulin	26.5	2	U937	Secreted	Yes (157)
ENO1	Enolase 1	16.6	2	U937	Secreted	?
HSP90B1	Endoplasmin	15.6	11	U937	Secreted	Yes (158-160)
CTSD	Cathepsin D	9.6	1	U937	Secreted	Yes
ARHGAP45	Rho GTPase-activating protein 45	8.5	1	U937	Secreted	?
GRN	Granulins	7.1	2	U937	Secreted	Yes (161)
ANXA6	Annexin A6	5.6	1	U937	Secreted	Yes (162)
FIBP	Acidic fibroblast growth factor intracellular-binding protein	16.5	1	U937	Intracellular	?
CTSL	Cathepsin L1	8.9	2	U937	Lysosome	?

Figures

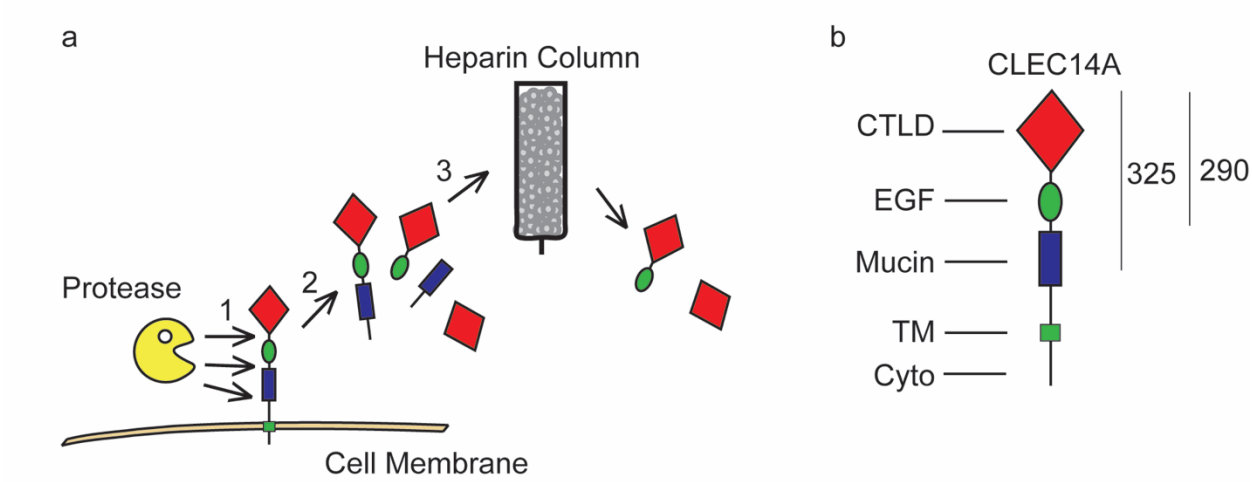


Figure 2.1. Isolation of heparin binding proteins by LPHAMS. a, Schematic representation of limited proteolysis screen. In this example, a protease cuts CLEC14A at exposed regions separating it into domains. These domains are passed through a heparin-Sepharose column, and those domains that contain a heparin-binding site bind more avidly to the column, leading to their enrichment. **b**, Subdomain structure of CLEC14A depicting the C-type lectin domain (CTLD), epidermal growth factor (EGF), mucin-like (MUCIN), transmembrane (TM) and cytoplasmic (CYTO) domains.

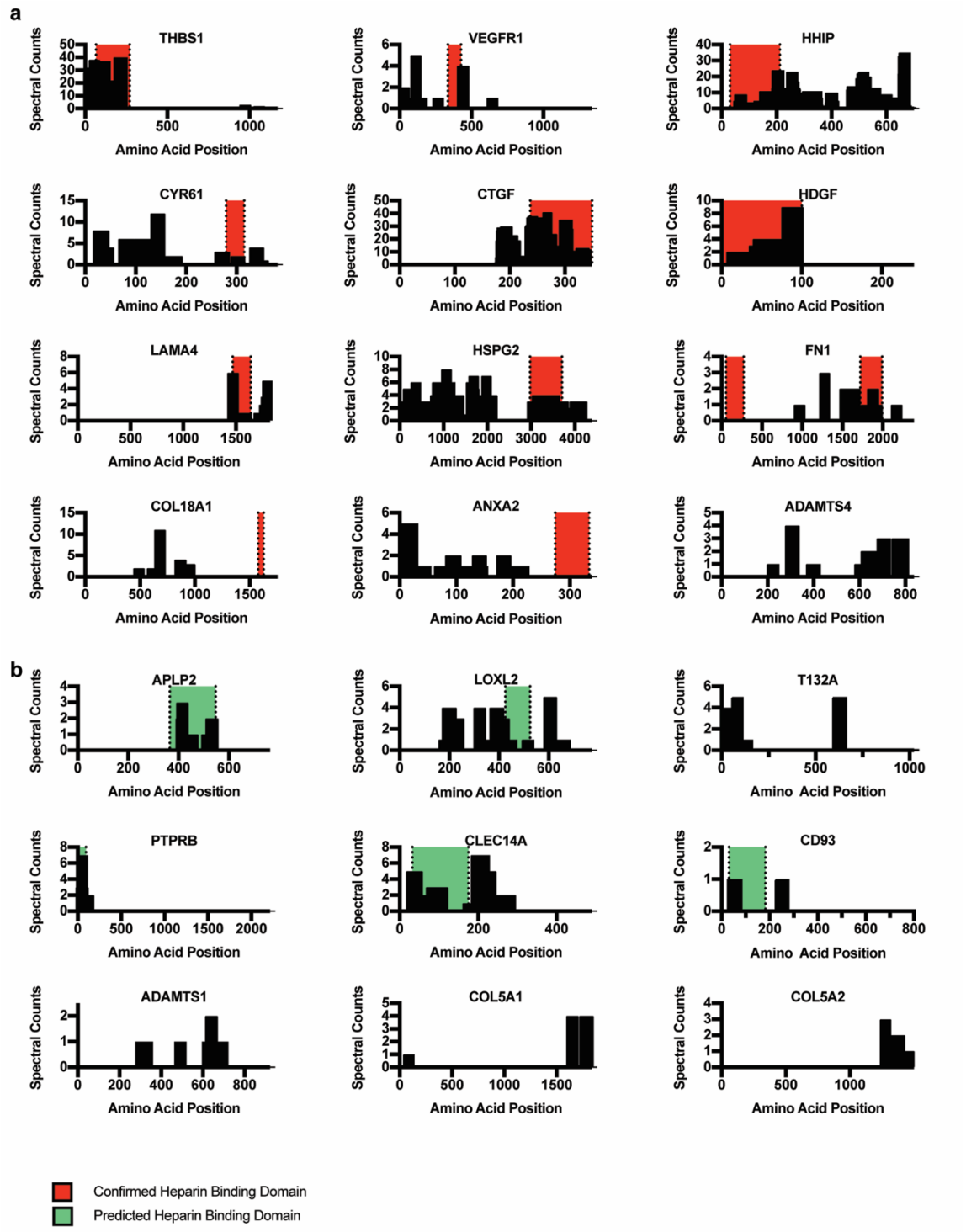


Figure 2.2. Limited proteolysis selects for ectodomains of cell surface proteins and maps to putative heparin binding sites. a,b, Alignment and mass spectral counts of peptides detected by LPHAMS (black domains). Red domains represent documented heparin-binding sites, whereas green domains represent putative heparin-binding sites based on LPHAMS and modeling studies. **a**, Peptide alignments of known heparin-binding proteins. **b**, Peptide alignments of previously unknown heparin-binding proteins.

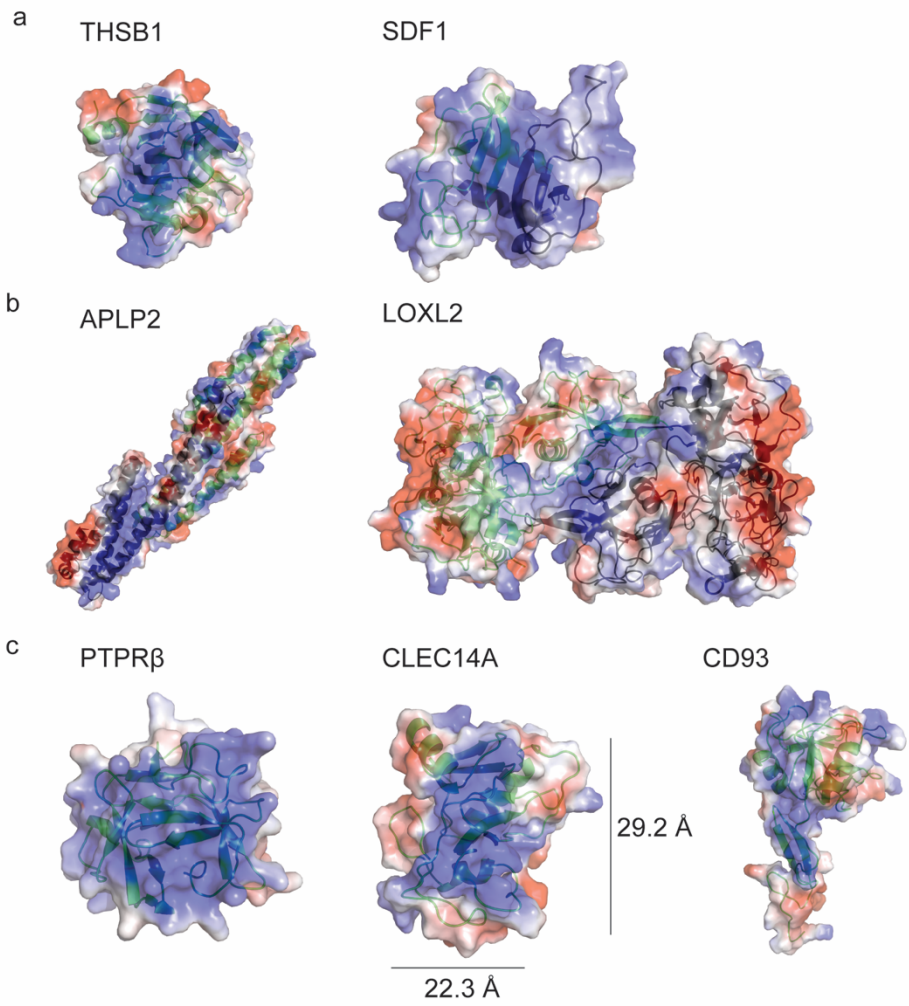


Figure 2.3. Electropotential plots and ribbon diagrams of either crystallized or modeled structures of selected candidate HSBPs. Blue represents regions of positive charge, red negative charge. **a,** Reported 3-diemnsional structures of known heparin-binding proteins THSB1 (PDB 2OUJ) and SDF1 (PDB 2NWG). **b,** Reported crystal structures of previously unknown heparin-binding proteins APLP2 (PDB 5TPT) and LOXL2 (PDB 5ZE3). **c,** Structural models of heparin-enriched protein domains of PTPR□, CLEC14A and CD93.

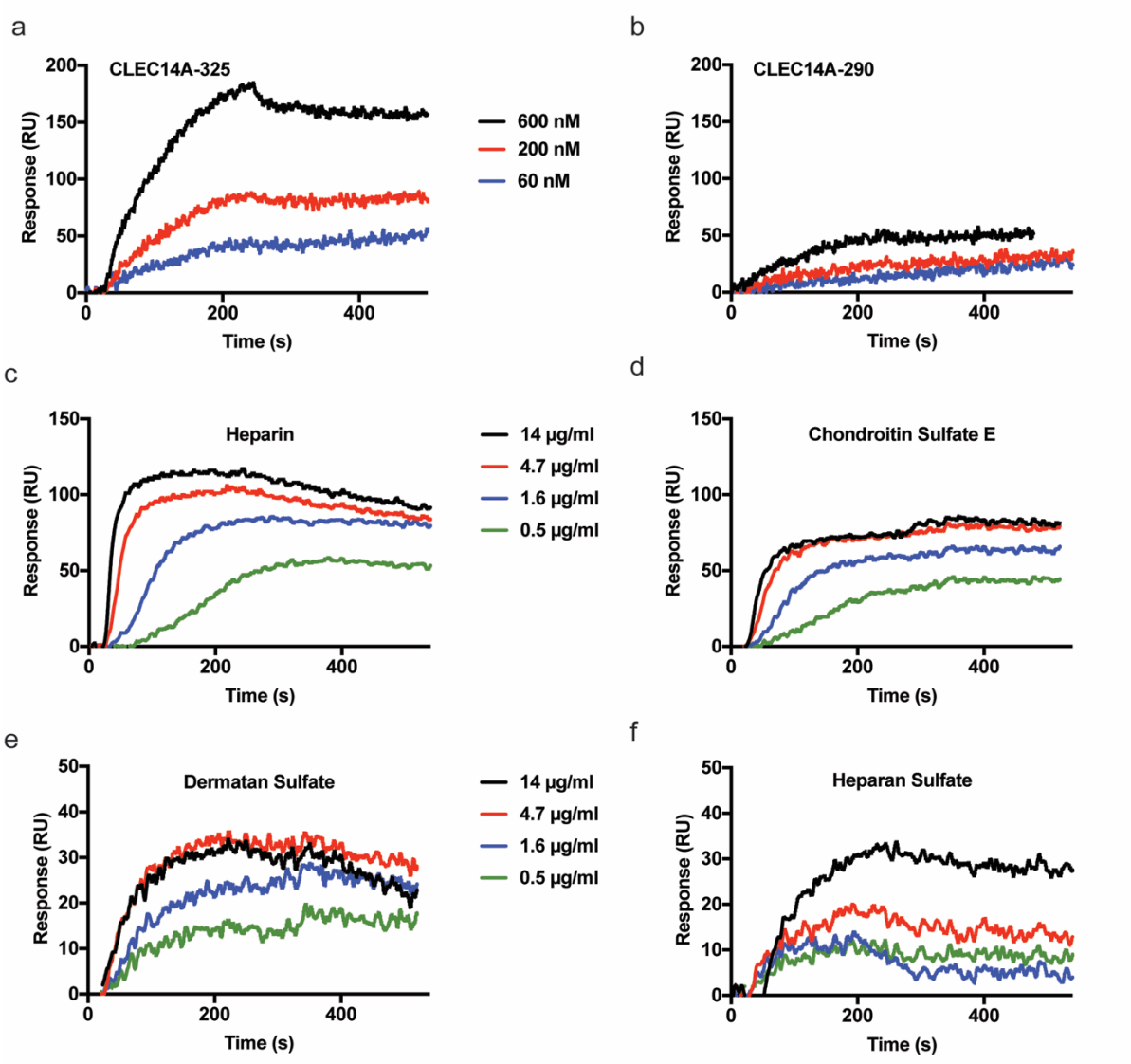


Figure 2.4. CLEC14A binds to several glycosaminoglycans measured by surface plasmon resonance. a,b, CLEC14A-325 and CLEC14A-290 binding to immobilized heparin. **c,d,e,f,** Surface plasmon resonance curves demonstrating binding of GAGs to immobilized CLEC14A-325. **c**, heparin; **d**, squid chondroitin sulfate E; **e**, dermatan sulfate; **f**, CHO heparan sulfate.

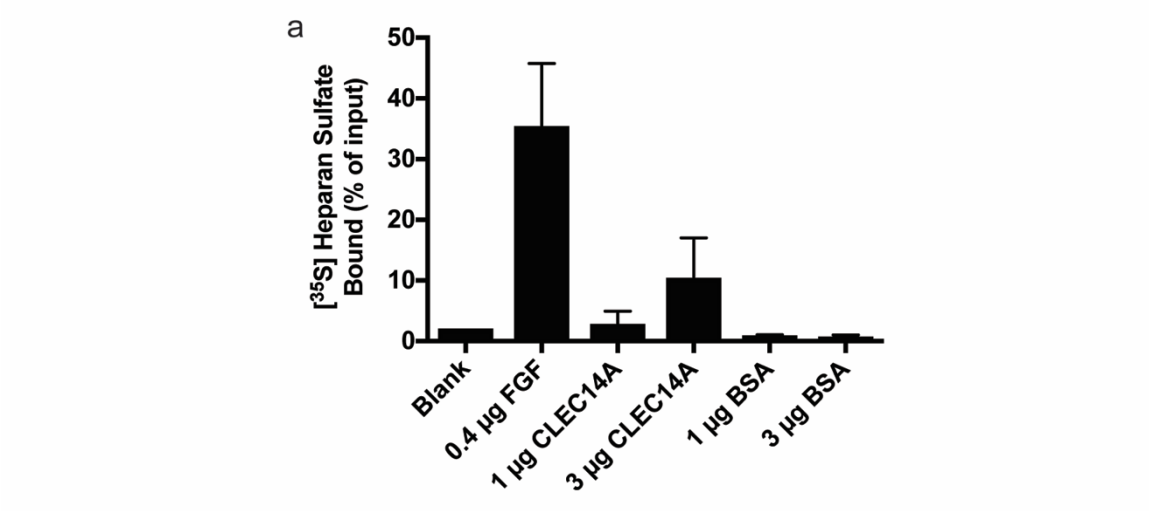


Figure 2.5. CLEC14A binds to recombinant heparan sulfate and to endothelial cells. **a,** Filter binding assay of HUVEC [^{35}S] heparan sulfate to FGF2, CLEC14A-325, and BSA. The assay was performed in triplicate and the average values ± standard deviation were determined.

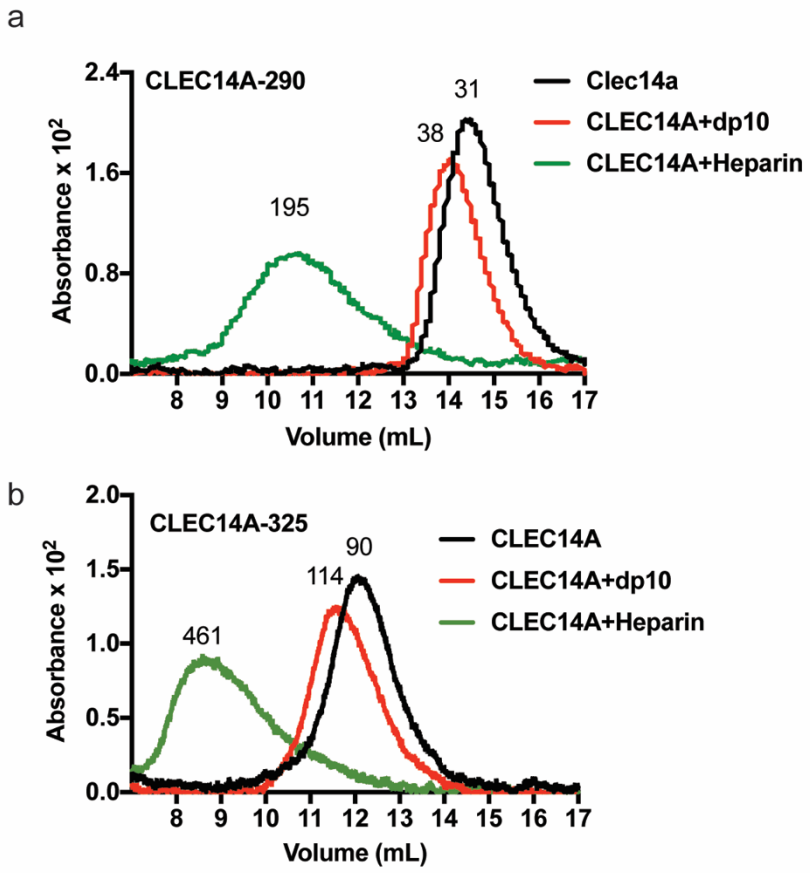


Figure 2.6. CLEC14A binds to heparin oligosaccharides. **a**, Recombinant CLEC14A-290 and b, CLEC14A-325 were analyzed by size exclusion chromatography in the absence or presence of dp10 heparin oligosaccharides or heparin. CLEC14A-290 behaves as a monomer, whereas CLEC14A-325 behaves as an oligomer, most likely a trimer. Both recombinant proteins bind heparin oligosaccharides, without affecting oligomerization.

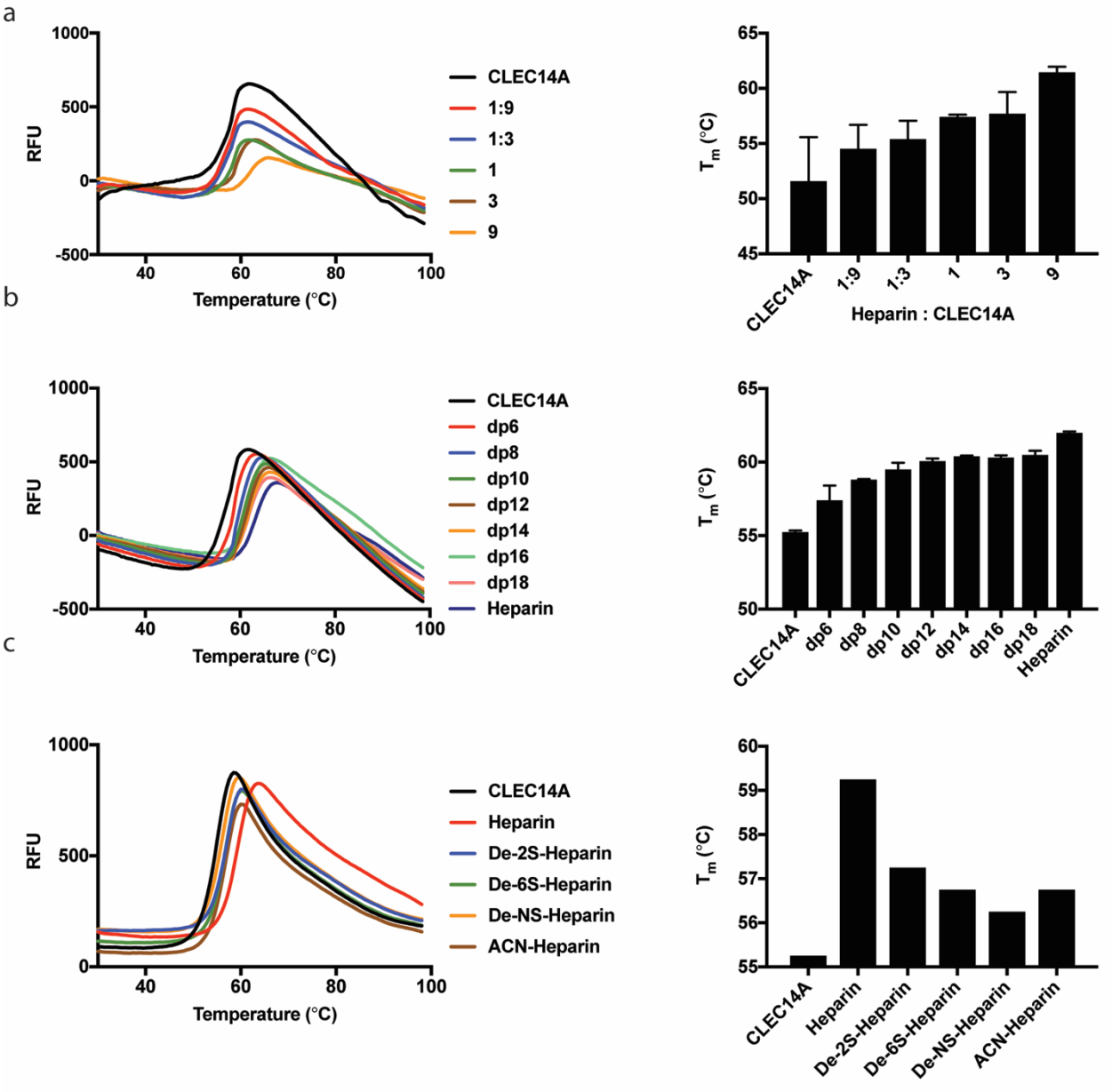


Figure 2.7. Heparin stabilizes CLEC14A to thermal denaturation. CLEC14A-325 was thermally denatured and the melting point was determined by Differential Scanning Fluorimetry (DSF). **a**, CLEC14A-325 (6 μM) was incubated with various concentrations of heparin to achieve the indicated molar ratios. **b**, Heparin oligosaccharides (dp6-18) were incubated with CLEC-325 at a 9:1 molar ratio and analyzed by DSF. **c**, CLEC14A-325 was incubated with chemically de-sulfated forms of heparin. Each condition was performed in triplicate and the data was analyzed by Prism (V5.0).

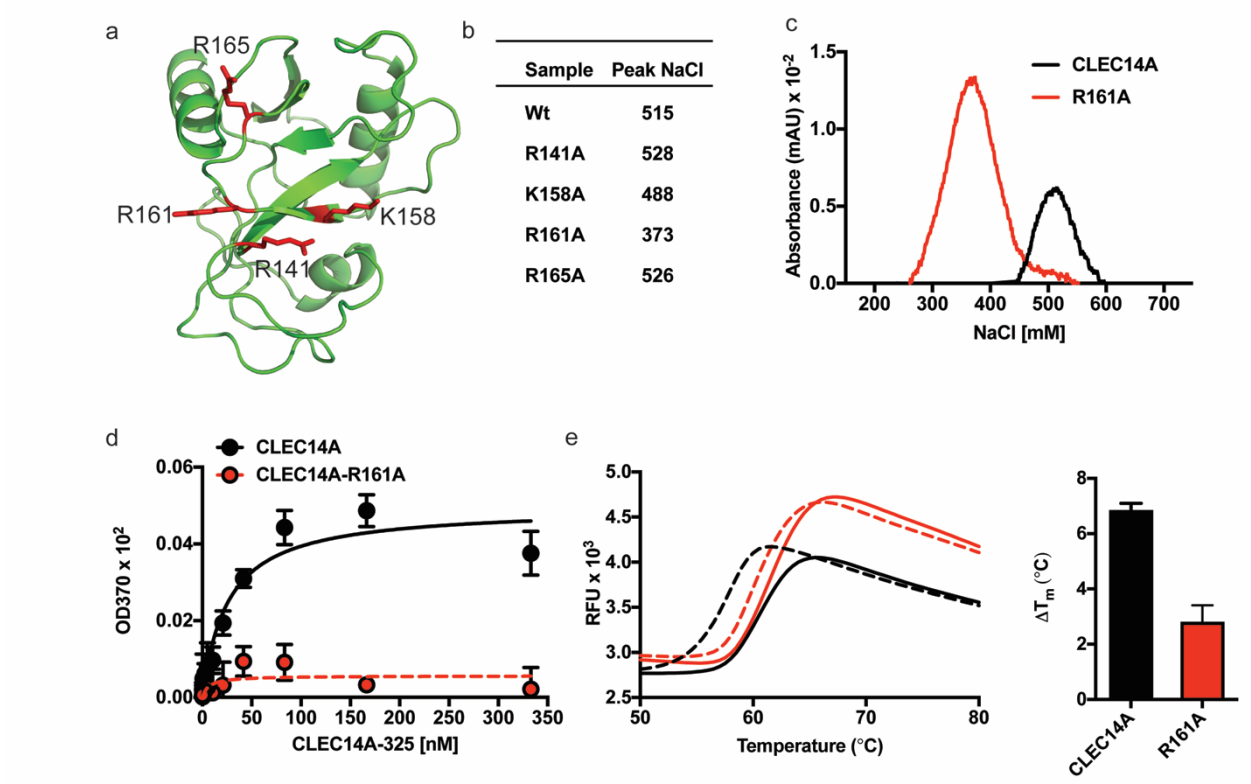
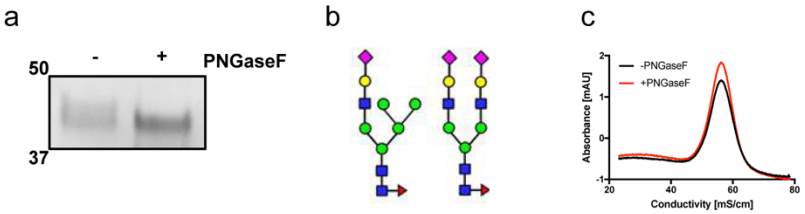
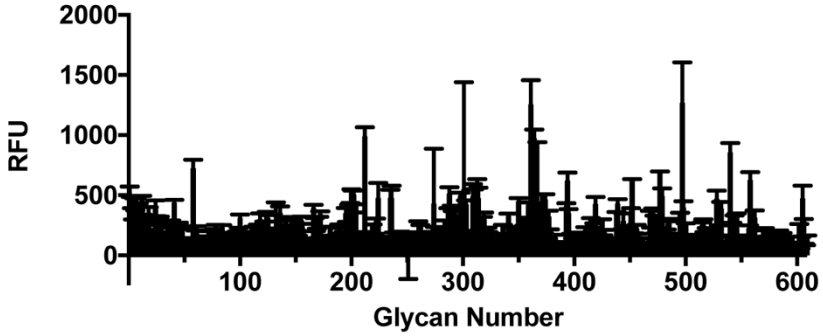


Figure 2.8. Mapping the heparin binding domain in CLEC14A. a, Schematic representation of the CLEC14A lectin domain (CTLD, robetta predictions). Arginine and lysine residues are shown as red stick representations on ribbon diagram. **b,** Elution of CLEC14A-325 mutants on heparin-Sepharose. Wild-type CLEC14A-325 and the indicated mutants were chromatographed on heparin-Sepharose and the concentration of NaCl required for elution was determined. **c,** Heparin-Sepharose chromatography of wild-type CLEC14A-325 and mutant R161A. **d,** Binding of CLEC14A-325 and mutant R161A to immobilized heparin. **e,** DSF of CLEC14A-325 and mutant R161A in the presence and absence of heparin. **f,** ΔT_m (°C) of wild-type and mutant CLEC14A induced by heparin.



Supplemental Figure 2.1. *N*-linked glycosylation of CLEC14A at Asn189 is dispensable for heparin-binding. a, Western blot of recombinant CLEC14A produced in HEK293-F cells treated with and without PNGase F. **b**, Predominant *N*-linked glycan structures attached to recombinant CLEC14A determined by mass spectrometry. **c**, Heparin affinity chromatography of CLEC14A before and after treatment with.



Supplemental Figure 2.2. Binding of CLEC14A to the CFG glycan array. CLEC14A-325 was used to probe the CFG glycan microarray of 609 different glycans. No significant binding was observed.

References

1. Xu, D., and Esko, J. D. (2014) Demystifying heparan sulfate-protein interactions. *Annu Rev Biochem* **83**, 129-157

2. Bjork, I., and Lindahl, U. (1982) Mechanism of the anticoagulant action of heparin. *Mol Cell Biochem* **48**, 161-182

3. Chiodelli, P., Bugatti, A., Urbinati, C., and Rusnati, M. (2015) Heparin/Heparan Sulfate Proteoglycans Glycomic Interactome in Angiogenesis: Biological Implications and Therapeutical Use. *Molecules (Basel, Switzerland)* **20**, 6342-6388

4. van Wijk, X. M., and van Kuppevelt, T. H. (2014) Heparan sulfate in angiogenesis: a target for therapy. *Angiogenesis* **17**, 443-462

5. Fuster, M. M., Wang, L., Castagnola, J., Sikora, L., Reddi, K., Lee, P. H., Radek, K. A., Schuksz, M., Bishop, J. R., Gallo, R. L., Sriramarao, P., and Esko, J. D. (2007) Genetic alteration of endothelial heparan sulfate selectively inhibits tumor angiogenesis. *J Cell Biol* **177**, 539-549

6. Wang, L., Fuster, M., Sriramarao, 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

7. Axelsson, J., Xu, D., Kang, B. N., Nussbacher, J. K., Handel, T. M., Ley, K., Sriramarao, P., and Esko, J. D. (2012) Inactivation of heparan sulfate 2-O-sulfotransferase accentuates neutrophil infiltration during acute inflammation in mice. *Blood* **120**, 1742-1751

8. Ori, A., Wilkinson, M. C., and Fernig, D. G. (2011) A systems biology approach for the investigation of the heparin/heparan sulfate interactome. *J Biol Chem* **286**, 19892-19904

9. Xu, D., Young, J., Song, D., and Esko, J. D. (2011) Heparan sulfate is essential for high mobility group protein 1 (HMGB1) signaling by the receptor for advanced glycation end products (RAGE). *J Biol Chem* **286**, 41736-41744

10. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539

11. Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145

12. Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B. H., Das, R., Grishin, N. V., and Baker, D. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **77 Suppl 9**, 89-99

13. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858

14. Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta* **1382**, 191-206

15. Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis. *Acta biochimica Polonica* **51**, 299-321

16. Olaya-Abril, A., Jimenez-Munguia, I., Gomez-Gascon, L., and Rodriguez-Ortega, M. J. (2014) Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. *J Proteomics* **97**, 164-176

17. Dixit, V. M., Grant, G. A., Santoro, S. A., and Frazier, W. A. (1984) Isolation and characterization of a heparin-binding domain from the amino terminus of platelet thrombospondin. *J Biol Chem* **259**, 10100-10105

18. Tan, K., Duquette, M., Liu, J. H., Zhang, R., Joachimiak, A., Wang, J. H., and Lawler, J. (2006) The structures of the thrombospondin-1 N-terminal domain and its complex with a synthetic pentameric heparin. *Structure (London, England : 1993)* **14**, 33-42

19. Tan, K., Duquette, M., Liu, J. H., Shanmugasundaram, K., Joachimiak, A., Gallagher, J. T., Rigby, A. C., Wang, J. H., and Lawler, J. (2008) Heparin-induced cis- and trans-dimerization modes of the thrombospondin-1 N-terminal domain. *J Biol Chem* **283**, 3932-3941

20. Holtz, A. M., Griffiths, S. C., Davis, S. J., Bishop, B., Siebold, C., and Allen, B. L. (2015) Secreted HHIP1 interacts with heparan sulfate and regulates Hedgehog ligand localization and function. *J Cell Biol* **209**, 739-758

21. Park, M., and Lee, S. T. (1999) The fourth immunoglobulin-like loop in the extracellular domain of FLT-1, a VEGF receptor, includes a major heparin-binding site. *Biochem Biophys Res Commun* **264**, 730-734

22. 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-1 alpha) binding to heparin. *J Biol Chem* **282**, 10018-10027

23. Bleul, C. C., Farzan, M., Choe, H., Parolin, C., Clark-Lewis, I., Sodroski, J., and Springer, T. A. (1996) The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829-833

24. Benecky, M. J., Kolvenbach, C. G., Amrani, D. L., and Mosesson, M. W. (1988) Evidence that binding to the carboxyl-terminal heparin-binding domain (Hep II) dominates the interaction between plasma fibronectin and heparin. *Biochemistry* **27**, 7565-7571

25. O'Reilly, M. S., Boehm, T., Shing, Y., Fukai, N., Vasios, G., Lane, W. S., Flynn, E., Birkhead, J. R., Olsen, B. R., and Folkman, J. (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* **88**, 277-285

26. Shao, C., Zhang, F., Kemp, M. M., Linhardt, R. J., Waisman, D. M., Head, J. F., and Seaton, B. A. (2006) Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J Biol Chem* **281**, 31689-31695

27. Kashiwagi, M., Enghild, J. J., Gendron, C., Hughes, C., Caterson, B., Itoh, Y., and Nagase, H. (2004) Altered proteolytic activities of ADAMTS-4 expressed by C-terminal processing. *J Biol Chem* **279**, 10109-10119

28. Fushimi, K., Troeberg, L., Nakamura, H., Lim, N. H., and Nagase, H. (2008) Functional differences of the catalytic and non-catalytic domains in human ADAMTS-4 and ADAMTS-5 in aggrecanolytic activity. *J Biol Chem* **283**, 6706-6716

29. Ball, D. K., Rachfal, A. W., Kemper, S. A., and Brigstock, D. R. (2003) The heparin-binding 10 kDa fragment of connective tissue growth factor (CTGF) containing module 4 alone stimulates cell adhesion. *J Endocrinol* **176**, R1-7

30. Brigstock, D. R., Steffen, C. L., Kim, G. Y., Vegunta, R. K., Diehl, J. R., and Harding, P. A. (1997) Purification and characterization of novel heparin-binding growth factors in uterine secretory fluids. Identification as heparin-regulated Mr 10,000 forms of connective tissue growth factor. *J Biol Chem* **272**, 20275-20282

31. Nakamura, H., Izumoto, Y., Kambe, H., Kuroda, T., Mori, T., Kawamura, K., Yamamoto, H., and Kishimoto, T. (1994) Molecular cloning of complementary DNA for a novel human hepatoma-derived growth factor. Its homology with high mobility group-1 protein. *J Biol Chem* **269**, 25143-25149

32. Sue, S. C., Chen, J. Y., Lee, S. C., Wu, W. G., and Huang, T. H. (2004) Solution structure and heparin interaction of human hepatoma-derived growth factor. *J Mol Biol* **343**, 1365-1377

33. Brown, J. C., Sasaki, T., Gohring, W., Yamada, Y., and Timpl, R. (1997) The C-terminal domain V of perlecan promotes beta1 integrin-mediated cell adhesion, binds heparin, nidogen and fibulin-2 and can be modified by glycosaminoglycans. *Eur J Biochem* **250**, 39-46

34. Yamaguchi, H., Yamashita, H., Mori, H., Okazaki, I., Nomizu, M., Beck, K., and Kitagawa, Y. (2000) High and low affinity heparin-binding sites in the G domain of the mouse laminin alpha 4 chain. *J Biol Chem* **275**, 29458-29465

35. Yamashita, H., Beck, K., and Kitagawa, Y. (2004) Heparin binds to the laminin alpha4 chain LG4 domain at a site different from that found for other laminins. *J Mol Biol* **335**, 1145-1149

36. Zhang, X., Wang, Q., Wu, J., Wang, J., Shi, Y., and Liu, M. (2018) Crystal structure of human lysyl oxidase-like 2 (hLOXL2) in a precursor state. *Proc Natl Acad Sci U S A* **115**, 3828-3833

37. Lee, S., Rho, S. S., Park, H., Park, J. A., Kim, J., Lee, I. K., Koh, G. Y., Mochizuki, N., Kim, Y. M., and Kwon, Y. G. (2017) Carbohydrate-binding protein CLEC14A regulates VEGFR-2- and VEGFR-3-dependent signals during angiogenesis and lymphangiogenesis. *J Clin Invest* **127**, 457-471

38. Noy, P. J., Lodhia, P., Khan, K., Zhuang, X., Ward, D. G., Verissimo, A. R., Bacon, A., and Bicknell, R. (2015) Blocking CLEC14A-MMRN2 binding inhibits sprouting angiogenesis and tumour growth. *Oncogene* **34**, 5821-5831

39. Zelensky, A. N., and Gready, J. E. (2005) The C-type lectin-like domain superfamily. *The FEBS journal* **272**, 6179-6217

40. Yu, W. H., Zhao, P., Draghi, M., Arevalo, C., Karsten, C. B., Suscovich, T. J., Gunn, B., Streeck, H., Brass, A. L., Tiemeyer, M., Seaman, M., Mascola, J. R., Wells, L., Lauffenburger, D. A., and Alter, G. (2018) Exploiting glycan topography for computational design of Env glycoprotein antigenicity. *PLoS Comput Biol* **14**, e1006093

41. Angel, P. M., Lim, J. M., Wells, L., Bergmann, C., and Orlando, R. (2007) A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Commun Mass Spectrom* **21**, 674-682

42. Foxall, C., Watson, S. R., Dowbenko, D., Fennie, C., Lasky, L. A., Kiso, M., Hasegawa, A., Asa, D., and Brandley, B. K. (1992) The three members of the selectin receptor family recognize a common carbohydrate epitope, the sialyl Lewis(x) oligosaccharide. *J Cell Biol* **117**, 895-902

43. Kreuger, J., Lindahl, U., and Jemth, P. (2003) Nitrocellulose filter binding to assess binding of glycosaminoglycans to proteins. *Methods Enzymol* **363**, 327-339

44. Maciag, T., Mehlman, T., Friesel, R., and Schreiber, A. B. (1984) Heparin binds endothelial cell growth factor, the principal endothelial cell mitogen in bovine brain. *Science* **225**, 932-935

45. Shing, Y., Folkman, J., Sullivan, R., Butterfield, C., Murray, J., and Klagsbrun, M. (1984) Heparin affinity: purification of a tumor-derived capillary endothelial cell growth factor. *Science* **223**, 1296-1299

46. Mohammadi, M., Olsen, S. K., and Ibrahimi, O. A. (2005) Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev* **16**, 107-137

47. Cummings, R. D., and McEver, R. P. (2015) C-Type Lectins. in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. eds.), Cold Spring Harbor Laboratory Press

48. Khan, S., Gor, J., Mulloy, B., and Perkins, S. J. (2010) Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. *J Mol Biol* **395**, 504-521

49. Burgess, W. H., Shaheen, A. M., Ravera, M., Jaye, M., Donohue, P. J., and Winkles, J. A. (1990) Possible dissociation of the heparin-binding and mitogenic activities of heparin-binding (acidic fibroblast) growth factor-1 from its receptor-binding activities by site-directed mutagenesis of a single lysine residue. *J Cell Biol* **111**, 2129-2138

50. Hauschka, P. V., Mavrakos, A. E., Iafrati, M. D., Doleman, S. E., and Klagsbrun, M. (1986) Growth factors in bone matrix. Isolation of multiple types by affinity chromatography on heparin-Sepharose. *J Biol Chem* **261**, 12665-12674

51. Klagsbrun, M., and Shing, Y. (1985) Heparin affinity of anionic and cationic capillary endothelial cell growth factors: analysis of hypothalamus-derived growth factors and fibroblast growth factors. *Proc Natl Acad Sci U S A* **82**, 805-809

52. Jang, J., Kim, M. R., Kim, T. K., Lee, W. R., Kim, J. H., Heo, K., and Lee, S. (2017) CLEC14a-HSP70-1A interaction regulates HSP70-1A-induced angiogenesis. *Sci Rep* **7**, 10666

53. Khan, K. A., Naylor, A. J., Khan, A., Noy, P. J., Mambretti, M., Lodhia, P., Athwal, J., Korzystka, A., Buckley, C. D., Willcox, B. E., Mohammed, F., and Bicknell, R. (2017) Multimerin-2 is a ligand for group 14 family C-type lectins CLEC14A, CD93 and CD248 spanning the endothelial pericyte interface. *Oncogene* **36**, 6097-6108

54. Noy, P. J., Swain, R. K., Khan, K., Lodhia, P., and Bicknell, R. (2016) Sprouting angiogenesis is regulated by shedding of the C-type lectin family 14, member A (CLEC14A) ectodomain, catalyzed by rhomboid-like 2 protein (RHBDL2). *FASEB J* **30**, 2311-2323

55. Zanivan, S., Maione, F., Hein, M. Y., Hernandez-Fernaud, J. R., Ostasiewicz, P., Giraudo, E., and Mann, M. (2013) SILAC-based proteomics of human primary endothelial cell morphogenesis unveils tumor angiogenic markers. *Mol Cell Proteomics* **12**, 3599-3611

56. Mura, M., Swain, R. K., Zhuang, X., Vorschmitt, H., Reynolds, G., Durant, S., Beesley, J. F., Herbert, J. M., Sheldon, H., Andre, M., Sanderson, S., Glen, K., Luu, N. T., McGettrick, H. M., Antczak, P., Falciani, F., Nash, G. B., Nagy, Z. S., and Bicknell, R. (2012) Identification and angiogenic role of the novel tumor endothelial marker CLEC14A. *Oncogene* **31**, 293-305

57. Zhuang, X., Cross, D., Heath, V. L., and Bicknell, R. (2011) Shear stress, tip cells and regulators of endothelial migration. *Biochem Soc Trans* **39**, 1571-1575

58. Rho, S. S., Choi, H. J., Min, J. K., Lee, H. W., Park, H., Park, H., Kim, Y. M., and Kwon, Y. G. (2011) Clec14a is specifically expressed in endothelial cells and mediates cell to cell adhesion. *Biochem Biophys Res Commun* **404**, 103-108

59. Kim, T. K., Park, C. S., Jang, J., Kim, M. R., Na, H. J., Lee, K., Kim, H. J., Heo, K., Yoo, B. C., Kim, Y. M., Lee, J. W., Kim, S. J., Kim, E. S., Kim, D. Y., Cha, K., Lee, T. G., and Lee, S. (2018) Inhibition of VEGF-dependent angiogenesis and tumor angiogenesis by an optimized antibody targeting CLEC14a. *Molecular oncology* **12**, 356-372

60. Moscoso, L. M., Chu, G. C., Gautam, M., Noakes, P. G., Merlie, J. P., and Sanes, J. R. (1995) Synapse-associated expression of an acetylcholine receptor-inducing protein, ARIA/heregulin, and its putative receptors, ErbB2 and ErbB3, in developing mammalian muscle. *Dev Biol* **172**, 158-169

61. Loeb, J. A., and Fischbach, G. D. (1995) ARIA can be released from extracellular matrix through cleavage of a heparin-binding domain. *J Cell Biol* **130**, 127-135

62. Li, Q., and Loeb, J. A. (2001) Neuregulin-heparan-sulfate proteoglycan interactions produce sustained erbB receptor activation required for the induction of acetylcholine receptors in muscle. *J Biol Chem* **276**, 38068-38075

63. Shikano, S., Bonkobara, M., Zukas, P. K., and Ariizumi, K. (2001) Molecular cloning of a dendritic cell-associated transmembrane protein, DC-HIL, that promotes RGD-dependent adhesion of endothelial cells through recognition of heparan sulfate proteoglycans. *J Biol Chem* **276**, 8125-8134

64. Yamada, K. M., Kennedy, D. W., Kimata, K., and Pratt, R. M. (1980) Characterization of fibronectin interactions with glycosaminoglycans and identification of active proteolytic fragments. *J Biol Chem* **255**, 6055-6063

65. Talts, J. F., Sasaki, T., Miosge, N., Gohring, W., Mann, K., Mayne, R., and Timpl, R. (2000) Structural and functional analysis of the recombinant G domain of the laminin alpha4 chain and its proteolytic processing in tissues. *J Biol Chem* **275**, 35192-35199

66. Yang, G. P., and Lau, L. F. (1991) Cyr61, product of a growth factor-inducible immediate early gene, is associated with the extracellular matrix and the cell surface. *Cell Growth Differ* **2**, 351-357

67. Chen, N., Chen, C. C., and Lau, L. F. (2000) Adhesion of human skin fibroblasts to Cyr61 is mediated through integrin alpha 6beta 1 and cell surface heparan sulfate proteoglycans. *J Biol Chem* **275**, 24953-24961

68. Brown, J. C., Mann, K., Wiedemann, H., and Timpl, R. (1993) Structure and binding properties of collagen type XIV isolated from human placenta. *J Cell Biol* **120**, 557-567

69. Wojciechowski, K., Chang, C. H., and Hocking, D. C. (2004) Expression, production, and characterization of full-length vitronectin in Escherichia coli. *Protein Expr Purif* **36**, 131-138

70. Jiang, M. S., Yang, X., Esposito, D., Nelson, E., Yuan, J., Hopkins, R. F., Broadt, T., Xiao, Z., Colantonio, S., Prieto, D. A., Welch, A. R., Creekmore, S. P., Mitra, G., and Zhu, J. (2015) Mammalian cell transient expression, non-affinity purification, and characterization of human recombinant IGFBP7, an IGF-1 targeting therapeutic protein. *Int Immunopharmacol* **29**, 476-487

71. Arai, T., Parker, A., Busby, W., Jr., and Clemmons, D. R. (1994) Heparin, heparan sulfate, and dermatan sulfate regulate formation of the insulin-like growth factor-I and insulin-like growth factor-binding protein complexes. *J Biol Chem* **269**, 20388-20393

72. Hodgkinson, S. C., Napier, J. R., Spencer, G. S., and Bass, J. J. (1994) Glycosaminoglycan binding characteristics of the insulin-like growth factor-binding proteins. *J Mol Endocrinol* **13**, 105-112

73. Tressel, T. J., Tatsuno, G. P., Spratt, K., and Sommer, A. (1991) Purification and characterization of human recombinant insulin-like growth factor binding protein 3

expressed in Chinese hamster ovary cells. *Biochem Biophys Res Commun* **178**, 625-633

74. Ehrlich, H. J., Keijer, J., Preissner, K. T., Gebbink, R. K., and Pannekoek, H. (1991) Functional interaction of plasminogen activator inhibitor type 1 (PAI-1) and heparin. *Biochemistry* **30**, 1021-1028

75. Kuno, K., Kanada, N., Nakashima, E., Fujiki, F., Ichimura, F., and Matsushima, K. (1997) Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene. *J Biol Chem* **272**, 556-562

76. Morimoto-Tomita, M., Uchimura, K., Werb, Z., Hemmerich, S., and Rosen, S. D. (2002) Cloning and characterization of two extracellular heparin-degrading endosulfatases in mice and humans. *J Biol Chem* **277**, 49175-49185

77. Dahms, S. O., Hoefgen, S., Roeser, D., Schlott, B., Guhrs, K. H., and Than, M. E. (2010) Structure and biochemical analysis of the heparin-induced E1 dimer of the amyloid precursor protein. *Proc Natl Acad Sci U S A* **107**, 5381-5386

78. Lawler, J. W., Slayter, H. S., and Coligan, J. E. (1978) Isolation and characterization of a high molecular weight glycoprotein from human blood platelets. *J Biol Chem* **253**, 8609-8616

79. Spencer, J. L., Stone, P. J., and Nugent, M. A. (2006) New insights into the inhibition of human neutrophil elastase by heparin. *Biochemistry* **45**, 9104-9120

80. Kainulainen, V., Wang, H., Schick, C., and Bernfield, M. (1998) Syndecans, heparan sulfate proteoglycans, maintain the proteolytic balance of acute wound fluids. *J Biol Chem* **273**, 11563-11569

81. Ledoux, D., Merciris, D., Barritault, D., and Caruelle, J. P. (2003) Heparin-like dextran derivatives as well as glycosaminoglycans inhibit the enzymatic activity of human cathepsin G. *FEBS Lett* **537**, 23-29

82. Flodgaard, H., Ostergaard, E., Bayne, S., Svendsen, A., Thomsen, J., Engels, M., and Wollmer, A. (1991) Covalent structure of two novel neutrophile leucocyte-derived proteins of porcine and human origin. Neutrophile elastase homologues with strong monocyte and fibroblast chemotactic activities. *Eur J Biochem* **197**, 535-547

83. Ohashi, K., Kisilevsky, R., and Yanagishita, M. (2002) Affinity binding of glycosaminoglycans with beta(2)-microglobulin. *Nephron* **90**, 158-168

84. Riera, M., Roher, N., Miro, F., Gil, C., Trujillo, R., Aguilera, J., Plana, M., and Itarte, E. (1999) Association of protein kinase CK2 with eukaryotic translation initiation factor eIF-2 and with grp94/endoplasmin. *Mol Cell Biochem* **191**, 97-104

85. Menoret, A., and Bell, G. (2000) Purification of multiple heat shock proteins from a single tumor sample. *J Immunol Methods* **237**, 119-130

86. Reed, R. C., Zheng, T., and Nicchitta, C. V. (2002) GRP94-associated enzymatic activities. Resolution by chromatographic fractionation. *J Biol Chem* **277**, 25082-25089

87. Zhou, J., Gao, G., Crabb, J. W., and Serrero, G. (1993) Purification of an autocrine growth factor homologous with mouse epithelin precursor from a highly tumorigenic cell line. *J Biol Chem* **268**, 10863-10869

88. 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**, 9935-9941

CHAPTER 3: C-TYPE LECTIN XIV FAMILY

C-Type Lectins

CLEC14A belongs to the C-Type Lectin superfamily. C-Type lectins are a class of proteins with broad physiological functions that have a characteristic C-Type lectin-like domains (CTLDs) (117). These proteins occur as membrane bound or secreted proteins as monomers or oligomers. C-Type lectins were first described to distinguish proteins that bound to carbohydrates in a Ca^{2+} dependent manner, with a characteristic fold defining the CTLD. The conserved fold consists of a compact module composed of a double loop structure with two alpha-helixes (red) flanking two pairs of beta sheets (Figure 3.1). Anti-parallel beta sheets close the loop structure at the N- and C-termini. The second loop (blue) extends from the core compact domain, exiting and entering at the same region where highly conserved cysteine residues form disulfide bridges at the base of the loop. The flexible second loop is the region predominantly responsible for binding to Ca^{2+}. From the structural studies of rat mannose binding protein A (MBP-A), two Ca2+ binding amino acid motifs emerged, “EPN” and “WND” (163). In the core of the CTLD, two alpha-helices are positioned in the periphery sandwiching the beta-sheet. It should be noted that not all CTLDs bind Ca^{2+} or carbohydrates and therefore “C-type Lectin” may be a misnomer in many of the family members.

C-type Lectin 14A Family

C-Type Lectin family members are subclassified based on their overall domain structure (117). Of the 17 families of C-Type Lectins, CLEC14A falls into the C-type Lectin XIV group comprised of thrombomodulin (TM), endosialin (CD248), and complement component C1q receptor (CD93) (117). These proteins are type-one transmembrane proteins and are structurally characterized by a N-terminal CTLD, followed by an epidermal growth factor (EGF) domain(s), a mucin-like region rich in serine and threonine, a transmembrane domain and a

short cytoplasmic tail (117) (Figure 3.2). All of the C-type Lectin XIV group members are expressed in the vasculature.

TM is an anticoagulant factor expressed at the surface of endothelial cells. Mice deficient in endothelial TM succumb to a spontaneous and lethal hypercoagulation and thrombosis (164). TM binds to thrombin preventing thrombin from converting fibrinogen into fibrin and shifting thrombin's catalytic activity to activate protein C (APC) to further activate an anti-coagulation response (165,166).

CD248 was first identified as a tumor stromal antigen (167-169), and was later shown to be expressed by pericytes to affect tumor growth and blood vessel maturation (170,171). CD248 on pericytes interacts with endothelial cells undergoing angiogenesis to promote vessel destabilization and regression (171). CD248 has been shown to bind to proteins such as Galectin-3-binding protein (LGALS3BP), fibronectin (FN), collagen Types I and IV, and multimerin-2 (MMRN2) (127,172,173). It is unknown if CD248 binds to carbohydrates.

CD93 is expressed by a variety of cells including neutrophils, monocytes, microglia and endothelial cells. It is upregulated in glioma and other tumor vessels (174-177), and high CD93 expression in glioma is correlated with poor survival (175). CD93 localizes to the filopodia of endothelial tip cells where it is thought to control cell-cell and cell-matrix adhesion (175,176,178). In conjunction with MMRN2, CD93 modulates fibronectin fibrillogenesis, promotes β_1 integrin signaling during tumor angiogenesis, and controls blood vessel architecture (175,179). It is also unknown whether CD93 interacts with carbohydrates.

CLEC14A is an endothelial cell surface protein highly increased in expression during tumor angiogenesis. CLEC14A has been shown to interact with heparin and chondroitin sulfate through the CTLD domain (Chapter 2), MMRN2, and HSP90-1A (126,127). Human CLEC14A is 490 amino acids with 67% sequence identity to mouse CLEC14A and does not contain Ca^{2+} binding motifs. While it is clear that CLEC14A regulates components of the angiogenic process, it is not clear how it preforms this function. This review focuses on CLEC14A and provides a

model for how CLEC14A and other family members may participate in tumor angiogenesis.

CLEC14A - A Call to Arms

CLEC14A is a gene found expressed during angiogenesis (129,130,180,181), including differentiation of endothelial progenitor cells and during tumor angiogenesis (130,181). A proteomic study found CLEC14A to be one of the most abundant proteins expressed during Matrigel-induced endothelial morphogenesis (129).

A combination of qPCR, western blotting, immunofluorescence and *in situ* hybridization has demonstrated CLEC14A expression primarily in endothelial cells during vascular development and tumor angiogenesis, but to a lesser extend in endothelial cells in adult tissues (130,132). Expression of CLEC14A appears to be restricted to endothelial cells of the developing vasculature and lymphatic system (115,130). Histological analysis of CLEC14A on tissue arrays consisting of ten human carcinomas with adjacent normal tissue, detected the presence and increased expression of CLEC14A in ovarian, liver, bladder, prostate, breast, kidney, pancreatic, stomach and esophageal tumors and low to no expression in adjacent healthy tissue (130). Collectively, these studies indicate that CLEC14A plays a role in angiogenesis and in particular pathological angiogenesis that takes place in the tumor microenvironment.

Localization of CLEC14A

In HUVEC, CLEC14A is found at the cell surface, at cell-cell contacts, and the leading edge of migrating cells (130,132). CLEC14A may also undergo shedding into the extracellular matrix by the protease rhomboid-like 2 protein (RHBDL2) (128,129). Moreover, added soluble CLEC14A localizes to the endothelial tips of developing blood vessels and interferes with migration, but whether binding occurs in a homotypic fashion to endogenous CLEC14A or to another ligand is not known (128). When overexpressed, CLEC14A induces the formation of

filopodia in human embryonic kidney 293 (HEK293) cells (130), and induces cell aggregation of Chinese hamster ovary (CHO) cells suggesting CLEC14A expression correlates with cell-cell contacts possibly through homotypic interactions. Inhibition of CLEC14A in HUVEC, by either siRNA knockdown (130,132) or CLEC14A blocking antibodies (116,130,182), reduces cell migration, filopodia formation, and endothelial morphogenesis, suggesting CLEC14A as a positive regulator of endothelial cell migration.

CLEC14A in tumor biology

The abundant expression of CLEC14A RNA and protein in the tumor vasculature suggested that that CLEC14A might be a biomarker of pathological angiogenesis. Studies of mouse (115,116) and zebrafish (130,183) models have confirmed this idea.

Zebrafish were the first used to examine the role of CLEC14A in vasculature development. Morpholino knockdown of zebrafish Clec14a homologue (zclec14a) disrupted intersegmental vessel (ISV) and dorsal longitudinal anastomotic vessel formation (Figure 3.3a) (130). The disrupted vessels were then rescued by treating zebrafish embryos with human CLEC14A cDNA, confirming the specificity of the silencing morpholino (130). When both clec14a and the zebrafish orthologue for CD93 were simultaneously genetically inactivated, intersomitic vessel formation was severely disrupted, suggesting that both CLEC14A and CD93 participate in angiogenesis in zebrafish (183).

Clec14a$^{-/-}$mice were generated by replacing the *clec14a* coding sequence with a *lacZ* reporter (116) *Clec14A*$^{-/-}$ mice displayed reduced angiogenic sprouting in an aortic ring assay and reduced invasion into FGF2-soaked subcutaneous polyether sponges (116). When the mice were challenged with Lewis Lung Carcinoma (LLC), tumor growth was drastically reduced as well. Upon inspection of the tumors, there was a remarkable reduction in vascular density and pericyte coverage suggesting that tumor growth was reduced due to loss of CLEC14A-dependent angiogenesis (Figure 3.3b, c).

Lee *et al* generated a different *Clec14a* deficient mouse model, termed CL14A-KO. These mice exhibited increased development of nonfunctional hemorrhage-prone vessels (115). CLEC14A deficiency led to increased angiogenesis in aortic ring outgrowth assays and Matrigel plug assays. Increased angiogenesis and hemorrhage also occurred in a model of oxygen-induced retinopathy (115). Reductions in tumor growth were also noted, but instead of observing a reduction in angiogenic sprouting, an increase in angiogenesis and hemorrhage was noted (115). Developmental effects were observed in E13.5 embryos, characterized by vascular dilation and hemorrhage in the dorsal and ventral areas of the hindbrain as well as increased vascular density in the yolk sacs and hindbrain. At E15.5, CL14A-KO mice exhibited increased lymphangiogenesis in the forelimbs. Despite these developmental vascular defects, there was no observable effect on survival.

It is not clear why the two CLEC14A mouse models showed differential effects on angiogenesis. While one mouse had an overall decrease in angiogenesis and the other an increase, the vessels in each model were noted to be non-functional. Both lines were backcrossed into C57BL/6N mice. The differences may be explained in part by experimental design. For aortic ring assays, Noy *et al.* used adult mice whereas Lee *et al.* used aortas from neonates. Noy et al injected LLC subcutaneously into the right flank of mice 8-10 weeks of age whereas Lee et al. injected LLC into the abdominal region of mice 6-7 weeks of age. The authors note “further detail study is required to elucidate the reasons for these differing results.”

VEGFR3 Signaling

Angiogenesis is regulated by fine tuning of vascular endothelial growth factor receptor R2/R3 (VEGFR2/3) signaling in the leading edge of endothelial tip cells (184-186). CLEC14A forms part of the VEGFR signaling complex (115) based on co-immunoprecipitation of either CLEC14A or VEGFR3 from human umbilical cord venous endothelial cells (HUVEC). Removal of CLEC14A led to a decrease in MMRN2 expression, and Notch and VEGFR3 signaling, while

increasing VEGFR2 signaling. Overexpression of CLEC14A had the opposite effect (115). CLEC14A deficiencies also caused defects in retinal angiogenesis (115). CL14A-KO mice exhibited increased blood vessel densities and reduced pericyte coverage in the retina consistent with the idea that CLEC14A acts to regulate vessel integrity. CL14A-KO mice bearing tumors had decreased survival correlated with hemorrhagic shock due to increased tumor angiogenesis and tumor leakiness. Moreover, blockade of VEGFR2 signaling during tumor growth using chemical inhibitors or VEGFR2 monoclonal antibodies rescued CLEC14A deficiencies in tumor models (115). Monoclonal antibodies targeting the CLEC14A-CTLD block VEGF-dependent angiogenesis (133). Thus, the levels of CLEC14A may regulate blood vessel stability by fine tuning VEGFR2/VEGFR3 signaling in endothelial cells and control blood vessel maturation by regulating endothelial-pericyte interactions.

CLEC14A-Multimerin 2 interactions

While it is clear that CLEC14A promotes endothelial migration and angiogenesis, its mechanism of action at the molecular level remains unclear. The protein has no known intrinsic kinase/phosphatase activity or obvious interacting partners. Zanivan *et al*. used an anti-CLEC14A polyclonal antibody to immunoprecipitate CLEC14A from HUVEC undergoing Matrigel-induced morphogenesis. Mass spectrometry analysis of the immunoprecipitated material revealed that Multimerin-2 (MMRN2) bound to CLEC14A. MMRN2 is a pan-endothelial extracellular matrix protein found in normal and tumor vasculature (187). MMRN2 consists of 949 amino acids and contains an Emilin (EMI) domain, a coiled-coiled region, an arginine-rich region, and a C-terminal C1q domain. MMRN2 is a dynamically regulated during angiogenesis and maintenance of blood vessels and is thought to titer/sequester VEGFA isoforms (188,189).

In an attempt to probe the functional relevance of the CLEC14A-MMRN2 interaction, Noy P.J., *et. al*. created a series of monoclonal antibodies targeting CLEC14A. One antibody, C4, blocked CLEC14A-MMRN2 interaction (116). Tumor bearing mice treated with the C4

antibody displayed reduced tumor burden and angiogenesis, demonstrating the importance of CLEC14A-MMRN2 interactions in tumor angiogenesis (116). Using a series of CLEC14A monoclonal antibodies and site directed mutagenesis, it was later determined that the loop-loop domain in the CTLD bound to the MMRN2 coiled-coiled region (AA 530-624) (Figures 3.4 and 3.5) (127). In the same study, other C-Type Lectin 14A family members were tested for their ability to bind MMRN2. Three of the four family members, CLEC14A, CD93 and CD248 bound to MMRN2 using the CTLD loop-loop region (127). However, the sites of binding differed amongst the C-Type Lectins. CD248, a pericyte protein, bound to the N-terminal portion of the MMRN2 coiled-coiled domain (AA 133-486), whereas CLEC14A and CD93, bound to the C-terminal portion of the coiled-coiled domain (AA 530-624) (127). A new model was proposed wherein CLEC14A-MMRN2-CD248 drives endothelial-pericyte interactions to mediate angiogenesis and MMRN2 acts as a glue to stick endothelial cells to pericytes (127).

CLEC14A-HSP70-1A Interactions

Heat Shock protein 70-1A (HSP70-1A) also interacts with the CTLD domain of CLEC14A (126). this interaction was discovered while producing human recombinant CLEC14A-FC fusion protein in HEK293F cells. Mass spectrometry identified the contaminant as HSP70-1A (126). HSP70-1A is a heat shock protein expressed at low to near undetectable levels in the cytoplasm in unstressed cells, but it is rapidly upregulated and translocated to the cell surface when cells are exposed to stress such as hypoxia or nutrient deprivation (190). Cancer cells are chronically exposed to harsh conditions and up-regulate HSP70-1A (191-194), which is associated with tumor progression, malignancy, metastasis, and poor survival (195-201). Recently, it was shown that HSP70-1A may accumulate at the surfaces of endothelial cells and stimulate endothelial migration and tube formation through ERK phosphorylation but the upstream receptors are unknown (202). *In vivo*, recombinant HSP70-1A drastically increases microvessel formation in Matrigel plugs in mice (202).

To understand how CLEC14A-HSP70-1A protein interactions affect endothelial biology, it was first established that HSP70-1A co-immunoprecipitates with CLEC14A in endothelial cells, and that endothelial HSP70-1A cell surface binding was CLEC14A-dependent. ELISA and biolayer interferometry based assays showed that HSP70-1A bound to CLEC14A with a Kd = 8 nM, with near undetectable off rates (126). HSP70-1A binding was further characterized and mapped to amino acids 43-69 in CLEC14A by using a series of fragmented CTLD CLEC14A-FC constructs (Figure 3.4 and 3.5) (126). Jang J *et al* proposed a model in which HSP70-1A controls endothelial morphogenesis and ERK phosphorylation by promoting CLEC14A mediated endothelial cell-cell contacts and thus angiogenesis (126).

CLEC14A binds heparin

CLEC14A was first identified to bind heparin in a screen designed to identify endothelial membrane anchored heparan sulfate binding proteins (Chapter 2). The CLEC14A-CTLD bound to heparin with a Kd = 23 nM via a single binding site. Site directed mutagenesis of CLEC14A at Arg161 (R161A) to alanine greatly reduced heparin binding (Figures 3.4 and 3.5). Biophysical studies showed that CLEC14A exists as a trimer (Figure 3.4), and heparin acts to stabilize CLEC14A in a size and sulfation dependent manner. CLEC14A can physically interact with other glycosaminoglycans including endothelial heparan sulfate and chondroitin sulfate E, but not with neutral or sialylated oligosaccharides (Chapter 2). However, the manner in which heparin regulates CLEC14A in biological settings remains unknown.

Identifying Novel CLEC14A protein binding partners

To better understand the network of proteins that interact with CLEC14A, and the potential role of glycosaminoglycan binding, a chemical cross-linking strategy using Sulfo-SBED was developed using recombinant human CLEC14A (Chapter 4). This method confirmed the interaction of CLEC14A with MMRN2 and expanded its interactome to include proteins involved

in extracellular matrix organization, integrin assembly and cell migration and morphogenesis. These proteins include fibronectin (FN1), integrin beta 1 (ITGB1), neuropillin1 (NRP1), inter-trypsin alpha inhibitor heavy chain H5 (ITIH5) and EGF-containing fibulin-like extracellular matrix protein 1 (EFEMP1), and alpha-taxilin (TXLNA). Protein-protein binding was altered in the CLEC14A-R161A mutant suggesting that the interaction with heparin may impact CLEC14A interacting partners.

Summary

CLEC14A is a single pass transmembrane protein consisting of a CTLD, EGF, endomucin domain followed by a transmembrane region and an unstructured cytoplasmic tail. The endomucin domain promotes trimerization of CLEC14A, potentially increasing valency of the CTLD for its ligands. A plausible model is provided in Figure 3.6. Biochemical data suggest MMRN2 bridges endothelial CLEC14A or CD93 to pericyte CD248 acting as a glue (127). This is consistent with the reduced endothelial-pericyte coverage and nonfunctional neovascularization observed in mice with a genetic ablation of either CLEC14A, MMRN2, CD248 or CD93 (115,116,171,175,179,203) . The abundance of MMRN2 in the extracellular matrix may also act as sink for extracellular VEGF-A and VEGF-C acting as a VEGFR2 rheostat to control blood vessel stability (129,189). CLEC14A complexes with VEGFR3 to fine tune endothelial tip-stalk cell cross talk via VEGFR and NOTCH signaling (115). Removal of CLEC14A has the reciprocal effect on VEGFR activation as overexpression suggesting CLEC14A may titer this response (115). The secretion of HSP70-1A from cancer cells (191-194) binds the CLEC14A-CTLD to stabilize endothelial junctions and promote inter-endothelial cell interactions (126,202). The protease RHBLD2 may cleave CLEC14A from the cell surface and release it into the extracellular space where it may bind to endothelial tip cells to potentially block CLEC14A protein-protein interactions (128). During cell migration, MMRN2 binding to CD93 induces fibronectin fibrillogenesis where it forms a complex with ITGB1 to stimulate

endothelial focal adhesions (179). Overall, the flux of CLEC14A and its binding partners protein expression, localization and processing suggests that CLEC14A is dynamically and kinetically involved in mediating angiogenesis.

Figures

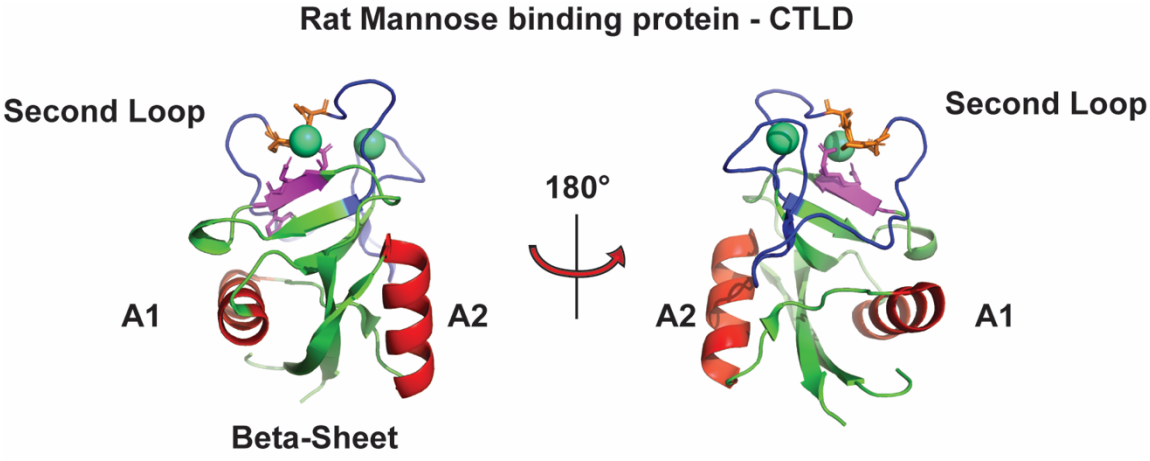


Figure 3.1: The structure of C-type Lectin domain (CTLD) for the Rat mannose binding protein (PDB: 1MSB), with two alpha-helices (red), the second loop region (blue), Ca^{2+} (teal), EPN motif (orange) and WPN motif (magenta).

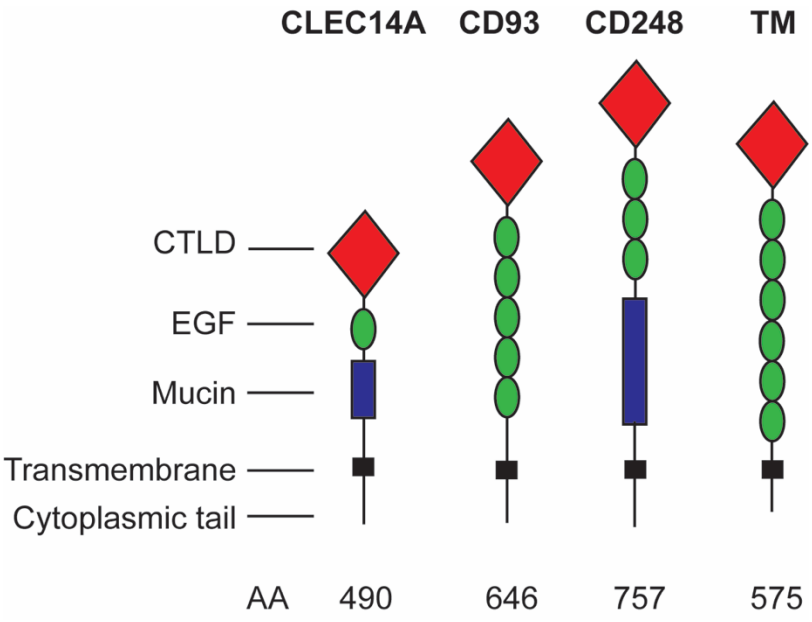


Figure 3.2: Schematic representation of the C-type lectin 14a family. CTLDs (red), EGF (green), and mucin like domains (blue).

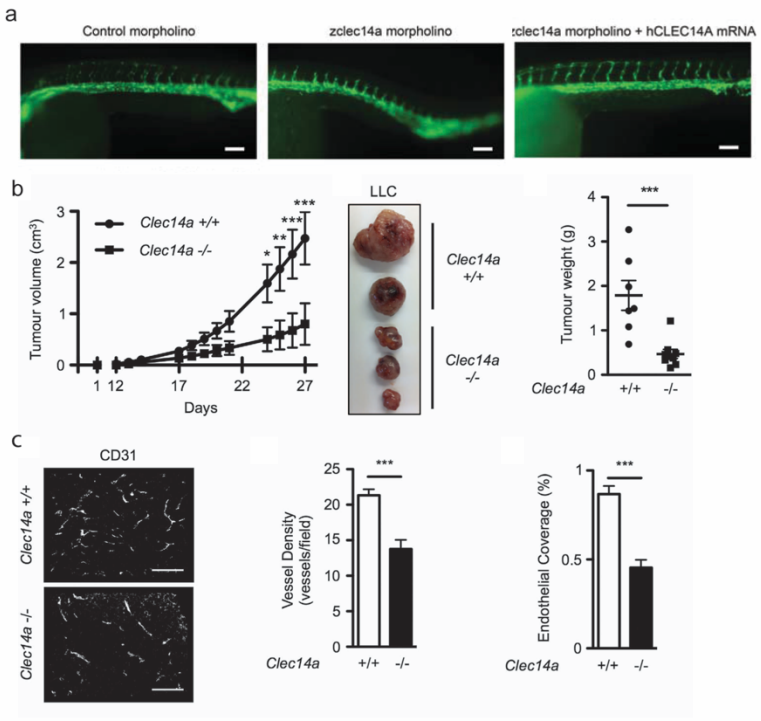


Figure3. 3: CLEC14A alters angiogenesis in vivo. (a) Morpholino knockdown and recovery of zebrafish CLEC14A during intersegmental vessel development. Adapted from Mura M. *et. al.* (2012). *Clec14a*$^{-/-}$ mice have reduced Lewis lung carcinoma tumor growth (b) with decreased blood vessel densities and pericyte coverage (c). Adapted from Noy PJ *et. al.* (2015).

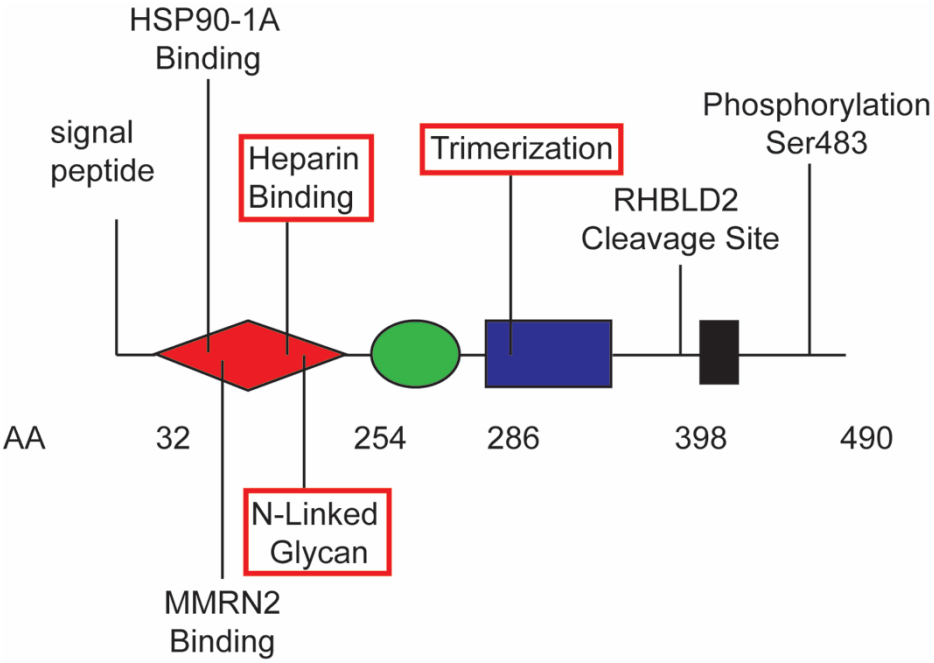


Figure 3.4: An annotated linear map of CLEC14A depicting known structural features. Red boxes represent novel findings from Chapter 2.

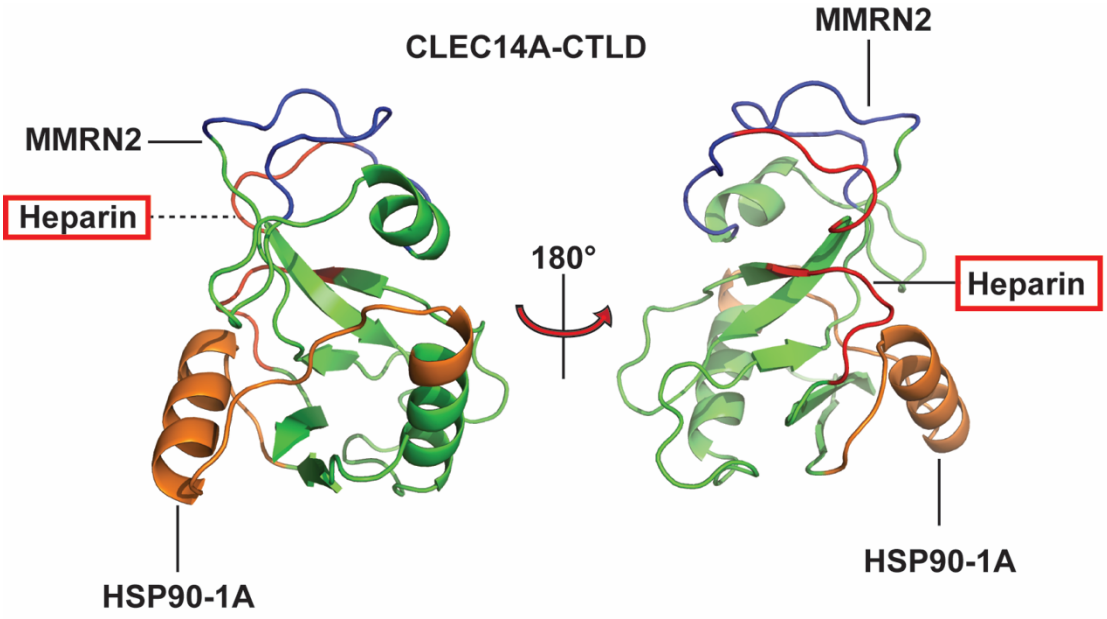


Figure 3.5: Homologue protein model of CLEC14A generated by robbetta labeling key regions of the CLEC14A-CTLD involved in ligand binding to MMRN2 (blue), HSP90-1A (orange) and heparin (red).

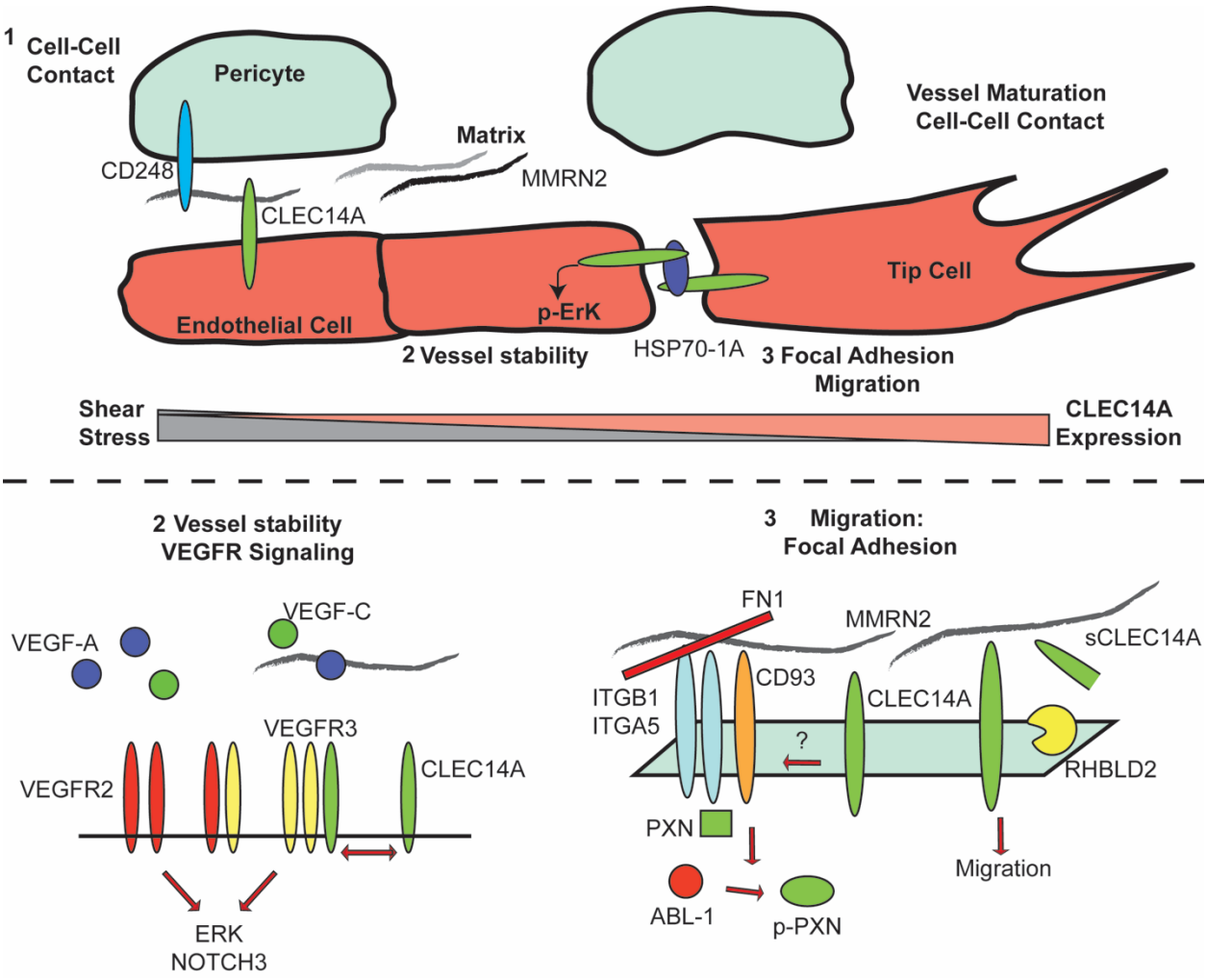


Figure 3.6: Overview of the proposed functions of CLEC14A during blood vessel formation. (1) CLEC14A regulates angiogenesis and vessel maturation by cell-cell interactions, vessel stability, focal adhesion and migration. Pericyte-endothelial cell interactions are mediated by pericyte CD248 and endothelial CLEC14A binding to mmrn2 to control blood vessel regression and maturation during angiogenesis. HSP70-1A bridges and stabilizes CLEC14A amongst endothelial cells to control endothelial ERK phosphorylation. (2) CLEC14A engages the VEGFR3 complex to fine tune and balance VEGFR2-VEGFR3 signaling for proper blood vessel maturation. MMRN2 may act to sequester and titer VEGF-A and VEGF-C. (3) CLEC14A regulates cell migration through interactions with MMRN2. CD93 in concert with MMRN2 promotes fibronectin fibril formation to allow proper endothelial focal adhesion assembly. It is unknown if CLEC14A may have a redundant role. CLEC14A may be cleaved by RHBLD2 to create soluble CLEC14A where it may localize to endothelial tip cells.

References

1. Zelensky, A. N., and Gready, J. E. (2005) The C-type lectin-like domain superfamily. *FEBS J* **272**, 6179-6217

2. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991) Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* **254**, 1608-1615

3. Isermann, B., Hendrickson, S. B., Zogg, M., Wing, M., Cummiskey, M., Kisanuki, Y. Y., Yanagisawa, M., and Weiler, H. (2001) Endothelium-specific loss of murine thrombomodulin disrupts the protein C anticoagulant pathway and causes juvenile-onset thrombosis. *J Clin Invest* **108**, 537-546

4. Griffin, J. H., Zlokovic, B. V., and Mosnier, L. O. (2015) Activated protein C: biased for translation. *Blood* **125**, 2898-2907

5. Zlokovic, B. V. (1997) Antithrombotic, procoagulant, and fibrinolytic mechanisms in cerebral circulation: implications for brain injury and protection. *Neurosurg Focus* **2**, e5

6. Rettig, W. J., Garin-Chesa, P., Healey, J. H., Su, S. L., Jaffe, E. A., and Old, L. J. (1992) Identification of endosialin, a cell surface glycoprotein of vascular endothelial cells in human cancer. *Proc Natl Acad Sci U S A* **89**, 10832-10836

7. Opavsky, R., Haviernik, P., Jurkovicova, D., Garin, M. T., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Bies, J., Garfield, S., Pastorekova, S., Oue, A., and Wolff, L. (2001) Molecular characterization of the mouse Tem1/endosialin gene regulated by cell density in vitro and expressed in normal tissues in vivo. *J Biol Chem* **276**, 38795-38807

8. MacFadyen, J. R., Haworth, O., Roberston, D., Hardie, D., Webster, M. T., Morris, H. R., Panico, M., Sutton-Smith, M., Dell, A., van der Geer, P., Wienke, D., Buckley, C. D., and Isacke, C. M. (2005) Endosialin (TEM1, CD248) is a marker of stromal fibroblasts and is not selectively expressed on tumour endothelium. *FEBS Lett* **579**, 2569-2575

9. Nanda, A., Karim, B., Peng, Z., Liu, G., Qiu, W., Gan, C., Vogelstein, B., St Croix, B., Kinzler, K. W., and Huso, D. L. (2006) Tumor endothelial marker 1 (Tem1) functions in the growth and progression of abdominal tumors. *Proc Natl Acad Sci U S A* **103**, 3351-3356

10. Simonavicius, N., Ashenden, M., van Weverwijk, A., Lax, S., Huso, D. L., Buckley, C. D., Huijbers, I. J., Yarwood, H., and Isacke, C. M. (2012) Pericytes promote selective vessel regression to regulate vascular patterning. *Blood* **120**, 1516-1527

11. Tomkowicz, B., Rybinski, K., Foley, B., Ebel, W., Kline, B., Routhier, E., Sass, P., Nicolaides, N. C., Grasso, L., and Zhou, Y. (2007) Interaction of endosialin/TEM1 with extracellular matrix proteins mediates cell adhesion and migration. *Proc Natl Acad Sci U S A* **104**, 17965-17970

12. Becker, R., Lenter, M. C., Vollkommer, T., Boos, A. M., Pfaff, D., Augustin, H. G., and Christian, S. (2008) Tumor stroma marker endosialin (Tem1) is a binding partner of metastasis-related protein Mac-2 BP/90K. *FASEB J* **22**, 3059-3067

13. Khan, K. A., Naylor, A. J., Khan, A., Noy, P. J., Mambretti, M., Lodhia, P., Athwal, J., Korzystka, A., Buckley, C. D., Willcox, B. E., Mohammed, F., and Bicknell, R. (2017) Multimerin-2 is a ligand for group 14 family C-type lectins CLEC14A, CD93 and CD248 spanning the endothelial pericyte interface. *Oncogene* **36**, 6097-6108

14. Dieterich, L. C., Mellberg, S., Langenkamp, E., Zhang, L., Zieba, A., Salomaki, H., Teichert, M., Huang, H., Edqvist, P. H., Kraus, T., Augustin, H. G., Olofsson, T., Larsson, E., Soderberg, O., Molema, G., Ponten, F., Georgii-Hemming, P., Alafuzoff, I., and Dimberg, A. (2012) Transcriptional profiling of human glioblastoma vessels indicates a key role of VEGF-A and TGFbeta2 in vascular abnormalization. *J Pathol* **228**, 378-390

15. Langenkamp, E., Zhang, L., Lugano, R., Huang, H., Elhassan, T. E., Georganaki, M., Bazzar, W., Loof, J., Trendelenburg, G., Essand, M., Ponten, F., Smits, A., and Dimberg, A. (2015) Elevated expression of the C-type lectin CD93 in the glioblastoma vasculature regulates cytoskeletal rearrangements that enhance vessel function and reduce host survival. *Cancer Res* **75**, 4504-4516

16. Galvagni, F., Nardi, F., Spiga, O., Trezza, A., Tarticchio, G., Pellicani, R., Andreuzzi, E., Caldi, E., Toti, P., Tosi, G. M., Santucci, A., Iozzo, R. V., Mongiat, M., and Orlandini, M. (2017) Dissecting the CD93-Multimerin 2 interaction involved in cell adhesion and migration of the activated endothelium. *Matrix Biol* **64**, 112-127

17. Masiero, M., Simoes, F. C., Han, H. D., Snell, C., Peterkin, T., Bridges, E., Mangala, L. S., Wu, S. Y., Pradeep, S., Li, D., Han, C., Dalton, H., Lopez-Berestein, G., Tuynman, J. B., Mortensen, N., Li, J. L., Patient, R., Sood, A. K., Banham, A. H., Harris, A. L., and Buffa, F. M. (2013) A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis. *Cancer Cell* **24**, 229-241

18. Galvagni, F., Nardi, F., Maida, M., Bernardini, G., Vannuccini, S., Petraglia, F., Santucci, A., and Orlandini, M. (2016) CD93 and dystroglycan cooperation in human endothelial cell adhesion and migration adhesion and migration. *Oncotarget* **7**, 10090-10103

19. Lugano, R., Vemuri, K., Yu, D., Bergqvist, M., Smits, A., Essand, M., Johansson, S., Dejana, E., and Dimberg, A. (2018) CD93 promotes beta1 integrin activation and fibronectin fibrillogenesis during tumor angiogenesis. *J Clin Invest* **128**, 3280-3297

20. Jang, J., Kim, M. R., Kim, T. K., Lee, W. R., Kim, J. H., Heo, K., and Lee, S. (2017) CLEC14a-HSP70-1A interaction regulates HSP70-1A-induced angiogenesis. *Sci Rep* **7**, 10666

21. Mura, M., Swain, R. K., Zhuang, X., Vorschmitt, H., Reynolds, G., Durant, S., Beesley, J. F., Herbert, J. M., Sheldon, H., Andre, M., Sanderson, S., Glen, K., Luu, N. T., McGettrick, H. M., Antczak, P., Falciani, F., Nash, G. B., Nagy, Z. S., and Bicknell, R. (2012) Identification and angiogenic role of the novel tumor endothelial marker CLEC14A. *Oncogene* **31**, 293-305

22. Zanivan, S., Maione, F., Hein, M. Y., Hernandez-Fernaud, J. R., Ostasiewicz, P., Giraudo, E., and Mann, M. (2013) SILAC-based proteomics of human primary

endothelial cell morphogenesis unveils tumor angiogenic markers. *Mol Cell Proteomics* **12**, 3599-3611

23. Delcourt, N., Quevedo, C., Nonne, C., Fons, P., O'Brien, D., Loyaux, D., Diez, M., Autelitano, F., Guillemot, J. C., Ferrara, P., Muriana, A., Callol, C., Herault, J. P., Herbert, J. M., Favre, G., and Bono, F. (2015) Targeted identification of sialoglycoproteins in hypoxic endothelial cells and validation in zebrafish reveal roles for proteins in angiogenesis. *J Biol Chem* **290**, 3405-3417

24. Maeng, Y. S., Choi, H. J., Kwon, J. Y., Park, Y. W., Choi, K. S., Min, J. K., Kim, Y. H., Suh, P. G., Kang, K. S., Won, M. H., Kim, Y. M., and Kwon, Y. G. (2009) Endothelial progenitor cell homing: prominent role of the IGF2-IGF2R-PLCbeta2 axis. *Blood* **113**, 233-243

25. Rho, S. S., Choi, H. J., Min, J. K., Lee, H. W., Park, H., Park, H., Kim, Y. M., and Kwon, Y. G. (2011) Clec14a is specifically expressed in endothelial cells and mediates cell to cell adhesion. *Biochem Biophys Res Commun* **404**, 103-108

26. Lee, S., Rho, S. S., Park, H., Park, J. A., Kim, J., Lee, I. K., Koh, G. Y., Mochizuki, N., Kim, Y. M., and Kwon, Y. G. (2017) Carbohydrate-binding protein CLEC14A regulates VEGFR-2- and VEGFR-3-dependent signals during angiogenesis and lymphangiogenesis. *J Clin Invest* **127**, 457-471

27. Noy, P. J., Swain, R. K., Khan, K., Lodhia, P., and Bicknell, R. (2016) Sprouting angiogenesis is regulated by shedding of the C-type lectin family 14, member A (CLEC14A) ectodomain, catalyzed by rhomboid-like 2 protein (RHBDL2). *FASEB J* **30**, 2311-2323

28. Ki, M. K., Jeoung, M. H., Choi, J. R., Rho, S. S., Kwon, Y. G., Shim, H., Chung, J., Hong, H. J., Song, B. D., and Lee, S. (2013) Human antibodies targeting the C-type lectin-like domain of the tumor endothelial cell marker clec14a regulate angiogenic properties in vitro. *Oncogene* **32**, 5449-5457

29. Noy, P. J., Lodhia, P., Khan, K., Zhuang, X., Ward, D. G., Verissimo, A. R., Bacon, A., and Bicknell, R. (2015) Blocking CLEC14A-MMRN2 binding inhibits sprouting angiogenesis and tumour growth. *Oncogene* **34**, 5821-5831

30. Du, J., Yang, Q., Luo, L., and Yang, D. (2017) C1qr and C1qrl redundantly regulate angiogenesis in zebrafish through controlling endothelial Cdh5. *Biochem Biophys Res Commun* **483**, 482-487

31. Zarkada, G., Heinolainen, K., Makinen, T., Kubota, Y., and Alitalo, K. (2015) VEGFR3 does not sustain retinal angiogenesis without VEGFR2. *Proc Natl Acad Sci U S A* **112**, 761-766

32. Benedito, R., Rocha, S. F., Woeste, M., Zamykal, M., Radtke, F., Casanovas, O., Duarte, A., Pytowski, B., and Adams, R. H. (2012) Notch-dependent VEGFR3 upregulation allows angiogenesis without VEGF-VEGFR2 signalling. *Nature* **484**, 110-114

33. Tammela, T., Zarkada, G., Nurmi, H., Jakobsson, L., Heinolainen, K., Tvorogov, D., Zheng, W., Franco, C. A., Murtomaki, A., Aranda, E., Miura, N., Yla-Herttuala, S., Fruttiger, M., Makinen, T., Eichmann, A., Pollard, J. W., Gerhardt, H., and Alitalo, K. (2011) VEGFR-3 controls tip to stalk conversion at vessel fusion sites by reinforcing Notch signalling. *Nat Cell Biol* **13**, 1202-1213

34. Kim, T. K., Park, C. S., Jang, J., Kim, M. R., Na, H. J., Lee, K., Kim, H. J., Heo, K., Yoo, B. C., Kim, Y. M., Lee, J. W., Kim, S. J., Kim, E. S., Kim, D. Y., Cha, K., Lee, T. G., and Lee, S. (2018) Inhibition of VEGF-dependent angiogenesis and tumor angiogenesis by an optimized antibody targeting CLEC14a. *Mol Oncol* **12**, 356-372

35. Christian, S., Ahorn, H., Novatchkova, M., Garin-Chesa, P., Park, J. E., Weber, G., Eisenhaber, F., Rettig, W. J., and Lenter, M. C. (2001) Molecular cloning and characterization of EndoGlyx-1, an EMILIN-like multisubunit glycoprotein of vascular endothelium. *J Biol Chem* **276**, 48588-48595

36. Andreuzzi, E., Colladel, R., Pellicani, R., Tarticchio, G., Cannizzaro, R., Spessotto, P., Bussolati, B., Brossa, A., De Paoli, P., Canzonieri, V., Iozzo, R. V., Colombatti, A., and Mongiat, M. (2017) The angiostatic molecule Multimerin 2 is processed by MMP-9 to allow sprouting angiogenesis. *Matrix Biol* **64**, 40-53

37. Lorenzon, E., Colladel, R., Andreuzzi, E., Marastoni, S., Todaro, F., Schiappacassi, M., Ligresti, G., Colombatti, A., and Mongiat, M. (2012) MULTIMERIN2 impairs tumor angiogenesis and growth by interfering with VEGF-A/VEGFR2 pathway. *Oncogene* **31**, 3136-3147

38. Lindquist, S. (1986) The heat-shock response. *Annu Rev Biochem* **55**, 1151-1191

39. Ralhan, R., and Kaur, J. (1995) Differential expression of Mr 70,000 heat shock protein in normal, premalignant, and malignant human uterine cervix. *Clin Cancer Res* **1**, 1217-1222

40. Malusecka, E., Zborek, A., Krzyzowska-Gruca, S., and Krawczyk, Z. (2001) Expression of heat shock proteins HSP70 and HSP27 in primary non-small cell lung carcinomas. An immunohistochemical study. *Anticancer Res* **21**, 1015-1021

41. Hwang, T. S., Han, H. S., Choi, H. K., Lee, Y. J., Kim, Y. J., Han, M. Y., and Park, Y. M. (2003) Differential, stage-dependent expression of Hsp70, Hsp110 and Bcl-2 in colorectal cancer. *J Gastroenterol Hepatol* **18**, 690-700

42. Abe, M., Manola, J. B., Oh, W. K., Parslow, D. L., George, D. J., Austin, C. L., and Kantoff, P. W. (2004) Plasma levels of heat shock protein 70 in patients with prostate cancer: a potential biomarker for prostate cancer. *Clin Prostate Cancer* **3**, 49-53

43. Ciocca, D. R., Clark, G. M., Tandon, A. K., Fuqua, S. A., Welch, W. J., and McGuire, W. L. (1993) Heat shock protein hsp70 in patients with axillary lymph node-negative breast cancer: prognostic implications. *J Natl Cancer Inst* **85**, 570-574

44. Santarosa, M., Favaro, D., Quaia, M., and Galligioni, E. (1997) Expression of heat shock protein 72 in renal cell carcinoma: possible role and prognostic implications in cancer patients. *Eur J Cancer* **33**, 873-877

45. Thomas, X., Campos, L., Mounier, C., Cornillon, J., Flandrin, P., Le, Q. H., Piselli, S., and Guyotat, D. (2005) Expression of heat-shock proteins is associated with major adverse prognostic factors in acute myeloid leukemia. *Leuk Res* **29**, 1049-1058

46. Du, X. L., Jiang, T., Wen, Z. Q., Gao, R., Cui, M., and Wang, F. (2009) Silencing of heat shock protein 70 expression enhances radiotherapy efficacy and inhibits cell invasion in endometrial cancer cell line. *Croat Med J* **50**, 143-150

47. Meng, L., Hunt, C., Yaglom, J. A., Gabai, V. L., and Sherman, M. Y. (2011) Heat shock protein Hsp72 plays an essential role in Her2-induced mammary tumorigenesis. *Oncogene* **30**, 2836-2845

48. Goloudina, A. R., Demidov, O. N., and Garrido, C. (2012) Inhibition of HSP70: a challenging anti-cancer strategy. *Cancer Lett* **325**, 117-124

49. Juhasz, K., Lipp, A. M., Nimmervoll, B., Sonnleitner, A., Hesse, J., Haselgruebler, T., and Balogi, Z. (2013) The complex function of hsp70 in metastatic cancer. *Cancers (Basel)* **6**, 42-66

50. Kim, T. K., Na, H. J., Lee, W. R., Jeoung, M. H., and Lee, S. (2016) Heat shock protein 70-1A is a novel angiogenic regulator. *Biochem Biophys Res Commun* **469**, 222-228

51. Bao, L., Tang, M., Zhang, Q., You, B., Shan, Y., Shi, S., Li, L., Hu, S., and You, Y. (2016) Elevated expression of CD93 promotes angiogenesis and tumor growth in nasopharyngeal carcinoma. *Biochem Biophys Res Commun* **476**, 467-474

CHAPTER 4: MAPPING CLEC14A-CELL SUFACE PROTEIN INTERACTIONS USING SULFO-SBED CROSS-LINKING

Keywords: Cross-linking, protein-protein interactions, proteomics, cell surface, glycan, heparin, focal adhesion, CLEC14A, neuropilin

Abstract

Protein interactions drive many cellular functions. Here we present a novel proteomic work flow termed RIPID to map cell surface interactions and protein interfaces by using chemical cross-linking affinity purification and mass spectrometry to identify and quantitate dynamic cell surface interactions. RIPID employs conjugation of a chemical cross-linker, sulfo-SBED biotin (Sulfo-N-hydroxysuccinimidyl-2-(6-[biotinamido]-2-(p-azido benzamido)-hexanoamido) ethyl-1,3'-dithioproprionate) to a bait protein to identify their cell surface ligands on live cells. In our study, we use C-type lectin 14a (CLEC14A), an endothelial transmembrane protein, to validate our approach and expand the CLEC14A interactome to 44 endothelial cell surface proteins. Furthermore, we use a protect and label strategy in conjunction with RIPID to map a CLEC14A-heparin protein interactome interface. Upon doing so, we identified 14 proteins that bind CLEC14A including MMRN2, ITIH5, NRP1, TXLNA, and EFEMP1 and show that the site critical for mediating CLEC14A-heparin interactions binds to NRP1 and ITIH5. Moreover, we show that CLEC14A regulates endothelial cell morphogenesis in vitro through these interactions. We suggest that heparan sulfate alters the CLEC14A protein binding landscape to promote endothelial tube morphogenesis.

Introduction

Cell surface proteins respond to extracellular cues to initiate multiple signaling cascades inside a cell to drive key biological functions. The ectodomains of cell surface proteins interact with other cell surface proteins including receptors, growth factors, adhesion proteins and matrix components. Whether embedded in the cell membrane or deposited in the extracellular matrix, cell surface proteins are surrounded by a dense mesh of glycans termed the glycocalyx (204). The glycocalyx is composed of all glycans at the cell surface. These glycans include heparan sulfate, a glycosaminoglycan that modulates the function of many extracellular proteins by either acting to tether, scaffold, allosterically regulate, catalyze, or oligomerize a protein (Chapter 1). Defining these protein-protein interactions (PPI) and protein-carbohydrate interactions (PCI) are essential to understanding many biological processes. However, identifying interacting partners and how glycans modulate the interactions remain a challenge.

For PPIs, groups have used various complementary methods including affinity purification-mass spectrometry (AP-MS), proximity-based labeling, cross-linking MS (XL-MS) and yeast two-hybrid (Y2H) screens(205). In AP-MS, endogenous protein or a bait protein is tagged and enriched using an affinity resin followed by liquid chromatography mass spectrometry (LC-MS/MS). In this approach, identification of PPIs are often limited to strong, and stable interactions that resist harsh extraction methods. PPIs that are weak and transient can be easily missed. To circumvent this issue, others have implemented proximity labeling strategies to covalently modify and tag adjacent proteins. By covalently X-linking and tagging proteins, one can identify both strong and transient PPIs by enriching them using an affinity-based pulldown, and identifying the proteins by LC-MS/MS. The methods used to implement proximity labeling fall into two categories, enzymatic modification and cross-linking. In methods employing enzymatic modification, a bait protein is fused to an enzyme such as the *Escherichia coli* biotin ligase BirA (BioID) or a ascorbate peroxidase (APEX) to biotinylate and tag adjacent proteins. APEX and BioID provide in maps of the interactions of proteomes of intracellular

compartments (206). However, limitations of APEX and BioID include the need to fuse to . To perform these methods, these enzymes are fused at the N- or C- terminal end of the bait protein and overexpress in cell, limiting resolution of a protein interface and introducing “cloning” scars. Moreover, BioID may only modify exposed lysines of target proteins and APEX may only modify tyrosines, tryptophans, histidines and cysteines. Other groups have taken up photo cross-linking to identify PPIs, and protein-RNA interactions. In these approaches, a target gene is modified to encode an artificial biorthogonal amino acid containing a photo reactive cross-linking group (207-210).

Despite the great advances these proteomic techniques have made in identifying intracellular PPIs, there are still many issues identifying ligands at the cell surface. In this study, we present a cross-linking proximity-labeling affinity purification mass spectrometry workflow (**RIPID**, Really Interesting Protocol Implemented by Daniel) to identify PPIs of ectodomains of cell surface proteins. To validate our approach, we used C-type Lectin 14A (CLEC14A), an endothelial cell surface protein involved in regulating angiogenesis, to map its respective PPI. We identified extracellular matrix and cell surface proteins including MMRN2, ITIH5, NRP1, TXLNA, and EFEMP1 as CLEC14A protein binding partners. Moreover, we show that alanine mutageneis of R161, a key site in CLEC14A-heparin binding site, alters CLEC14A PPIs. We suggest that CLEC14A-heparan sulfate binding may act to partition CLEC14A protein interactions to fine tune and control endothelial cell adhesion, migration and morphogenesis.

Results

CLEC14A-R161A exhibits dominant-negative activity during tube formation

In Chapter 2, we found that CLEC14A bound heparin with high nanomolar affinity and arginine 161 plays a key role in binding. To test if CLEC14A-heparin interactions was critical in mediating endothelial tube formation, HUVEC were transfected with cDNAs encoding full length CLEC14A or full length CLEC14A-R161A. Overexpressing cells were then challenged to form

tubes in Matrigel. After 8 hours, HUVEC typically forms an anastomosing network of tubules (Supplemental Figure 4.1a). Expression of wildtype CLEC14A had no significant effect on tube formation compared to HUVEC transfected with cDNA encoding GFP. However, overexpression of CLEC14A-R161A led to a significant decrease in tube formation (Supplemental Figure 4.1a). Similarly, expression of the CLEC14A ectodomain as a secreted protein (eCLEC14A) had no effect on vessel formation whereas expression of eCLEC14A-R161A inhibited tube formation (Supplemental Figure 4.1b). These data suggest that soluble CLEC14A-R161A exhibits dominant negative activity during endothelial tube formation. We suspected that R161 was a site critical to CLEC14A function, and that this site was involved in heparan sulfate mediated protein-protein interactions.

One potential function of heparan sulfate-protein interactions is to modulate PPIs. At the time our studies began, MMRN2 was the only known CLEC14A protein binding protein (116,129). To test if heparin may alter CLEC14A-MMRN2 binding, we first immunoprecipitated recombinant MMRN2 and His-tagged CLEC14A with an anti-his antibody. Recombinant MMRN2 bound to CLEC14A (figure 4.2a,b). To test if heparin can block this interaction, we pretreated CLEC14A with heparin. Heparin dose dependently blocked CLEC14A from binding to MMRN2. Binding of MMRN2 was independent of the heparin-binding site based on the ability of CLEC14A-R161A to engage MMRN2 to the same extent as the wildtype protein. When heparin did not block MMRN2-CLEC14A-R161A binding (supplementary figure 4.2a,b). Heparin size dependently blocks MMRN2-CLEC14A interactions suggesting that heparin binding to CLEC14A is independent of the CLEC14A MMRN2 binding site. (supplementary figure 4.2c).

Suflo-SBED conjugation workflow

To test whether if heparan sulfate may alter CLEC14A cell surface protein-protein interactions, we first needed to develop and identify the endothelial CLEC14A protein binding repertoire. We implemented a chemical cross-linking proximity-labeling affinity purification

coupled with high resolution mass spectrometry workflow. In this method, a bait protein is conjugated with a biotinylated bifunctional crosslinker, such as Sulfo-SBED (sulfo-N-hydroxysuccinimidyl-2-(6-[biotinamido]-2-(p-azido benzamido)-hexanoamido) ethyl-1,3'-dithioproprionate) (figure 4.1a). We chose Sulfo-SBED due to its chemical properties: an NHS (N-hydroxysuccinimide) moiety to conjugate to exposed lysines residues; an aryl-azide group to photo-cross link bait to cell surface proteins; a biotin moiety to purify and enrich cross-linked prey proteins; and a disulfide bridge to separate bait and prey proteins for purification, enrichment and mass spectrometry analysis (figure 4.1). First recombinant CLEC14A (rCLEC14A) was conjugated with excess Sulfo-SBED in solution to form $CLEC14A_{ss}$ (CLEC14A-Sulfo-SBED). $CLEC14A_{ss}$ conjugate was then purified using a NAX5 column (figure 4.1b).

Chemical crosslinking of recombinant CLEC14A to the surface of HUVECs reveals cell surface interacting partners

As CLEC14A known function occurs in the vasculature (115,116,130,132) and has a limited number of known PPIs (116,126,128,129), we adapted our mapping strategy to HUVEC (human umbilical vein endothelial cells) and asked if RIPID can identify CLEC14A PPIs on cultured endothelial cells. $CLEC14A_{ss}$ was applied to HUVEC cell monolayers for 5 minutes at 37 °C followed by treatment with and without UV light to photo-crosslink (XL) $CLEC14A_{ss}$ to HUVEC cell surface proteins (figure 4.1b). Cells were then lysed, and prey proteins were isolated and enriched by streptavidin-affinity chromatography, trypsinized and subjected to LC-MS/MS analysis and label free quantification (LFQ) . Principle component analysis revealed clustering of three groups compromised of the experimental conditions: cross-linked $CLEC14A_{ss}$, $CLEC14A_{ss}$ and lysates (figure 4.2a) indicating the enrichment of specific proteins. We next plotted the intensities of XL- $CLEC14A_{ss}$ versus the lysate to identify proteins significantly enriched by RIPID (figure 4.1b). RIPID identified 41 CLEC14A endothelial cell

surface protein-protein interactions (Table 4.1). Extracellular matrix proteins including Mulitmerin-2 (MMRN2), Multimerin-1 (MMRN1), Fibronectin (FN1), and Laminin (LAMA4), and membrane bound cell surface proteins including Endoglin (ENG), Platelet and endothelial cell adhesion molecule 1 (PECAM1), integrin subunit beta 1 (ITGB1), integrin subunit alpha 5 (ITGA5), and integrin subunit alpha 5 (ITGA3) were identified and significantly enriched. Photo-cross linking increased the number of tagged cell surface proteins by 24 including Neuropilin-1 (NRP1), CD44 antigen (CD44), Ephrin type-A receptor 2 (EPHA2), Intercellular adhesion molecule 2 (ICAM2), and CD166 antigen (ALCAM) suggesting photo-crosslinking increases the sensitivity of identifying PPI (Figure 4.2d). RIPID enhances the number of identified PPIs.

CLEC14A is a known heparan sulfate binding protein, and residues found to interact with heparin have been determined (Chapter 2). While one principal function of heparan sulfate is to modulate PPIs, the manner in which heparan sulfate regulates endothelial CLEC14A function is uncertain. We next asked if RIPID, may be used to map PPI on unique faces of a protein. Although there is not an available structure for CLEC14A, the modeled and predicted heparin binding site contains a lysine. To prevent any NHS conjugation of lysine's involved in heparin binding, we implemented a protect and label strategy. First rCLEC14A was bound to a heparin Sepharose column to protect lysine residues in the heparin-binding site from conjugation (Chapter 2) (figure 4.3a). Then Sulfo-SBED was added to the rCLEC14A heparin Sepharose column to construct $CLEC14A_{HP}$ (rCLEC14A-heparin protected-Sulfo-SBED). After several washes, the $CLEC14A_{HP}$ conjugates were eluted. We then used a mutant form of CLEC14A with significantly lower capacity to bind heparin under high salt conditions, CLEC14A-R161A. CLEC14A-R161A-Sulfo-SBED conjugation followed the same flow though as rCLEC14A. We then applied the $CLEC14A_{HP}$ to confluent HUVEC and photo-cross-linked to CLEC14A PPIs. When comparing cross-linked rCLEC14A versus the lysate, seventy-five proteins were significantly different compared to the untreated lysate controls (Table 4.2). Fifteen proteins were identified displaying more than 2-fold enrichment compared with the lysate (Table 4.2). Of

the fifteen proteins, five are extracellular proteins (Inter-alpha-trypsin inhibitor heavy chain 5 (ITIH5), neuropilin-1 (NRP1), Taxilin A (TXLNA), EGF Containing Fibulin Extracellular Matrix Protein 1 (EFEMP1) and MMRN2. In addition to CLEC14A itself, which was used as a bait, MMRN2, a known CLEC14A PPI was among the highest enriched proteins, confirming the feasibility and validity of the approach (127,129). However, photo cross-linking of CLEC14A-R161A was significantly reduction in ITIH5, NRP1, and TXLNA while maintaining an enrichment of MMRN2 and EFEMP1 (figure 4.3a,b). These data suggest that R161 is critical in mediating binding of CLEC14A to ITIH5, NRP1 and TXLNA, and that RIPID may be used to map PPIs on unique sites of proteins.

Sulfo-SBED Cross-linking identifies high and low affinity binding proteins.

To validate RIPID and confirm CLEC14A PPIs, we performed surface plasmon resonance (SPR) on three CLEC14A protein binding partners identified by RIPID. MMRN2 bound to immobilized CLEC14A with a K_d = 0.04 μM affinity (Figure 4.4a). NRP1 bound with a K_d =0.11 μM affinity (Figure 4.4b) and was characterized with fast on rate and slow off rate while EFEMP1 bound a K_d = 0.8 μM with fast on and off rates (Figure 4.4c). Thus, RIPID may be used to identify and map PPIs with high and low affinities.

Sulfo-SBED targets genetically interact with CLEC14A to regulate endothelial morphogenesis.

The majority of CLEC14A protein binding partners are implicated in extracellular matrix organization, integrin interactions, cell migration and endothelial morphogenesis (figure 4.3C). Furthermore, recent studies have shown that CLEC14A is a regulator of endothelial morphogenesis, migration and angiogenesis (115,116,130,132). To test if CLEC14A PPIs regulate endothelial morphogenesis on Matrigel, we silenced the expression of CLEC14A,

MMRN2, NRP1, ITIH5, EFEMP1 and TXLNA using one or two individual siRNAs per gene. After transfecting HUVEC with the respective siRNA, HUVEC were plated on Matrigel and allowed to form sprouts. Compared to the scramble control, expression of CLEC14A was decreased by 82.6% and 42.1%, and NRP1 was knockdown 90.7% and 69% (supplementary figure 4.4a,b). MMRN2 was reduced 97.4% and EFEMP1 was reduced 95.4 and 60.8% (supplementary figure 4.4c,d). Knockdown of CLEC14A inhibited endothelial tube formation by decreasing overall sprout length and preventing the formation of interconnected networks as measured by the number of meshes (figure 4.4a,b). Knockdown of MMRN2, EFEMP1 and ITIH5 lead to an overall decrease in total sprout length and decreased the number of meshes whereas knockdown of NRP1 altered endothelial morphogenesis by maintaining overall sprout length but decreasing the number of meshes (figure 4.4a,b). To test for genetic interactions between CLEC14A and NRP1 during endothelial morphogenesis, HUVEC were transfected with siRNAs targeting NRP1 in the presence of absence of siRNA targeting either a scramble siRNA or a CLEC14A siRNA. Transfected HUVEC were then plated on Matrigel to measure endothelial tube formation (figure 4.4c). Knockdown of NRP1 reduced the overall sprout length and decreased the number of meshes (figure 4.4d). However, when CLEC14A was knocked down in the presence of a NRP1 knockdown, the overall length and number of meshes were restored (figure 4.4d) suggesting that CLEC14A and NRP1 interact to regulate endothelial morphogenesis.

Discussion

Here we describe a proteomics workflow utilizing the small molecule Sulfo-SBED to identify and map PPIs at the cell surface and identify regulators of endothelial morphogenesis. There are 4 moieties on Sulfo-SBED that make it suitable to identify PPIs: an NHS moiety to conjugate Sulfo-SBED to bait proteins, an aryl-azide to covalently link PPIs, a biotin handle to enrich bait PPIs, and a disulfide bridge to deplete bait proteins from mass spectrometry analysis

(figure 4.1). We used CLEC14A, a relatively uncharacterized endothelial cell surface protein, to demonstrate RIPID capabilities in identifying CLEC14A PPIs, and demonstrate CLEC14A involvement in endothelial morphogenesis.

We developed a novel proteomic workflow to identify protein-protein interactions and took advantage of our CLEC14A heparin deficient mutants to ask two questions. What are the CLEC14A protein binding partners, and which proteins are dependent on R161A? We were able to partition CLEC14A protein binding partners into two categories: R161A dependent versus independent. MMRN2 was reconfirmed as a CLEC14A protein binder, and independent of the mutation. However, heparin size dependently blocks MMRN2-CLEC14A interactions through steric interactions. NRP1 and ITIH5 is categorized as a heparin dependent binder. ITIH5 is an uncharacterized protein, and its role in endothelial biology is unknown. In contrast, NRP1 has been extensively characterized an important orphan receptor in mediating endothelial function by interacting with VEGFR and integrin pathways (211). The manner and location in which CLEC14A binds to NRP1 remains to be determined. If CLEC14A mediates NRP1 function is unknown.

At the time our CLEC14A studies began, MMRN2 was the only known CLEC14A binding protein. We decided to pursue and describe the CLEC14A proteome in endothelial cells. We first needed to identify and expand CLEC14A protein binding partners. Zanivan *et al* has performed CLEC14A immunoprecipitation followed by mass spectrometry to identify CLEC14A protein binding partners using antibodies targeting endogenous CLEC14A (129). They found MMRN2, LAMA4, LAMC1, VIM, HSPG2, FN1, and DBNN1 as the major binding partners of endogenous CLEC14A in HUVEC undergoing tube morphogenesis on Matrigel. In our approach, we used conjugated Sulfo-SBED to recombinant CLEC14A-325 to identify protein binding partners in HUVEC cultured on gelatin. We identified extracellular matrix proteins MMRN2, MMRN1, FN1, and LAMA4, and membrane bound cell surface proteins including ENG, PECAM1, ITGB1, ITGA5, and ITGA3 as part of the CLEC14A proteome (Table 4.1).

Between our differing methods, the only overlapping protein binding partner was MMRN2. RIPID is able to enrich more proteins than traditional immunoprecipitation.

To test if heparin binding can alter protein-protein interactions, we protected the CLEC14A heparin binding site from Sulfo-SBED conjugation. The CLEC14A heparin binding region contains a lysine residue that when conjugated, may block protein interactions from occurring at this site (chapter 2). To protect this region of the protein, we conjugated CLEC14A while on a heparin column (figure 4.3a). Upon doing so, we identified 14 proteins with including MMRN2, ITIH5, NRP1, TXLNA, and EFEMP1 (Table 4. 2). When compared to CLEC14A-R161A, ITIH5, NRP1, and TXLNA was no longer enriched suggesting that R161A is critical to bind these proteins (Figure 4.3e,f). The protein binding repertoire of CLEC14A changes when the heparin binding site is mutated suggesting that heparin not only stabilizes CLEC14A, but heparan sulfate acts to stabilize protein complexes and promote binding.

Sulfo-SBED is an ideal candidate cross-linking molecule for identifying PPIs. Protein conjugations are specific to lysine residues, and these lysine residues can be protected by blocking the site during conjugation to allow specific portions of the protein to be free from any steric hindrance the Sulfo-SBED may cause. When we conjugated Suflo-SBED to CLEC14A, the identified protein binding partners skewed based on whether we performed an in-solution conjugation or used a protect and label strategy to keep the heparin binding site open (Figure 4.1b and 4.3a). If we expanded the protect and label strategy to include other CLEC14A binding partners, Sulfo-SBED may provide valuable means to comprehensively map CLEC14A protein-interfaces or may be used a tool for general protein-protein interactions. By applying iterations of the RIPID workflow, one may be able to reassemble how the protein interactions are partitioned across the cell surface. For example, after identifying the CLEC14A PPIs such as MMRN2 and NRP1, Sulfo-SBED conjugation on CLEC14A PPIs would allow us to comprehensively reassemble protein complexes.

While there are conflicting mouse models of CLEC14A function in vivo, CLEC14A ultimately fine tunes pathological angiogenesis by regulating cell migration, cell matrix interactions and VEGFR signaling (115,116,126,128,133,182). Here we applied proteomics strategies to understand CLEC14A molecular functions by identifying CLEC14A protein binding partners. These binding partners are key players in endothelial migration and morphogenesis (Supplementary figure 4.3C), consistent with the observed functions of CLEC14A. Moreover, we identified a suggestive role for CLEC14A-heparan sulfate binding as our data suggests that heparan sulfate may act to partition CLEC14A PPIs.

With the identification of the CLEC14A protein binding site and CLEC14A protein binding partners, the CLEC14A-R161A mutation presents a unique opportunity to study CLEC14A biology. One may knockin the R161A mutation into mice to test for CLEC14A dependency on heparan sulfate, NRP1 and ITIH5 while maintaining CLEC14A dependence on MMRN2. Whereas previous models ask what happens to angiogenic systems in the absence CLEC14A, a *Clec14a-R161A* model asks what happens when replace a normal CLEC14A with a mutated CLEC14A. Does the R161 site allow CLEC14A to partition differing biological responses during angiogenesis?

Experimental Procedures

Sulfo-SBED Biotin Label Conjugation

200 µg of recombinant CLEC14A-325 was passed down a heparin Sepharose gravity column and washed with PBS. 1 mg of Sulfo-SBED Biotin Label (Thermo Scientific) was passed down the column and incubated for 30 minutes. After the column was washed with dPBS with glycine [0.1 M], pH7, followed by another PBS wash. The Sulfo-SBED-Biotin CLEC14A-325 conjugate was eluted with PBS 1M NaCl.

Sulfo-SBED Cross-linking proteomics.

HUVEC cells were cultured on gelatin with EGM-2. Cells were washed two times with M199, and then incubated with 30ug of Sulfo-SBED Biotin CLEC14A-325 or CLEC14A-325-R161A conjugates for 5 minutes at 37°C. After cells were washed with serum free media and placed on ice. Cells were cross-linked for 15 minutes using a UV Stratalinker 1800 (Stratagene). Cells were then washed 2 more times with PBS, lysed and harvested with RIPA buffer.

Sample preparation

Proteins were quantified by BCA assay (Thermo Scientific) as per manufacturer recommendations. A total of 3 mg of protein extract from each sample was used for affinity purification of biotinylated proteins. Affinity purification was carried out in a Bravo AssayMap platform (Agilent) using AssayMap streptavidin cartridges (Agilent). Briefly, cartridges were first primed with 50 mM ammonium bicarbonate, and then proteins were slowly loaded onto the streptavidin cartridge. Background contamination was removed by extensively washing the cartridges with 8M urea, 50 mM ammonium bicarbonate. Finally, cartridges were washed with Rapid digestion buffer (Promega, Rapid digestion buffer kit) and proteins were subjected to on-cartridge digestion with mass spec grade Trypsin/Lys-C Rapid digestion enzyme (Promega, Madison, WI) at 70ºC for 2h. Digested peptides were then desalted in the Bravo platform using AssayMap C18 cartridges and the organic solvent was removed in a SpeedVac concentrator prior to LC-MS/MS analysis.

LC-MS/MS analysis

Dried peptides were reconstituted with 2% acetonitrile, 0.1% formic acid, quantified by modified BCA peptide assay (Thermo Fisher Scientific) and analyzed by LC-MS/MS using a Proxeon EASY nanoLC system (Thermo Fisher Scientific) coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). Peptides were separated using an analytical C_{18} Acclaim PepMap column 0.075 x 500 mm, 2µm particles (Thermo Scientific) in a 93-min linear

gradient of 2-28% solvent B at a flow rate of 300nL/min. The mass spectrometer was operated in positive data-dependent acquisition mode. MS1 spectra were measured with a resolution of 70,000, an AGC target of 1e6 and a mass range from 350 to 1700 m/z. Up to 12 MS2 spectra per duty cycle were triggered, fragmented by HCD, and acquired with a resolution of 17,500 and an AGC target of 5e4, an isolation window of 1.6 m/z and a normalized collision energy of 25. Dynamic exclusion was enabled with duration of 20 sec.

Expression and purification of CLEC14A. We synthesized mammalian codon optimized (Genewiz) human CLEC14A-325 (residues 1-325) DNA and cloned them into pcDNA3.1A (+) (Invitrogen) with a C-terminal His_6 tag. To produce recombinant protein, HEK293F cells (1.5-2 $x10^6$ cells/ml) were transfected 2.5 μg/ml of plasmid DNA using PEI (9 μg/ml) in FreeStyle 293 Expression Medium (Gibco). One day later the cells were treated with valproic acid (2 mM), and 5 days after the initial transfection, the conditioned medium was mixed with cOmplete, EDTA-free Protease Inhibitor (Roche). Recombinant protein was purified by chromatography on a 1ml Ni^{2+} Sepharose 6 Fast Flow column (GE LifeSciences). Samples were loaded with FreeStyle 293 Expression Medium supplemented with 30 mM imidazole, washed with 30 mM Imidazole, 0.5 M NaCl, 20 mM Tris buffer (pH 7.4) and recombinant protein was eluted using 0.5 M NaCl, 0.3 M imidazole in 20 mM Tris buffer (pH 7.4). The protein was further purified by size exclusion chromatography (HiLoad 16/60 Superdex 200, prep grade. GE LifeSciences) in 0.3 M NaCl, 5% glycerol in 20 mM Tris buffer (pH 7.4). Mutant CLEC14A was purified in the same manner as wild-type CLEC14A.

Surface Plasmon Resonance. A Nicoya OpenSPR was used to generate binding curves for CLEC14A binding to NRP1, MMRN2, and EFEMP1. CLEC14A was immobilized on a Nicoya carboxyl sensor using Nicoya amine coupling kit. Carboxyl sensors were functionalized using 0.2 ml of a 1:1 mix of N-hydroxysuccinimide (0.1 M) and 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC, 0.4 M) before coupling to recombinant CLEC14A

under flow conditions. Ethanolamine was used to block remaining active sites on the chip. All surfaces were washed with SPR buffer (20 mM HEPES, 150 mM NaCl, 5mM $CaCl_2$, 17 mM NaN_3, 5 mM $MgCl_2$, 0.1% BSA, and 0.05% Tween20 pH 7.2) and regenerated with 20 mM HEPES buffer (pH 7.2) containing 3 M NaCl. Ligands were allowed to associate with the chip at a flow rate of 20 μl/min in SPR buffer for 4 min, and allowed to dissociate for 5 min. Regeneration buffer was used before each injection of ligand to clean the surface chip.

Endothelial Cell Sprouting Assay.

Matrigel (BD Biosciences) was added to each well of a 96-well plate and allowed to polymerize at 37°C for 30 minutes. HUVEC (5,000/well) at passage 4 to passage 6 were added to quadruplicate wells in 100 μl of EGM-2. After 8 hours, endothelial sprouting was viewed under phase-contrast light microscopy and measured using the angiogenesis analyzer plugin for ImageJ.

Cell Adhesion-Strength Assay

Adhesion strength was measured using a previously described method (212). Briefly, Glass coverslips (25 mm, Fisher Scientific) were sonicated in ethanol and pure water before coating with 10 μg/ml fibronectin (isolated from serum) for 1 hour at room temperature. HUVEC were then plated at a density of 50,000 cells per coverslip and cultured in EGM-2 for 24 hours. To spin, coverslips were attached to a custom built spinning disk device dipped in a temperature (37 ºC) controlled spinning buffer (PBS supplemented 4.5 g/L of dextrose). Cells were spun at defined angular velocities for 5 minutes and subjected to 4% PFA for fixation immediately after spinning. To calculate adhesion strength, whole coverslipps were imaged at 10x magnification on a Nikon (Melville, NY) Ti-S microscope (~1000 individual images stitiched together with Metamorph 7.6 software and custom macros) and analyzed using a custom-written MATLAB program. Cell densities as a function of radial position were used to calculate adhesion strength.

Immunofluorescence

Glass slides were coated with 10 μg/ml fibronectin (Sigma) for 1 hour at room temperature. HUVEC were plated at a density of 25,000 cells per well and cultured in EGM-2. Cells were fixed with 4% PFA 24 hours later, and washed twice with PBS. Cells were washed and blocked in 5% goat serum, 0.5% BSA, 0.1% Saponin in PBS for 1 hour.

CLEC14A Immunoprecipitation

CLEC14A was then incubated in the presence or absence of recombinant MMRN2 (ABNOVA) in PBST for 30 minutes at room temperature. CLEC14A-MMRN2 complexes were then pulled down with using magnetic Protein-G Dynabeads beads (Thermo) conjugated with THE™ His Tag antibody (GenScript). Beads were washed 3X with PBST, and protein was then eluted with citric acid, NuPage LDS buffer (Invitrogen). Western blot was used to detect the presence of CLEC14A and MMRN2. Blots were analyzed using ImageJ.

Author contributions.

D.R.S. contributed to the conception and design of the work, acquisition, analysis, and interpretation of data, and to the drafting of the manuscript; A.G.T. contributed to the design of the work, analysis, and interpretation of data; A.B. and C.D.P., contributed to acquisition and interpretation of the data; J.D.E. contributed to the conception and design of the work, analysis, and interpretation of data, and to the drafting of the manuscript.

Acknowledgements.

This work was supported in part by the Program of Excellence in Glycoscience P01 HL107150 and grant P01 HL131474 from the NIH (to J.D.E.), a supplement to P01 HL131474 (to D.R.S.).

Chapter 4, in part is being prepared for submission. Toledo, A.G., Painter, C.D., Esko J.D. I am the primary investigator and author of this paper.

Competing interests. The authors declare that thy have no conflict of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Table

Table 4.1: Cell Surface CLEC14A protein binding partners identified by RIPID using in-solution Sulfo-SBED Conjugation. LFQ intensities for cross-linked Suflo-SBED conjugated CLEC14A (XL-CLEC14A$_{ss}$), non-crosslinked, a R161A mutant CLEC14A, and cell lysates.

Gene name	Protein names	XL-CLEC14A$_{ss}$	XL-CLEC14A$_{ss}$-R161A	CLEC14A$_{ss}$	Lysate	MS/MS count
FN1	Fibronectin	24.7617	24.5278	23.8417	22.5982	155
CLEC14A	C-type lectin domain family 14 member A	25.6538	23.4663	23.2238	0	17
MMRN2	Multimerin-2	22.8852	23.3262	23.2886	20.9751	39
ITGB1	Integrin beta-1	23.1256	23.3229	17.9982	0	37
MCAM	Cell surface glycoprotein MUC18	22.889	22.918	19.1953	0	29
MMRN1	Multimerin-1	22.7914	22.8034	23.0822	20.5254	87
ENG	Endoglin	22.2537	22.2222	18.6344	0	19
ITGA5	Integrin alpha-5	22.4759	22.0206	18.0435	0	13
TAGLN2	Transgelin-2	21.9366	21.828	0	0	18
LAMC1	Laminin subunit gamma-1	21.4492	21.5248	20.8775	0	39
CLIC1	Chloride intracellular channel protein	21.6605	21.4464	19.1543	0	16
PECAM1	Platelet endothelial cell adhesion molecule	21.7436	21.3725	18.9957	0	25
ITGA2	Integrin alpha-2	21.162	21.2222	0	0	19
VWF	von Willebrand factor	21.6158	21.2166	23.234	20.2295	36
LAMB1	Laminin subunit beta-1	20.712	20.8263	19.6369	0	21
LGALS1	Galectin-1	22.3994	20.8205	19.6431	0	10
HEL-S-269	Protein disulfide-isomerase	22.2542	20.4333	18.5904	0	24
DRIP4	Programmed cell death 6-interacting protein	19.7289	20.169	19.2968	0	12
LAMA4	Laminin subunit alpha-4	19.9976	19.9532	19.3917	0	14
PLOD2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	20.1743	19.9339	0	0	7
UTRN	Utrophin	18.5501	19.9199	20.071	0	16
ICAM2	Intercellular adhesion molecule 2	0	19.8495	0	0	8
EPHA2	Ephrin type-A receptor 2	19.7837	19.7087	0	0	9
CD44	CD44 antigen	0	19.6702	0	0	8
NRP1	Neuropilin-1	20.2135	19.0115	0	0	4
COL18A1	Collagen alpha-1(XVIII) chain	18.9601	18.9253	19.527	0	5
MAGED2	Melanoma-associated antigen D2	18.9682	18.6949	18.4891	0	6
NOS3	Nitric oxide synthase	17.9095	18.4888	0	0	4
ITGB3	Integrin beta-3	0	16.8892	0	0	5
ACE	Angiotensin-converting enzyme	16.5563	0	0	0	3
ALCAM	CD166 antigen	16.8336	0	0	0	4
BASP1	Brain acid soluble protein 1	19.6579	0	0	0	3
C1QBP	Complement component 1 Q subcomponent-binding protein	18.5138	0	0	0	5
C9orf88	Niban-like protein 1	17.068	0	0	0	3
CD81	CD81 antigen	16.5596	0	0	0	4
CD9	CD9 antigen	19.7713	0	0	0	3
CDH5	Cadherin-5	16.9124	0	0	0	3

Table 4.1: Cell Surface CLEC14A protein binding partners identified by RIPID using in-solution Sulfo-SBED Conjugation, continued.

Gene name	Protein names	XL-CLEC14A$_{ss}$	XL-CLEC14A$_{ss}$-R161A	CLEC14A$_{ss}$	Lysate	MS/MS count
CLIC4	Chloride intracellular channel protein 4	17.8397	0	0	0	3
CTGF	Connective tissue growth factor	18.6097	0	0	0	4
EDIL3	EGF-like repeat and discoidin I-like domain-containing protein 3	18.7148	0	0	0	5
ITGA3	Integrin alpha-3	19.0901	0	0	0	5
ITGAV	Integrin alpha-V	17.9796	0	0	0	7
LTBP2	Latent-transforming growth factor beta-binding protein 2	17.8949	0	0	0	6
NID1	Nidogen-1	17.7413	0	0	0	4

Table 4.2: CLEC14A protein binding partners identified by RIPID using heparin protected conjugation. [a] Values are normalized LFQ intensities over the lysate.

Gene Name	Protein names	CLEC14A[a]	R161A[a]	MS/MS count
CLEC14A	C-type lectin domain family 14 member A	117.89	78.25	258.00
ITIH5	Inter-alpha-trypsin inhibitor heavy chain H5	7.32	1.10	11.00
NRP1	Neuropilin-1	4.18	1.11	19.00
TXLNA	Alpha-taxilin	3.89	0.93	11.00
TXNDC12	Thioredoxin domain-containing protein 12	3.29	0.94	13.00
MMRN2	Multimerin-2	2.90	2.01	95.00
NDUFB7	NADH dehydrogenase	2.81	1.76	26.00
GGH	Gamma-glutamyl hydrolase	2.49	1.36	74.00
MARCKS	Myristoylated alanine-rich C-kinase substrate	2.46	1.73	129.00
SAR1B	GTP-binding protein SAR1b	2.34	2.06	28.00
VAC14	Protein VAC14 homolog	2.24	1.27	9.00
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	2.20	1.77	114.00
NDUFB8	NADH dehydrogenase	2.15	1.82	34.00
MYL6	Myosin light polypeptide 6	2.13	2.08	159.00
MMGT1	Membrane magnesium transporter 1	2.07	1.50	21.00
TNFSF4	Tumor necrosis factor ligand superfamily member 4	1.92	1.51	16.00
CPNE1	Copine-1	1.85	1.87	1.01
NDUFB9	NADH dehydrogenase	1.84	1.45	0.79
MEMO1	Protein MEMO1	1.79	1.74	0.97
AGPAT5	1-acyl-sn-glycerol-3-phosphate acyltransferase epsilon	1.71	1.33	0.78
HLA-B	HLA class I histocompatibility antigen, B-38 alpha chain	1.63	1.06	0.65
NT5DC1	5'-nucleotidase domain-containing protein 1	1.63	1.10	0.67

Table 4.2: CLEC14A protein binding partners identified by RIPID using heparin protected conjugation, continued.

Gene Name	Protein names	CLEC14A[a]	R161A[a]	MS/MS count
NIPSNAP1	Protein NipSnap homolog 1	1.61	1.56	0.97
MGST3	Microsomal glutathione S-transferase 3	1.56	1.52	0.97
CTSC	Dipeptidyl peptidase 1	1.56	1.48	0.95
RTN4	Reticulon-4	1.55	1.57	1.01
CFL1	Cofilin-1	1.50	1.22	0.82
SLC25A24	Calcium-binding mitochondrial carrier protein SCaMC-1	1.48	1.11	0.75
SDHB	Succinate dehydrogenase iron-sulfur subunit, mitochondrial	1.46	1.37	0.94
RAB6A	Ras-related protein Rab-6A	1.44	1.32	0.92
ACAA2	3-ketoacyl-CoA thiolase, mitochondrial	1.42	1.26	0.89
MRPS6	28S ribosomal protein S6, mitochondrial	1.42	1.51	1.07
VCL	Vinculin	1.41	1.19	0.84
ACTG1	Actin, cytoplasmic 2	1.39	1.25	0.90
ATP6V1E1	V-type proton ATPase subunit E 1	1.27	1.27	1.00
TRAP1	Heat shock protein 75 kDa, mitochondrial	1.20	1.00	0.84
ATP13A1	Manganese-transporting ATPase 13A1	1.12	1.11	1.00
SQSTM1	Sequestosome-1	1.06	1.19	1.12
FLOT1	Flotillin-1	0.84	0.62	0.73
RNPS1	RNA-binding protein with serine-rich domain 1	0.79	2.08	2.63
IKBIP	Inhibitor of nuclear factor kappa-B kinase-interacting protein	0.78	0.77	0.98
UCHL5	Ubiquitin carboxyl-terminal hydrolase isozyme L5	0.69	0.56	0.81
NLN	Neurolysin, mitochondrial	0.67	0.59	0.89
EIF2AK2	Interferon-induced, double-stranded RNA-activated protein kinase	0.64	0.71	1.11
PDE12	2',5'-phosphodiesterase 12	0.62	0.65	1.04
SEC63	Translocation protein SEC63 homolog	0.58	0.68	1.17

Table 4.2: CLEC14A protein binding partners identified by RIPID using heparin protected conjugation, continued.

Gene Name	Protein names	CLEC14A[a]	R161A[a]	MS/MS count
CDK17	Cyclin-dependent kinase 17	0.56	0.57	1.02
DDX6	Probable ATP-dependent RNA helicase DDX6	0.56	0.51	0.91
CPT2	Carnitine O-palmitoyltransferase 2, mitochondrial	0.54	0.50	0.92
HNRNPR	Heterogeneous nuclear ribonucleoprotein R	0.52	1.01	1.95
SSRP1	FACT complex subunit SSRP1	0.52	1.20	2.33
PCBP2	Poly(rC)-binding protein 2	0.50	0.58	1.17
RBM34	RNA-binding protein 34	0.50	1.02	2.05
DHX30	Putative ATP-dependent RNA helicase DHX30	0.49	0.58	1.18
NADK2	NAD kinase 2, mitochondrial	0.47	0.55	1.17
DUSP3	Dual specificity protein phosphatase 3	0.45	0.36	0.81
NOP2	Probable 28S rRNA	0.43	1.03	2.38
H2AFY	Core histone macro-H2A.1	0.43	1.24	2.91
FAM120A	Constitutive coactivator of PPAR-gamma-like protein 1	0.40	0.67	1.69
RSL1D1	Ribosomal L1 domain-containing protein 1	0.38	0.93	2.45
SNF8	Vacuolar-sorting protein SNF8	0.38	0.43	1.13
HIST1H1D	Histone H1.3	0.36	1.95	5.39
HIST1H1B	Histone H1.5	0.34	1.85	5.48
HIST1H4A	Histone H4	0.33	1.00	3.01
TOP2B	DNA topoisomerase 2-beta	0.29	0.56	1.95
HIST1H1C	Histone H1.2	0.28	2.06	7.43
HIST1H1E	Histone H1.4	0.26	1.41	5.33
DDX18	ATP-dependent RNA helicase DDX18	0.26	0.65	2.50
NUMA1	Nuclear mitotic apparatus protein 1	0.12	0.32	2.76
RALY	RNA-binding protein Raly	0.05	0.67	12.48

Figures

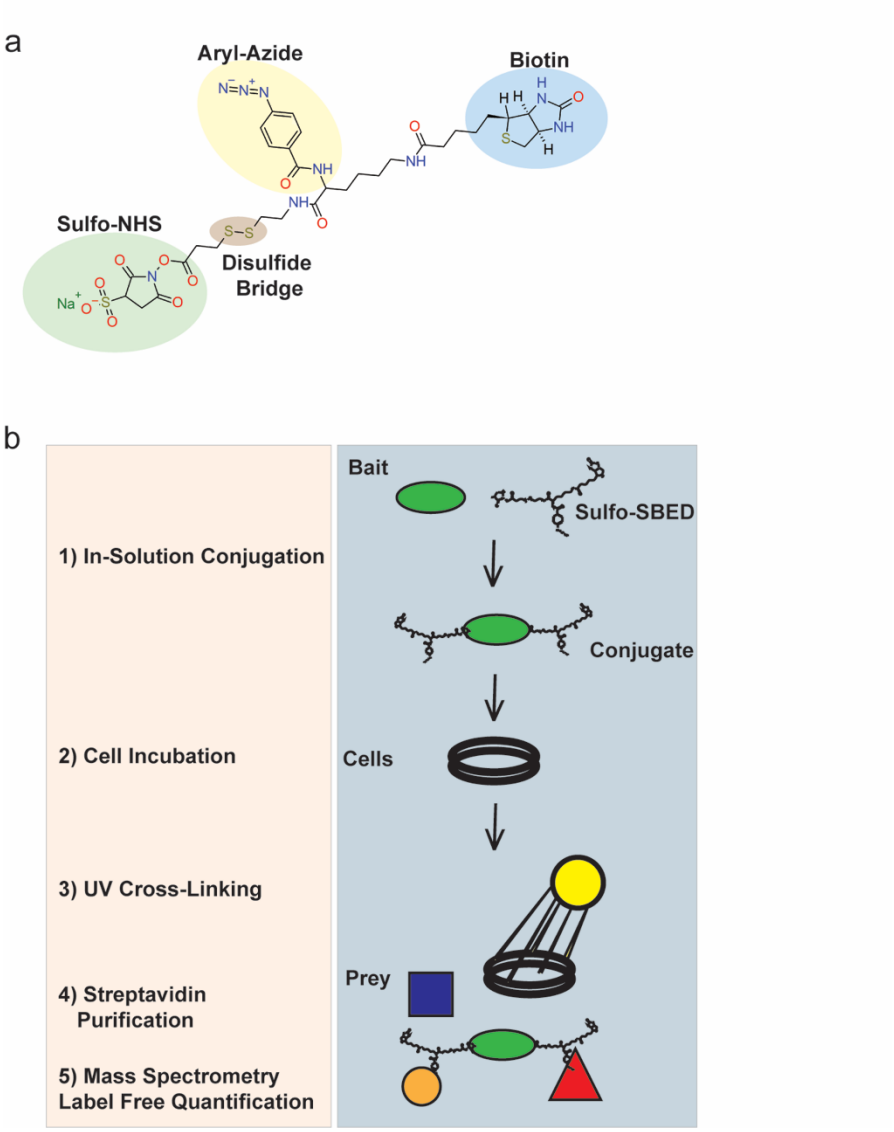


Figure 4.1: Suflo-SBED proteomic workflow. (a) Chemical Structure of Sulfo-SBED. (b) Sulfo-SBED proteomic workflow. Bait protein is conjugated in solution with excess Sulfo-Sbed. The conjugated protein is then incubated on cells and photo-crosslinked. Cells are then lysaed, and the tagged proteins are purified and enriched using streptavidin for mass spectrometry and label free quantification.

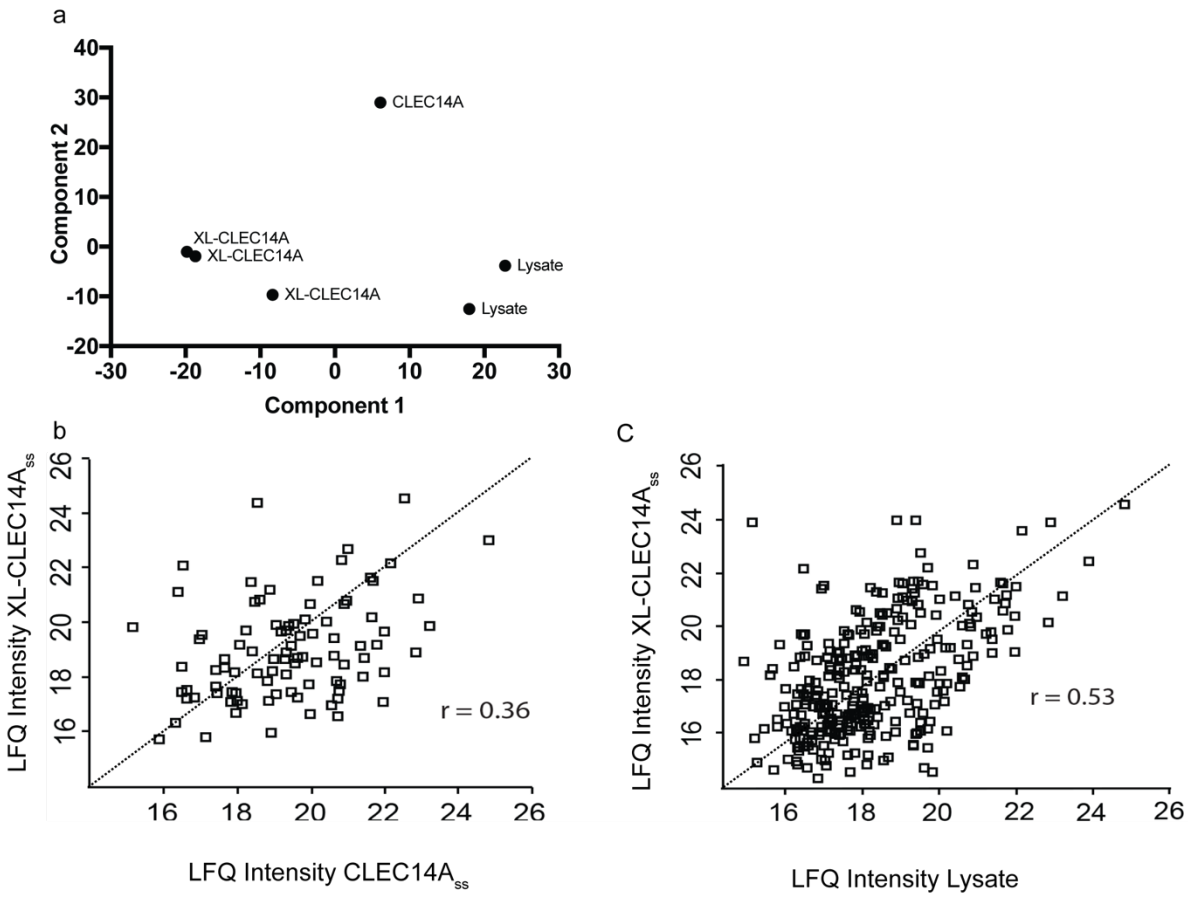


Figure 4.2: Validation of CLEC14A-Sulfo-SBED Workflow. (a) Principle component analysis of the variance due to differing experimental conditions from cross-linking CLEC14A-Sulfo-SBED to cultured HUVEC. (c) LFQ plot of proteomic hits from cross-linked CLEC14A$_{ss}$ versus HUVEC cell lysates. (d) LFQ plot of proteomic hits from cross-linked CLEC14A$_{ss}$ versus non-cross linked CLEC14A$_{ss}$.

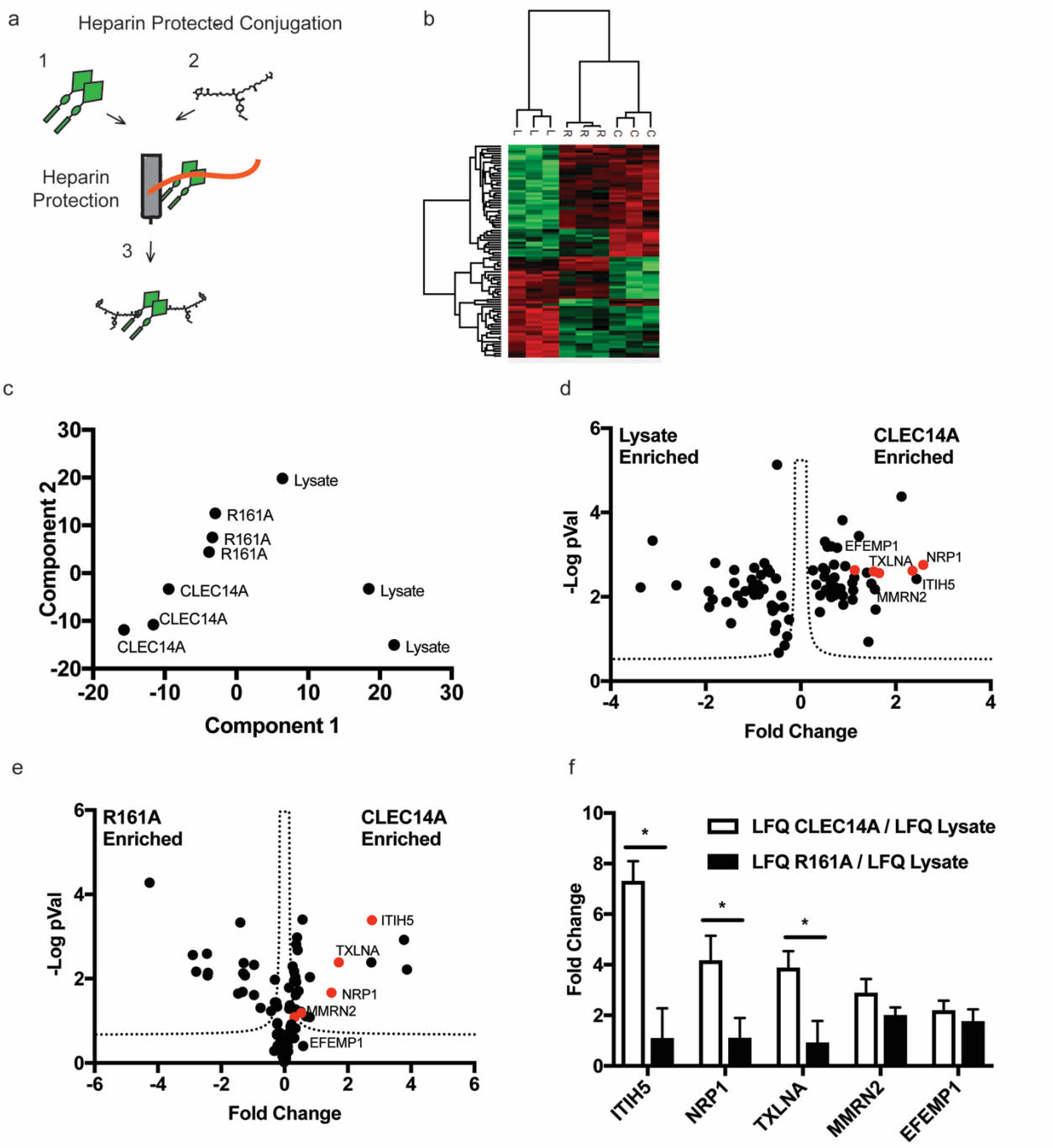


Figure 4.3: Heparin protected conjugation. (a) Protect and label Sulfo-SBED conjugation workflow. First CLEC14A or CLEC14A-R161A is bound to a heparin column to protect and hide lysine residues exposed at surface of CLEC14A. Then excess Sulfo-SBED is passed through the column to conjugate to lysines not involved in heparin binding. Newly conjugated CLEC14A-Sulfo-SBED and CLEC14A-R161A-Sulfo-SBED is then eluted from the heparin column to perform cross-linking proteomics on cultured HUVEC. (b) Heat map of proteins identified by RIPID. (c) Principle component analysis of experimental conditions. (d) Volcano plot of CLEC14A-Sulfo-SBED versus HUVEC lysate. (e) Volcano plot of CLEC14A-Sulfo-SBED versus CLEC14A-R161A-Sulfo-SBED. (f) quantitation of label free quantification of extracellular proteins identified by RIPID.

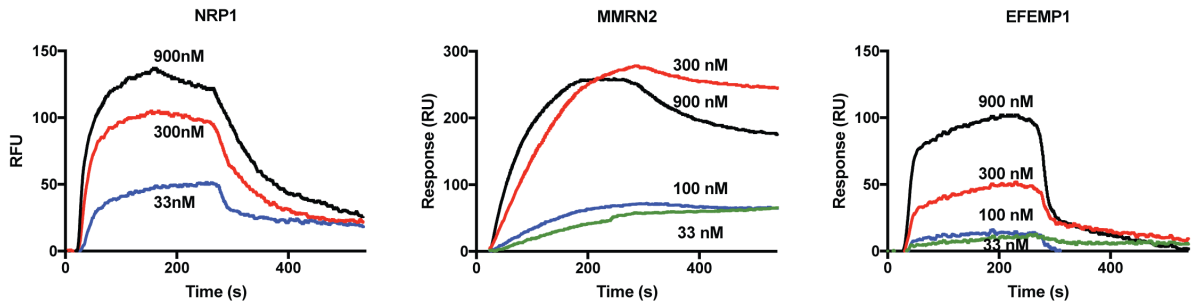


Figure 4.4: CLEC14A binds targets identified by RIPID. Surface plasmon resonance of CLEC14A binding proteins MMRN2 (a), NRP1 (b), EFEMP1 (c) flown over immobilized CLEC14A.

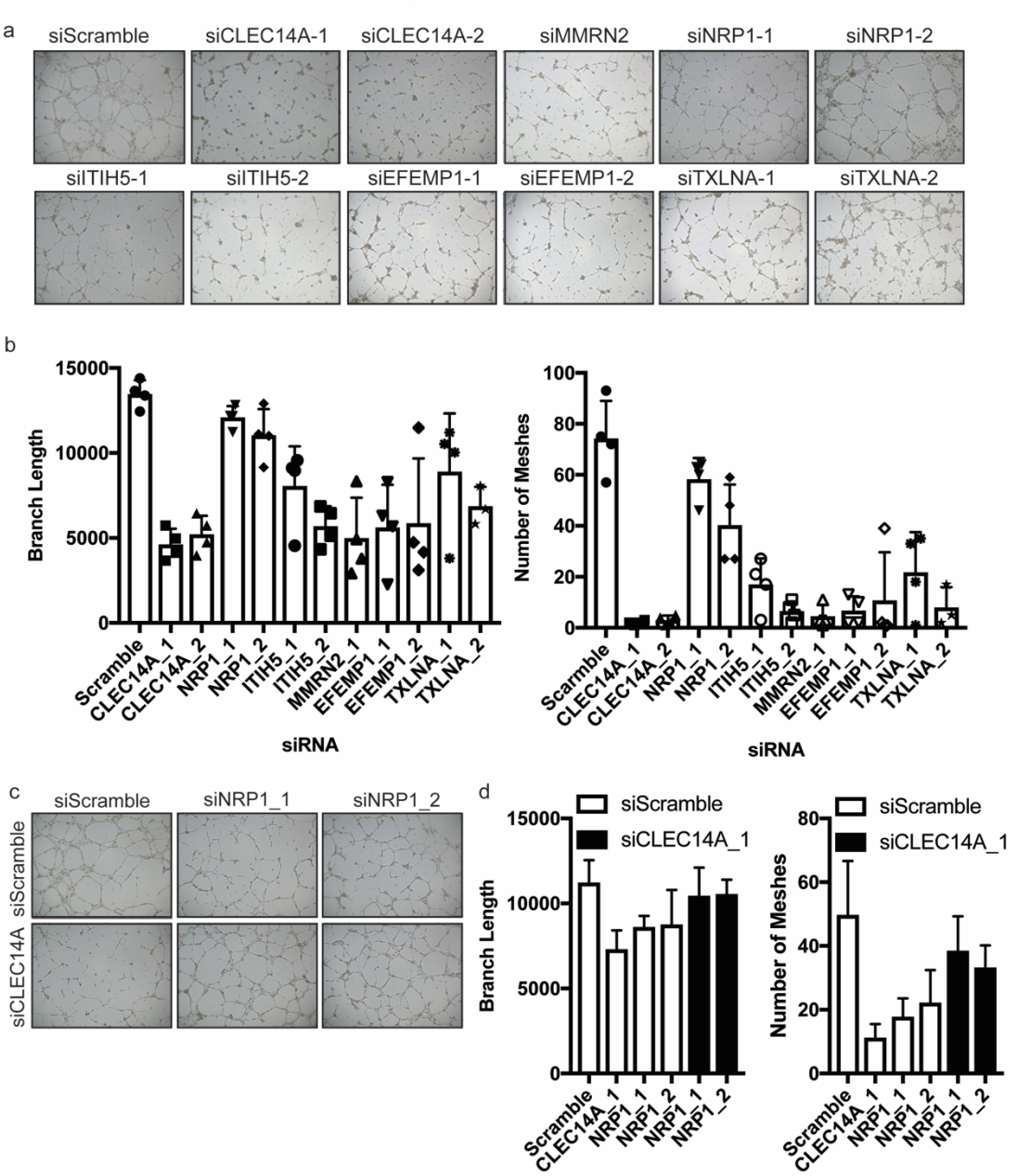
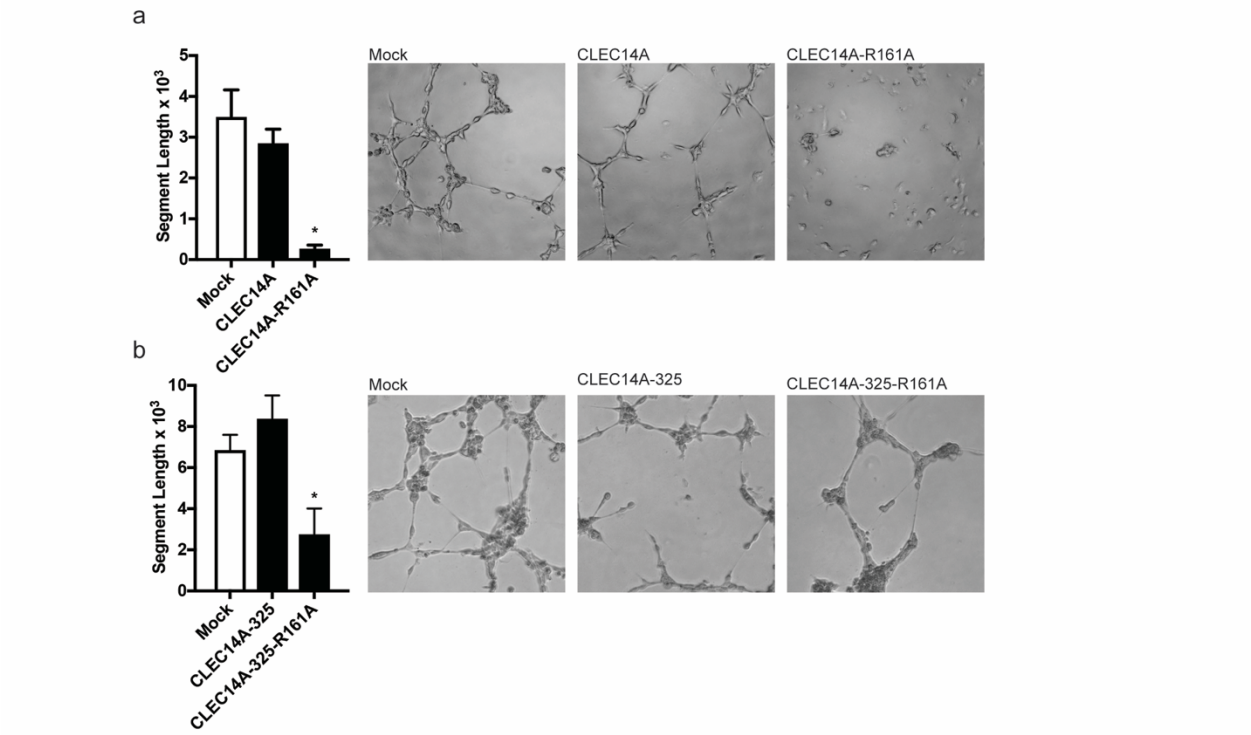
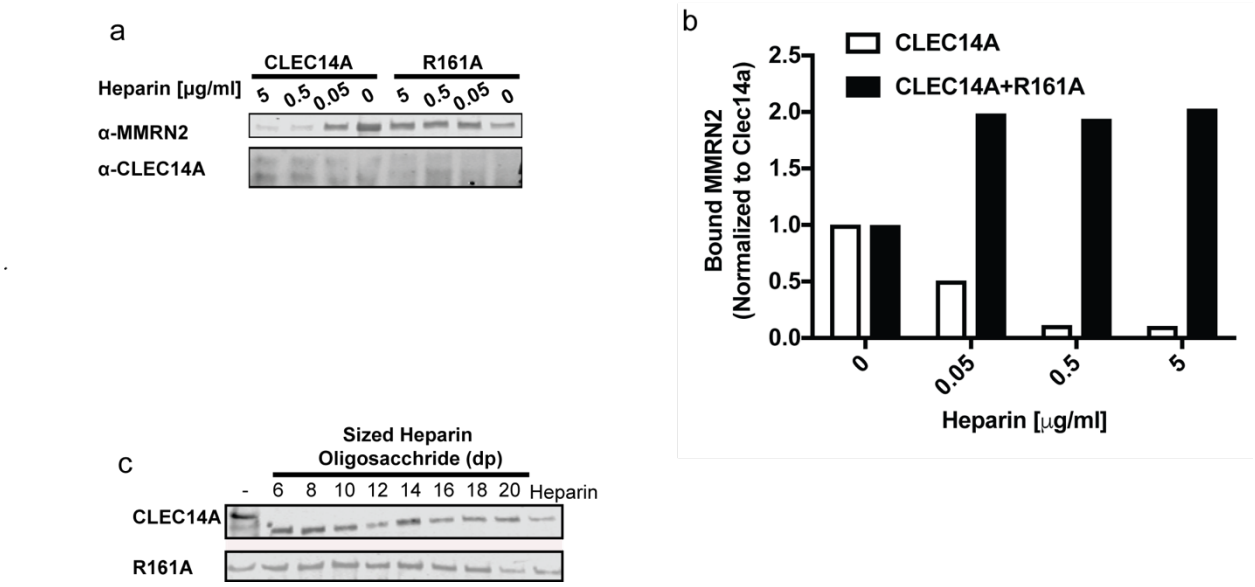


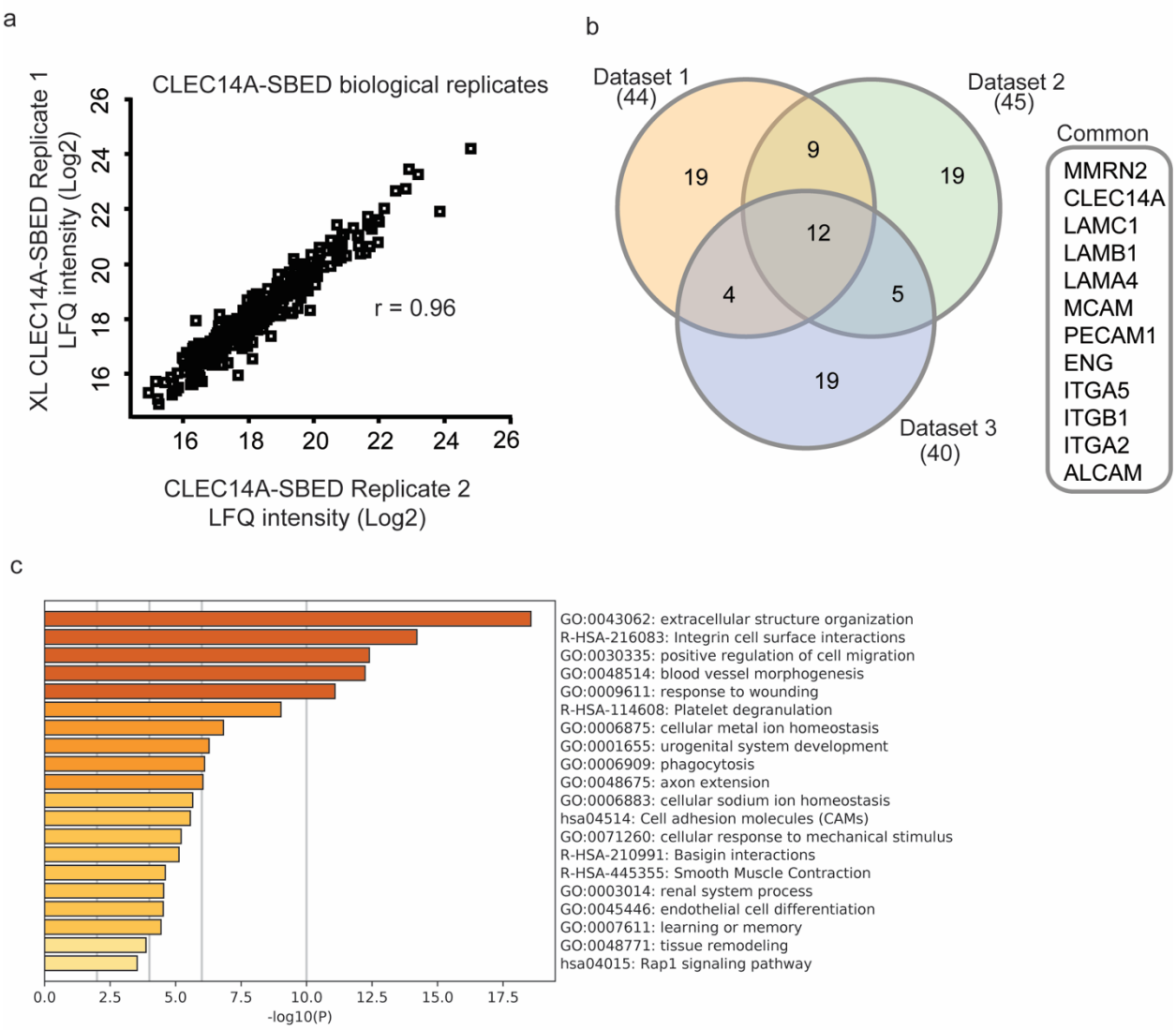
Figure 4.5: Sulfo-SBED targets genetically interact to regulate endothelial morphogenesis. Images of Endothelial Matrigel morphogenesis of HUVEC silenced with RIPID target genes (a), and quantification of the branch length and the number of meshes (b). Images of endothelial morphogenesis of HUVEC treated with siRNA targeting CLEC14A with or without treatment of siRNA targeting NRP1 (c), and quantification of branch length and mesh formation (d). All experiments were performed in triplicates.



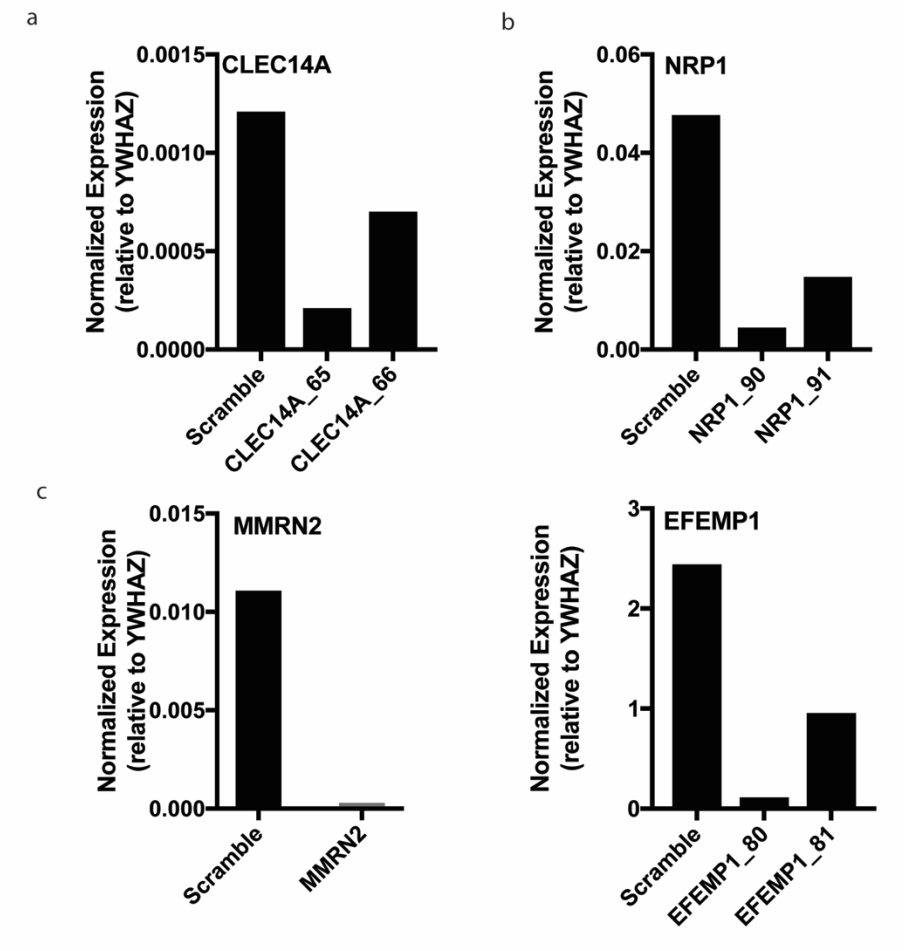
Supplemental Figure 4.1: CLEC14A-R161A has a dominant negative effect on endothelial tube formation in vitro. (a) Quantification of mean sprout length of HUVEC cultured on Matrigel transfected with GFP (control), full length CLEC14A, or full length CLEC14A-R161. (b), HUVEC were transfected cDNAs encoding the soluble ectomains CLEC14A-325 or CLEC14A-325-R161A. Representative phase contrast images of the cultures are shown.



Supplemental Figure 4.2: **Heparin blocks CLEC14A-MMRN2 binding.** (a) Incubation of MMRN2 with and without heparin to His-Tagged CLEC14A or CLEC14A-R161A immobilized to protein G beads conjugated with anti-His antibodies. (C) Binding of CLEC14A-MMRN2 with sized heparin oligomers (dp, degree of polymerization) as measured by western blot.



Supplemental Figure 4.3: RIPID Identifies CLEC14A protein binding partners. (a) LFQ intensities of identified proteins from cross-linked CLEC14A versus non-cross-linked CLEC14A. (b) overlap of identified CLEC14A-protein interactions from three datasets identified by RIPID. (c) Gene-ontology analysis of the CLEC14A-protein interactions identified by RIPID and protect and label proteomic workflows.



Supplemental Figure 4.4: siRNA Knockdown efficiencies in HUVEC. qPCR analysis of the gene expression of the respective siRNA target CLEC14A (a), NRP1 (b), MMRN2 (c) and EFEMP1 (d) in HUVEC. Each gene was targeted with two siRNAs.

References

1. Reitsma, S., Slaaf, D. W., Vink, H., van Zandvoort, M. A., and oude Egbrink, M. G. (2007) The endothelial glycocalyx: composition, functions, and visualization. *Pflugers Arch* **454**, 345-359

2. Mehta, V., and Trinkle-Mulcahy, L. (2016) Recent advances in large-scale protein interactome mapping. *F1000Res* **5**

3. Liu, X., Salokas, K., Tamene, F., Jiu, Y., Weldatsadik, R. G., Ohman, T., and Varjosalo, M. (2018) An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat Commun* **9**, 1188

4. Tian, Y., Jacinto, M. P., Zeng, Y., Yu, Z., Qu, J., Liu, W. R., and Lin, Q. (2017) Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. *J Am Chem Soc* **139**, 6078-6081

5. Wang, W., Li, T., Felsovalyi, K., Chen, C., Cardozo, T., and Krogsgaard, M. (2014) Quantitative analysis of T cell receptor complex interaction sites using genetically encoded photo-cross-linkers. *ACS Chem Biol* **9**, 2165-2172

6. Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Lin, S., Meng, R., Wang, C., and Chen, P. R. (2016) Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label. *Nat Commun* **7**, 12299

7. Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Xie, X., Lin, S., Hao, Z., Zheng, H., and Chen, P. R. (2017) Genetically encoded releasable photo-cross-linking strategies for studying protein-protein interactions in living cells. *Nat Protoc* **12**, 2147-2168

8. Zanivan, S., Maione, F., Hein, M. Y., Hernandez-Fernaud, J. R., Ostasiewicz, P., Giraudo, E., and Mann, M. (2013) SILAC-based proteomics of human primary endothelial cell morphogenesis unveils tumor angiogenic markers. *Mol Cell Proteomics* **12**, 3599-3611

9. Jang, J., Kim, M. R., Kim, T. K., Lee, W. R., Kim, J. H., Heo, K., and Lee, S. (2017) CLEC14a-HSP70-1A interaction regulates HSP70-1A-induced angiogenesis. *Sci Rep* **7**, 10666

10. Rho, S. S., Choi, H. J., Min, J. K., Lee, H. W., Park, H., Park, H., Kim, Y. M., and Kwon, Y. G. (2011) Clec14a is specifically expressed in endothelial cells and mediates cell to cell adhesion. *Biochem Biophys Res Commun* **404**, 103-108

11. Mura, M., Swain, R. K., Zhuang, X., Vorschmitt, H., Reynolds, G., Durant, S., Beesley, J. F., Herbert, J. M., Sheldon, H., Andre, M., Sanderson, S., Glen, K., Luu, N. T., McGettrick, H. M., Antczak, P., Falciani, F., Nash, G. B., Nagy, Z. S., and Bicknell, R. (2012) Identification and angiogenic role of the novel tumor endothelial marker CLEC14A. *Oncogene* **31**, 293-305

12. Noy, P. J., Lodhia, P., Khan, K., Zhuang, X., Ward, D. G., Verissimo, A. R., Bacon, A., and Bicknell, R. (2015) Blocking CLEC14A-MMRN2 binding inhibits sprouting angiogenesis and tumour growth. *Oncogene* **34**, 5821-5831

13. Lee, S., Rho, S. S., Park, H., Park, J. A., Kim, J., Lee, I. K., Koh, G. Y., Mochizuki, N., Kim, Y. M., and Kwon, Y. G. (2017) Carbohydrate-binding protein CLEC14A regulates VEGFR-2- and VEGFR-3-dependent signals during angiogenesis and lymphangiogenesis. *J Clin Invest* **127**, 457-471

14. Noy, P. J., Swain, R. K., Khan, K., Lodhia, P., and Bicknell, R. (2016) Sprouting angiogenesis is regulated by shedding of the C-type lectin family 14, member A (CLEC14A) ectodomain, catalyzed by rhomboid-like 2 protein (RHBDL2). *FASEB J* **30**, 2311-2323

15. Khan, K. A., Naylor, A. J., Khan, A., Noy, P. J., Mambretti, M., Lodhia, P., Athwal, J., Korzystka, A., Buckley, C. D., Willcox, B. E., Mohammed, F., and Bicknell, R. (2017) Multimerin-2 is a ligand for group 14 family C-type lectins CLEC14A, CD93 and CD248 spanning the endothelial pericyte interface. *Oncogene* **36**, 6097-6108

16. Niland, S., and Eble, J. A. (2019) Neuropilins in the Context of Tumor Vasculature. *Int J Mol Sci* **20**

17. Kim, T. K., Park, C. S., Jang, J., Kim, M. R., Na, H. J., Lee, K., Kim, H. J., Heo, K., Yoo, B. C., Kim, Y. M., Lee, J. W., Kim, S. J., Kim, E. S., Kim, D. Y., Cha, K., Lee, T. G., and Lee, S. (2018) Inhibition of VEGF-dependent angiogenesis and tumor angiogenesis by an optimized antibody targeting CLEC14a. *Mol Oncol* **12**, 356-372

18. Ki, M. K., Jeoung, M. H., Choi, J. R., Rho, S. S., Kwon, Y. G., Shim, H., Chung, J., Hong, H. J., Song, B. D., and Lee, S. (2013) Human antibodies targeting the C-type lectin-like domain of the tumor endothelial cell marker clec14a regulate angiogenic properties in vitro. *Oncogene* **32**, 5449-5457

19. Fuhrmann, A., Banisadr, A., Beri, P., Tlsty, T. D., and Engler, A. J. (2017) Metastatic State of Cancer Cells May Be Indicated by Adhesion Strength. *Biophys J* **112**, 736-745

CHAPTER 5: CONCLUSION

Heparan Sulfate Proteoglycans

All animal cells express heparan sulfate proteoglycans where they are either embedded in the cell membrane or secreted into the extracellular matrix to preform key cellular and physiological processes (1). Only 17 proteins are known to include heparan sulfate attachment sites. At the cell surface, heparan sulfate proteoglycans interact with both soluble and extracellular matrix and ectodomains of other cell surface proteins to either tether, promote oligomerization, induce allosteric interactions, promote stability or act as a scaffold (2).

To date, there are over 500 known heparan sulfate binding proteins (HSBPs) (Chapter 1). Many of these proteins were identified using heparin affinity chromatography coupled with mass spectrometry. This approach led to identify and purify unknown growth factors that drove key biological processes such as angiogenesis (3-13). However, later the focus shifted to identifying HSPBs in attempts to understand which proteins heparan sulfate interacts with to elicit its function. While heparin affinity chromatography screening was successful in identifying many of the growth factors we know today, it lacked the ability to identify membrane bound proteins. The problems arose through the biophysical nature of heparin (and other glycosaminoglycans) as a large, highly charged, multivalent polysaccharide. Heparin columns can become saturated with highly abundant and charged intracellular proteins such as DNA and RNA binding proteins, obscuring and reducing the amount of lowly abundant membrane proteins from both heparin affinity purification and mass spectrometry identification.

To circumvent this problem, we created a novel proteomic work flow termed LPHAMS (Limited proteolysis heparin affinity mass spectrometry). The fundamental principle of LPHAMs is to use limited proteolysis on intact living cells to enrich, isolate, purify and partition ectodomains from cells to flow over a heparin column (Chapter 2). Upon doing so, LPHAMS led to the identification of 75 HSBPs and has the advantage to identify the domain that binds to heparin. Using LPHAMS on endothelial cells, we identified 75 HSPBs, and 20 membrane bound

HSPBs from three endothelial cell types (mouse lung endothelial cell, mouse brain endothelial cell, and human umbilical vein endothelial cell) using predominantly two proteases, Proteinase K and Chymotrypsin (Chapter 2).

Proteases have a wide variety of specificities, reactivities and require a unique set of parameters for optimal activity (14,15). Proteinase K is a broad-spectrum protease that prefers cleaving peptide bonds of aliphatic and aromatic amino acids whereas chymotrypsin prefers large hydrophobic amino acids. However, by subjecting proteases to sub-optimal conditions, proteases may ignore their sequence preferences and will cleave exposed hinges or loops. The rate of proteolysis on individual proteins will based on their susceptibility to proteases and accessibility to exposed hinges or loops. One may overcome this by preforming a comprehensive analysis using an expanded repertoire of proteases with various activities may open new windows to identify HSBPs.

LPHAMS can be modified to include other GAGS such as chondroitin sulfate, dermatan sulfate and heparan sulfate as well as other glycans. It would be interesting to apply LPHAMS to identify chondroitin sulfate binding partners but also compare and footprint the cell surface protein repertoire of cells towards their preferred glycosaminoglycan partners. LPHAMs may provide a new avenue to identify novel glycosaminoglycan protein binding partners in a variety of cells.

CLEC14A

In Chapter 2, we designed a proteomic workflow to screen and identify membrane bound heparan sulfate binding proteins. A candidate heparan sulfate binding protein was CLEC14a. To this end, we investigated the biophysically interactions of heparin or heparan sulfate with CLEC14A (Chapter 2) and whether heparin may alter CLEC14A protein interactions (Chapter 4).

CLEC14A binds to heparin with at high affinities. Heparin binding to CLEC14A is size dependent and the binding site is equivalent to a heparin length of twelve saccharides. Common properties of heparin-protein interactions include the ability of heparin to oligomerize or stabilize a protein. For CLEC14A, we first needed to characterize its behavior. We asked whether CLEC14A behaves as a monomer or a multimer. To achieve this, we used size exclusion chromatography coupled with multi-angle light scattering to determine that the endomucin domain of CLEC14A induces trimerization of soluble CLEC14A. The addition of heparin of various lengths to either monomeric or trimer CLEC14A suggests a model where CLEC14A tethers to heparan sulfate like a string on a bead. In terms of thermal stability, CLEC14A melts at 55 ºC and the addition of heparin causes CLEC14A to become significantly thermal resistant by 10 ºC. Biophysically, heparin binds CLEC14A with high affinities to tether and string CLEC14A while also increasing CLEC14A stability.

The C-Type Lectin superfamily controls a broad spectrum of physiological processes (16). They all contain a C-type lectin domain (CTLD), and are further stratified into families based on their domain structures. The C-Type Lectin XIV family consists of four proteins: thrombomodulin (TM), endosialin (CD248), and complement component C1q receptor (CD93), and C-type lectin 14a (CLEC14A). These proteins are single pass transmembrane proteins compromised of a N-terminal CTLD, epidermal growth factor (EGF) domain(s), a mucin-like region rich in serine and threonine, a transmembrane domain and a short cytoplasmic tail. In regards to the vasculature, TM, CD93 and CLEC14A are expressed and localized to endothelial cells whereas CD248 localizes to pericytes (17-22). CD248, CD93, and CLEC14A have been shown to play a role in tumor angiogenesis (20,22-29).

Another manner in which CLEC14A controls blood vessel maturation during tumor angiogenesis is through VEGFR signaling (23,30). Within endothelial cells, acts in vascular homeostasis by fine tuning VEGFR2/3 signaling by forming protein complexes with VEGFR3 (23). Removal of CLEC14A leads to a decrease VEGFR3 signaling, while increasing VEGFR2

signaling, and the overexpression of CLEC14A has the opposite effect (23). What CLEC14A is doing in the VEGFR3 complex is unknown. How CLEC14A biophysically interacts with VEGFR3 is unknown.

Another common feature of heparin-protein interactions is to modulate protein-protein interactions by either promoting or blocking interactions. At the time our CLEC14A studies began, MMRN2 was the only known CLEC14A binding protein. We decided to pursue and describe the CLEC14A proteome in endothelial cells. To do so, we developed a novel proteomic workflow to identify protein-protein interactions and took advantage of our CLEC14A heparin deficient mutants to ask two questions. What are the CLEC14A protein binding partners, and which proteins are dependent on heparin or bind to the heparin binding site? We were able to partition CLEC14A protein binding partners into two categories: heparin dependent versus independent. MMRN2 was reconfirmed as a CLEC14A protein binder, and independent of the mutation. However, heparin size dependently blocks MMRN2-CLEC14A interactions through steric interactions. NRP1 and ITIH5 is categorized as a heparin dependent binder. ITIH5 is an uncharacterized protein, and its role in endothelial biology is unknown. In contrast, NRP1 has been extensively characterized an important orphan receptor in mediating endothelial function. The manner and location in which CLEC14A binds to NRP1 remains to be determined. If CLEC14A mediates NRP1 function is unknown.

With the identification of the CLEC14A protein binding site and CLEC14A protein binding partners, the R161A CLEC14A mutant presents a unique opportunity to study CLEC14A biology. A R161A knockin mutation in mice creates an unique model. Whereas previous models ask what happens to angiogenic systems in the absence CLEC14A, a *Clec14a-R161A* model asks what happens when replace a normal CLEC14A with a faulty CLEC14A. Would a *Clec14a-R161A* mice reveal the dependence CLEC14A on heparan sulfate, ITIH5, and NRP1 binding independent of affecting MMRN2 interactions? Does the R161 site allow CLEC14A to partition differing biological responses in the context of VEGFR signaling against Cell-Cell interactions?

Sulfo-SBED

Cells respond to external cues through their cell surface receptors to sense the environment and initiate a response to drive key biological processes. To do so, the ectodomains of cell surface proteins interact with adjacent cell surface proteins whether they be membrane bound, secreted factors or matrix components to elicit a response. Identifying protein-protein interactions (PPIs) are key to understanding many biological processes.

While there have been many advances in methodology to identify PPIs, these methods have been employed to identify intracellular protein interactions. Moreover, many of these methods perform well to identify strong and abundance protein interactions but are not ideal to capture weak or transient interactions. We decided to develop a novel proteomic work flow to by-pass problems many preexisting proteomic workflows encounter using chemical cross-linking (Chapter 4).

Sulfo-SBED is an ideal candidate cross-linking molecule for identifying PPIs. Protein conjugations are specific to lysine residues, and these lysines residues can be protected by blocking the site during conjugation to allow specific portions of the protein to be free from any steric hindrance the Sulfo-SBED may cause. When we conjugated Suflo-SBED to CLEC14A, the identified protein binding partners skewed based on whether we performed an in-solution conjugation or used a protect and label strategy to keep the heparin binding site open (Chapter 4). If we expanded the protect and label strategy to include other CLEC14A binding partners, Sulfo-SBED may provide valuable means to comprehensively map CLEC14A protein-interfaces or may be used a tool for general protein-protein interactions. Moreover, with the availability of CLEC14A monoclonal blocking antibodies with various biological responses (24,30,31), the implementation of these antibodies would allow us to skew CLEC14A protein interactions.

The proximity based labeling of Suflo-SBED allows the identification of protein binding partners up to 20Å. This allows for identification of direct protein interactions and leads to the

possibility and opportunity of identifying and reassembling large protein complexes. While we identified multiple CLEC14A protein binding partners, we are unable to decipher how these protein interactions are occurring. We established binary interactions. However, by applying iterations of the Suflo-SBED workflow, one may be able to reassemble how the protein interactions are partitioned across the cell surface. For example, after identifying the CLEC14A PPIs such as MMRN2 and NRP1, Sulfo-SBED conjugation on CLEC14A PPIs would allow us to comprehensively reassemble protein complexes.

Many biological processes such as angiogenesis are kinetic and dynamic in nature. During angiogenesis cells migrate, cell-cell contacts break down and resemble and the extracellular matrix is being remodeled. A protein may be involved in one or many of these processes but deciphering how they are involved proves difficult. For CLEC14A, a time course to identify how CLEC14A PPIs change over time in regard to angiogenesis and inflammation would aid in understanding on how CLEC14A regulates a complex process.

Heparan Sulfate and Chondroitin sulfate E are both glycosaminoglycans that interact with CLEC14A (Chapter 2). Genetic or enzymatic manipulation of the glycosaminoglycans on endothelial cells to apply RIPID to would provide another means to understand how these glycans dynamically mediate protein interactions. Moreover, removal of these glycosaminoglycans would provide a means to test if glycosaminoglycans act to partition PPI.

In regards to Carbohydrate-protein interactions (CPIs), efforts have been few and the methods less widespread at the proteomics level (32). Most efforts have used AP-MS to identify protein binding partners of a particular class of glycan or carbohydrate moiety. In regard to O- and N-linked glycans, notable efforts have been made in identifying glycan binding proteins (GBP) using XL-MS. Glycan-protein interactions have wide and varied binding affinities and specificities, and valences making cross-linking an ideal method to identify GBPs. Bioorthogonal monosaccharides were designed to be taken up by cells and incorporate into glycan chains.

Another approach has been to take artificial glycopolymers with cross-linking moieties to identify GBPs (33). However, efforts in the field of the glycosaminoglycan have been limited to AP-MS.

Approximately 1-5% of heparin or heparan sulfate is comprised of de-acetlyated N-acetyl-glucosamine with an ammonium group ($-NH_3^+$) corresponding to a ratio of 1 per 20-50 saccharides. Sulfo-SBED can be conjugated at the ($-NH_3^+$) sites along the heparan sulfate chain. Furthermore, the frequency at which modifications may occur creates an unique opportunity to implement Sulfo-SBED to identify HSPBs as the heparin binding site of many HSBPs typically spans 6-20 saccharides. Once Suflo-SBED has been established, heparan sulfate finger printing can then be implemented to determine how heparan sulfate of various structures lead to differing proteomes.

References

1. Sarrazin, S., Lamanna, W. C., and Esko, J. D. (2011) Heparan sulfate proteoglycans. *Cold Spring Harb Perspect Biol* **3**

2. Xu, D., and Esko, J. D. (2014) Demystifying heparan sulfate-protein interactions. *Annu Rev Biochem* **83**, 129-157

3. Maciag, T., Mehlman, T., Friesel, R., and Schreiber, A. B. (1984) Heparin binds endothelial cell growth factor, the principal endothelial cell mitogen in bovine brain. *Science* **225**, 932-935

4. Shing, Y., Folkman, J., Sullivan, R., Butterfield, C., Murray, J., and Klagsbrun, M. (1984) Heparin affinity: purification of a tumor-derived capillary endothelial cell growth factor. *Science* **223**, 1296-1299

5. Klagsbrun, M., and Shing, Y. (1985) Heparin affinity of anionic and cationic capillary endothelial cell growth factors: analysis of hypothalamus-derived growth factors and fibroblast growth factors. *Proc Natl Acad Sci U S A* **82**, 805-809

6. Shing, Y., Folkman, J., Haudenschild, C., Lund, D., Crum, R., and Klagsbrun, M. (1985) Angiogenesis is stimulated by a tumor-derived endothelial cell growth factor. *J Cell Biochem* **29**, 275-287

7. Hauschka, P. V., Mavrakos, A. E., Iafrati, M. D., Doleman, S. E., and Klagsbrun, M. (1986) Growth factors in bone matrix. Isolation of multiple types by affinity chromatography on heparin-Sepharose. *J Biol Chem* **261**, 12665-12674

8. Brigstock, D. R., Steffen, C. L., Kim, G. Y., Vegunta, R. K., Diehl, J. R., and Harding, P. A. (1997) Purification and characterization of novel heparin-binding growth factors in uterine secretory fluids. Identification as heparin-regulated Mr 10,000 forms of connective tissue growth factor. *J Biol Chem* **272**, 20275-20282

9. Ori, A., Wilkinson, M. C., and Fernig, D. G. (2011) A systems biology approach for the investigation of the heparin/heparan sulfate interactome. *J Biol Chem* **286**, 19892-19904

10. Xu, D., Young, J., Song, D., and Esko, J. D. (2011) Heparan sulfate is essential for high mobility group protein 1 (HMGB1) signaling by the receptor for advanced glycation end products (RAGE). *J Biol Chem* **286**, 41736-41744

11. Zhang, Y., Jiang, N., Jia, B., Chang, Z., Zhang, Y., Wei, X., Zhou, J., Wang, H., Zhao, X., Yu, S., Song, M., Tu, Z., Lu, H., Yin, J., Wahlgren, M., and Chen, Q. (2014) A comparative study on the heparin-binding proteomes of Toxoplasma gondii and Plasmodium falciparum. *Proteomics* **14**, 1737-1745

12. Hsiao, F. S., Sutandy, F. R., Syu, G. D., Chen, Y. W., Lin, J. M., and Chen, C. S. (2016) Systematic protein interactome analysis of glycosaminoglycans revealed YcbS as a novel bacterial virulence factor. *Sci Rep* **6**, 28425

13. Thacker, B. E., Seamen, E., Lawrence, R., Parker, M. W., Xu, Y., Liu, J., Vander Kooi, C. W., and Esko, J. D. (2016) Expanding the 3-O-Sulfate Proteome--Enhanced Binding of Neuropilin-1 to 3-O-Sulfated Heparan Sulfate Modulates Its Activity. *ACS Chem Biol* **11**, 971-980

14. Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis. *Acta Biochim Pol* **51**, 299-321

15. Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta* **1382**, 191-206

16. Zelensky, A. N., and Gready, J. E. (2005) The C-type lectin-like domain superfamily. *FEBS J* **272**, 6179-6217

17. Isermann, B., Hendrickson, S. B., Zogg, M., Wing, M., Cummiskey, M., Kisanuki, Y. Y., Yanagisawa, M., and Weiler, H. (2001) Endothelium-specific loss of murine thrombomodulin disrupts the protein C anticoagulant pathway and causes juvenile-onset thrombosis. *J Clin Invest* **108**, 537-546

18. MacFadyen, J., Savage, K., Wienke, D., and Isacke, C. M. (2007) Endosialin is expressed on stromal fibroblasts and CNS pericytes in mouse embryos and is downregulated during development. *Gene Expr Patterns* **7**, 363-369

19. MacFadyen, J. R., Haworth, O., Roberston, D., Hardie, D., Webster, M. T., Morris, H. R., Panico, M., Sutton-Smith, M., Dell, A., van der Geer, P., Wienke, D., Buckley, C. D., and Isacke, C. M. (2005) Endosialin (TEM1, CD248) is a marker of stromal fibroblasts and is not selectively expressed on tumour endothelium. *FEBS Lett* **579**, 2569-2575

20. Mura, M., Swain, R. K., Zhuang, X., Vorschmitt, H., Reynolds, G., Durant, S., Beesley, J. F., Herbert, J. M., Sheldon, H., Andre, M., Sanderson, S., Glen, K., Luu, N. T., McGettrick, H. M., Antczak, P., Falciani, F., Nash, G. B., Nagy, Z. S., and Bicknell, R. (2012) Identification and angiogenic role of the novel tumor endothelial marker CLEC14A. *Oncogene* **31**, 293-305

21. Rho, S. S., Choi, H. J., Min, J. K., Lee, H. W., Park, H., Park, H., Kim, Y. M., and Kwon, Y. G. (2011) Clec14a is specifically expressed in endothelial cells and mediates cell to cell adhesion. *Biochem Biophys Res Commun* **404**, 103-108

22. Langenkamp, E., Zhang, L., Lugano, R., Huang, H., Elhassan, T. E., Georganaki, M., Bazzar, W., Loof, J., Trendelenburg, G., Essand, M., Ponten, F., Smits, A., and Dimberg, A. (2015) Elevated expression of the C-type lectin CD93 in the glioblastoma vasculature regulates cytoskeletal rearrangements that enhance vessel function and reduce host survival. *Cancer Res* **75**, 4504-4516

23. Lee, S., Rho, S. S., Park, H., Park, J. A., Kim, J., Lee, I. K., Koh, G. Y., Mochizuki, N., Kim, Y. M., and Kwon, Y. G. (2017) Carbohydrate-binding protein CLEC14A regulates VEGFR-2- and VEGFR-3-dependent signals during angiogenesis and lymphangiogenesis. *J Clin Invest* **127**, 457-471

24. Noy, P. J., Lodhia, P., Khan, K., Zhuang, X., Ward, D. G., Verissimo, A. R., Bacon, A., and Bicknell, R. (2015) Blocking CLEC14A-MMRN2 binding inhibits sprouting angiogenesis and tumour growth. *Oncogene* **34**, 5821-5831

25. Lugano, R., Vemuri, K., Yu, D., Bergqvist, M., Smits, A., Essand, M., Johansson, S., Dejana, E., and Dimberg, A. (2018) CD93 promotes beta1 integrin activation and fibronectin fibrillogenesis during tumor angiogenesis. *J Clin Invest* **128**, 3280-3297

26. Galvagni, F., Nardi, F., Spiga, O., Trezza, A., Tarticchio, G., Pellicani, R., Andreuzzi, E., Caldi, E., Toti, P., Tosi, G. M., Santucci, A., Iozzo, R. V., Mongiat, M., and Orlandini, M. (2017) Dissecting the CD93-Multimerin 2 interaction involved in cell adhesion and migration of the activated endothelium. *Matrix Biol* **64**, 112-127

27. Tosi, G. M., Caldi, E., Parolini, B., Toti, P., Neri, G., Nardi, F., Traversi, C., Cevenini, G., Marigliani, D., Nuti, E., Bacci, T., Galvagni, F., and Orlandini, M. (2017) CD93 as a Potential Target in Neovascular Age-Related Macular Degeneration. *J Cell Physiol* **232**, 1767-1773

28. Nanda, A., Karim, B., Peng, Z., Liu, G., Qiu, W., Gan, C., Vogelstein, B., St Croix, B., Kinzler, K. W., and Huso, D. L. (2006) Tumor endothelial marker 1 (Tem1) functions in the growth and progression of abdominal tumors. *Proc Natl Acad Sci U S A* **103**, 3351-3356

29. Simonavicius, N., Ashenden, M., van Weverwijk, A., Lax, S., Huso, D. L., Buckley, C. D., Huijbers, I. J., Yarwood, H., and Isacke, C. M. (2012) Pericytes promote selective vessel regression to regulate vascular patterning. *Blood* **120**, 1516-1527

30. Kim, T. K., Park, C. S., Jang, J., Kim, M. R., Na, H. J., Lee, K., Kim, H. J., Heo, K., Yoo, B. C., Kim, Y. M., Lee, J. W., Kim, S. J., Kim, E. S., Kim, D. Y., Cha, K., Lee, T. G., and Lee, S. (2018) Inhibition of VEGF-dependent angiogenesis and tumor angiogenesis by an optimized antibody targeting CLEC14a. *Mol Oncol* **12**, 356-372

31. Ki, M. K., Jeoung, M. H., Choi, J. R., Rho, S. S., Kwon, Y. G., Shim, H., Chung, J., Hong, H. J., Song, B. D., and Lee, S. (2013) Human antibodies targeting the C-type lectin-like domain of the tumor endothelial cell marker clec14a regulate angiogenic properties in vitro. *Oncogene* **32**, 5449-5457

32. Palaniappan, K. K., and Bertozzi, C. R. (2016) Chemical Glycoproteomics. *Chem Rev* **116**, 14277-14306

33. Wibowo, A., Peters, E. C., and Hsieh-Wilson, L. C. (2014) Photoactivatable glycopolymers for the proteome-wide identification of fucose-alpha(1-2)-galactose binding proteins. *J Am Chem Soc* **136**, 9528-9531