Discovery of Small Molecule Inhibitors of Hyaluronan Binding at Cell Receptor CD44

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

LI-KAI LIU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISOR: Barry C. Finzel, Ph.D.

June 2015

Acknowledgements

Dedication

I dedicated my dissertation to my loving parents and sister, who have always encouraged my love of learning, and to my fiancée Ka-yan, who have inspired me and supported me throughout this journey.

Abstract

Selective inhibitors of hyaluronan (HA) binding to the cell surface receptor CD44 will have value as probes of CD44-mediated signaling, and have potential as therapeutic agents in chronic inflammation, cardiovascular disease and cancer. Using biophysical binding assays, fragment screening, and crystallographic characterization of complexes with the CD44 HA binding domain, we have discovered an inducible pocket adjacent to the HA binding groove into which small molecules may bind. Fragment combination and iterations of structure-driven design has led to identification of a series of 1,2,3,4-tetrahydroisoquinolines as the first non-glycosidic inhibitors of the CD44-HA interaction. The affinity of these molecules for the CD44 HA binding domain parallels their ability to interfere with CD44 binding to polymeric HA *in vitro*. X-ray crystallographic complexes with lead compounds are described and compared to a new complex with a short HA tetrasaccharide, in order to establish the tetrahydroisoquinoline pharmacophore as an attractive starting point for lead optimization.

Table of Contents

List of Tables, Schemes and Appendices

List of Figures

List of Abbreviations

HA	Hyaluronan
p-HA	Polymeric hyaluronan
o-HA	Hyaluronan oligosaccharide
ECM	Extracellular matrix
HABD	Hyaluronan binding domain
GlcNAc	N-acetylglucosamine
GlcUA	Glucuronic acid
THIQ	1,2,3,4-tetrahydroisoquinoline
BZD	2,3,4,5-tetrahydro-1H-benzodiazepine
TSG-6	Tumor necrosis factor-alpha stimulated gene-6
CD44	Cluster of differentiation 44
LYVE-1	Lymphatic Vessel Endothelial Hyaluronan Receptor-1
RHAMM	Receptor for hyaluronan-mediated motility
SPR	Surface plasmon resonance
DSF	Differential scanning fluorimetry

CHAPTER 1

INTRODUCTION

1.1 Hyaluronan as a signaling molecule

The glycosaminoglycan hyaluronan (HA), is a high molecular weight co-polymer consisting of repeats of *N*-acetylglucosamine (GlcNAc) and glucuronic acid (GlcUA) linked by a glucuronidic bond (Figure 1-1). HA is an integral component of the extracellular matrix (ECM) within all tissues, providing structural and functional integrity to cells and tissues [1]. Cells must constantly orchestrate a balance between HA production and turnover in order to shape the molecular environment in which they live [2, 3]. Many organs maintain high concentrations of HA that are needed for hydration, space filling and to provide a structural framework through which cells can migrate or blood vessels grow [4].

HA can be synthesized and released by many cell types, and mesenchymal cells such as stem cells are believed to be the major source of HA. Importantly, HA remodeling changes cell behavior [4]. HA is synthesized by membrane-bound synthases on the inner surface of the plasma membrane, and the chains are extruded into pericellular space or dumped into the matrix through a pore-like structure [5]. Mammalian HA is synthesized by three HA synthases (HAS) [3] and is actively produced during wound healing and tissue repair [6]. Native high molecular weight or polymeric HA (p-HA) of 1,000 to 10,000 kDa, is the predominant isoform under physiological conditions. It promotes epithelial structural integrity and is anti-inflammatory [1, 7]. It also serves as a common thread to bring adhesion receptors into close proximity to amplify biological effects [8].

Catabolism of HA impacts multiple biological processes such as tissue repair, immune cell activation and migration [9], as well as the progression of many vascular diseases [10] and cancer [4]. About one-third of total body HA is replaced daily, with a metabolic half-life of 1 to 1.5 days [11]. Native p-HA is rapidly degraded in tissues to fragments of 10 to 100 kDa that are used, recycled and supplemented with newly synthesized HA. HA degradation is enhanced under certain pathological conditions such as arthritis or cancer that causes an imbalance of HA turnover [6]. Smaller fragments of HA serve as both signaling molecules in the matrix, and as a product of tissue injury; cells must be able to recognize changes in HA composition in order to shape the environment in which they live [3, 12]. Vascular injury or pathogen invasion can raise the proportion of HA fragments that are interpreted as danger signals by cells of the innate immune system [13]. HA fragmentation has been linked to cell differentiation [14], angiogenesis [6, 15], tumorigenesis [16, 17], and in resistance to cancer chemotherapy [2, 18].
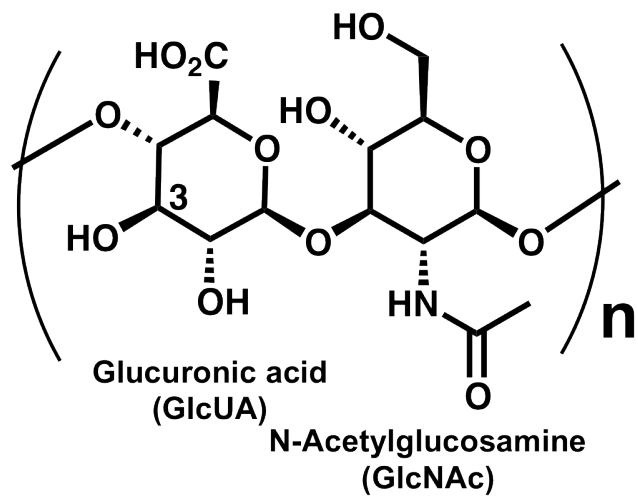
Figure 1-1. Chemical structure of hyaluronan (HA) that comprises glucuronic acid (GlcUA) and N-acetylglucosamine (GlcNAc) disaccharide repeats.

1.2 The Hyaladherins: Hyaluronan-binding Proteins

The roles of HA are mediated through diverse hyaladherins, or hyaluronan-binding proteins (HABP) that exhibit significant differences in their tissue expression, cellular localization and regulation, as well as affinity and specificity to HA [4]. Some HABPs such as CD44 or LYVE-1 are associated with cell membranes, while others including TSG-6 and other adhesion receptors are found in either the ECM or cytoplasm. RHAMM was discovered as the first cell-associated non-Link-module hyaladherin in the cytoplasm and nucleus, or transiently expressed on the cell surface of leukocytes and fibroblasts, regulating HA-mediated cell migration and proliferation in normal and tumor cells [19]. Other cell surface receptors such as Toll-like Receptors 2 and 4 are not hyaladherins, but preferentially recognize very small HA-derived oligosaccharides resulting from pathogen infection as danger signals that initiate innate immune responses [3, 20, 21]. HA of different sizes are recognized by different receptors to trigger diverse biological responses [22]. The molecular basis for understanding HA recognition by these proteins is a key to understanding the biology of HA and the possibilities for therapeutic intervention.

Structurally, many HABPs contain a lectin-like domain of about 100 residues called the 'Link module' that may exist as a single or tandemly repeated HA binding region. The Link protein was the first identified HABP isolated from cartilage [23]. It is comprised of two contiguous Link modules and an immunoglobulin domain subsequently found in the G1 domain of aggrecan, versican, neurocan and brevican [19]. These are tissue-specific proteoglycans present within extracellular space that reshapes HA polymers to provide

cartilage lubrication, vessel elasticity and tissue integrity in the skin or brain. Neurocan and brevican also contribute to neuronal development and brain tumor growth [3].

Other Link module proteins include the membrane-associated receptor LYVE-1 [24]. It is produced in lymph vessel endothelium and is implicated in HA degradation and cell trafficking within lymphatic vessels and lymph nodes. TSG-6 is the smallest HABP that contains a single Link module secreted into ECM in response to inflammatory stimuli [25]. TSG-6 can also attach covalently to HA using a Link-independent mechanism to drive the formation of periodic filamentous complexes that are likely to promote leukocyte adhesion [26]. Some hyaladherins lacking the Link module (e.g. RHAMM) contain the BX_7B motif believed to be the site for interacting with HA and mediating cell migration and proliferation in many normal or cancer cells [27].

CD44 is the most characterized HABP, both structurally and functionally. Sequence alignment suggests that CD44 is a close homolog of LYVE-1 and TSG-6 that contains the Link module (Figure 1-2). CD44 is a membrane-associated glycoprotein widely expressed in different cell types in a variety of different isoforms (Figure 1-3). The isoform of expression and the cell type dictate the HA binding properties of CD44. Most stromal cells (e.g. smooth muscle cells, fibroblasts), immune cells (e.g. microphages, leukocytes) and epithelial cells need the standard isoform of CD44 to mediate cell migration, adhesion and proliferation [19, 28]. Constitutively active variant isoforms of CD44 are abundantly expressed in many cancer cells and act as co-receptors to promote cell survival and chemo-resistance [2, 8]. All variant isoforms of CD44 include a consensus functional ectodomain at the *N*-terminus [28] – the hyaluronan-binding domain (HABD).

6

The HABD (17 kDa) contains a canonical link module stabilized by extended sequences. It is glycosylated and is known to be sufficient for the binding of hyaluronan [29].

```
CD44_HUMAN  |1UUH        20----AQIDLNITCRFAGVFH-VEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALSIGF
CD44_MURINE|2JCP         23---MNQIDLNVTCRYAGVFH-VEKNGRYSISRTEAADLCQAFNSTLPTMDQMKLALSKGF
LYVE1_HUMAN             24SLRAEELSIQVSCRIMGIT-LVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALKASF
TSG-6_HUMAN|2PF5         1---------------GVYHREARSGKYKLTYAEAKAVCEFEGGHLATYKQLEAARKIGF
                                        *:      :... .:. :**    *.      *    *:: * . .*
CD44_HUMAN  |1UUH        ETCRYGFI-EGHVVIPRIHPNSICAANNTGVYIL-T--SNTSQYDTYCFNASAPPEEDCT
CD44_MURINE|2JCP         ETCRYGFI-EGNVVIPRIHPNAICAANHTGVYILVT--SNTSHYDTYCFNASAPPEEDCT
LYVE1_HUMAN             ETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPV--SRQFAAYCYNSSDTWTNSCI
TSG-6_HUMAN|2PF5         HVCAAGWMAKGRVGYPIVKPGPNCGFGKTGIIDYGIRLNRSERWDAYCYNPHAK------
                         ..*  *:: .* *    : *   *.   .*:       : :: :**:*
CD44_HUMAN  |1UUH        SVTDLPNAFDGPITITIVN-------RDGTRYVQKGEYRTNPEDIYPSNPTDDDV
CD44_MURINE|2JCP         SVTDLPNSFDGPVTITIVN-------RDGTRYSKKGEYRTHQEDIDAS-------
LYVE1_HUMAN             PEIITT---KDPIFNTQTATQTTEFIVSDSTYSVASPYSTIPAPTTTP-P-----
TSG-6_HUMAN|2PF5         ------------------------------------------------------
```

Figure 1-2. Sequence alignment of the Link domain from human and murine CD44, LYVE-1 and TSG-6. The extended sequence outside the canonical Link module is highlighted in yellow. Identical residues are highlighted in red. Sequence similarities across the HABPs are noted in symbols below the sequences.

HABD
HA binding, N- and O-Glycosylation

Membrane proximal region
Splice variants, Glycosylation sites for association with heparan or chondrointin sulfate

Cytoplasmic domain
Phosphorylation site, cytoskeleton interaction site
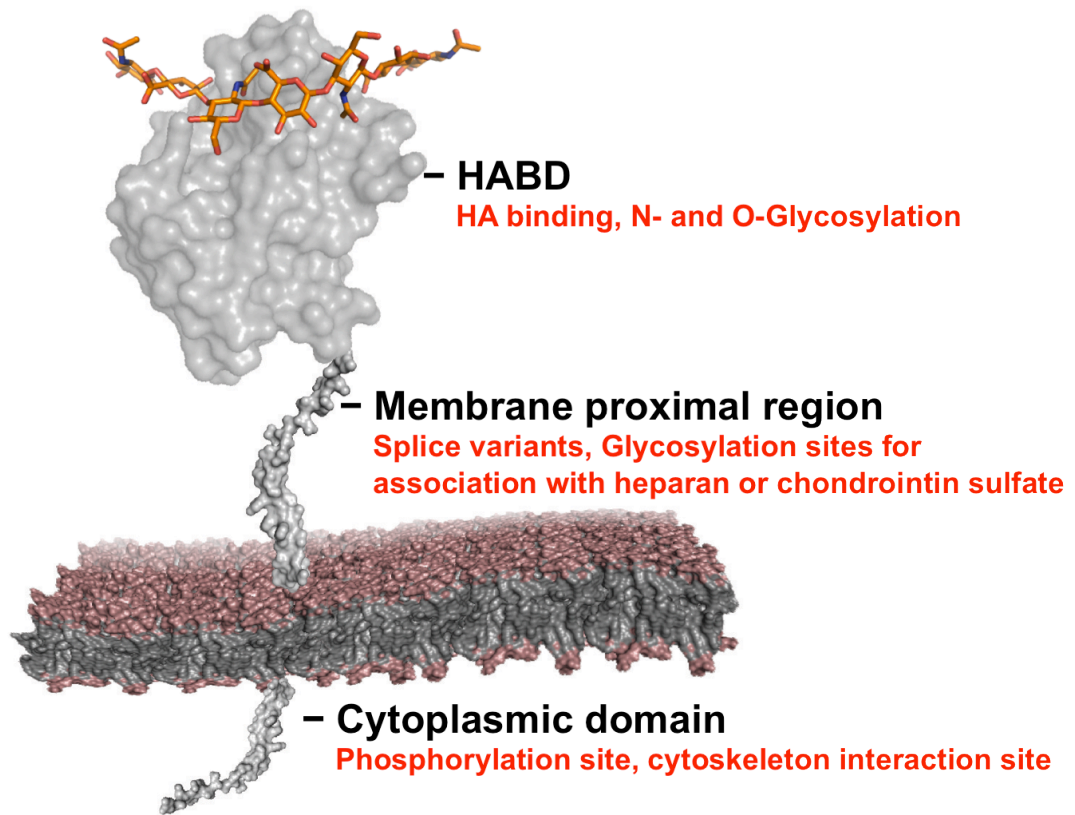
Figure 1-3. Domain organization of the full-length CD44.

Site-specific de-glycosylation of HABD on Asn25 and Asn120 activates the HA binding activity of "standard" CD44 [30]. Alternative isoforms arising from alternative splicing of variable exons differ in the size of potential glycosylation sites in the membrane-proximal region, and the form of glycosaminoglycans to which they attach (e.g. heparin or chondroitin sulfate) [4]. Glycosylation can facilitate assembly of protein complexes with receptor kinases and growth factors to initiate signal transduction [2]. Through cytoskeletal attachment, the cytoplasmic domain also mediates the signaling complex formation and the receptor density on the cell surface [19].

Importantly, a great deal is known about the structure of the HABD and the interaction it makes with oligosaccharides of HA from structural biology [31, 32]. A recombinantly expressed and refolded HABD is able to bind immobilized high-molecular weight of HA with an affinity similar to that of the glycosylated ectodomain of CD44 purified from eukaryotic cells [29]. Crystallographic research has provided a description of the structure of the human HABD [31], and a complex of the murine HABD with a bound HA octasaccharide (HA_8) [32]. The complex identifies an extended HA binding site across a shallow groove of the HABD.

1.3 Antagonism of the CD44-HA interactions

CD44 is a major cell-surface receptor of both native polymeric HA and its fragments. It is widely accepted that CD44-HA interactions are important to both immune stimulation

and cell adhesion, but the response to HA binding is dictated by polysaccharide chain-length [21]. Degraded HA can trigger signaling cascades by competing with native HA for CD44 [33-35]. The binding to endogenous polymeric HA (p-HA; > 1000 kDa) promotes epithelial structural integrity and is generally considered to be anti-inflammatory [1,7]. The anti-inflammatory action of native HA may facilitate co-localization of multiple CD44 or with other receptors that reinforce signal strength [2, 24, 36]. CD44 binding by HA fragments (< 1000 kDa) can trigger signaling cascades that result in stimulation of the innate immune response [6], the recruitment of immune cells [7], tumor growth [12, 16, 17, 35, 37], HA synthesis [38] and the disruption of cell-surface receptor clustering [36] that supports tissue homeostasis. HA oligosaccharides (o-HA) as short as HA_4 or HA_6 can also influence a variety of biological processes [22], presumably because they can disrupt binding to endogenous p-HA [33-35].

Different hyaladherins have preferences for HA of different lengths. Different receptors may recognize secondary structures of HA oligomers imparted by different chain lengths [19, 22]. The study of the molecular details of CD44-mediated signaling has been hampered, however, by the existence of multiple mechanisms by which cells respond to changing HA populations [21, 24, 39]. CD44 apparently partners with RHAMM, integrins and other receptors to regulate activation of pathways that stimulate growth, motility and, in the case of malignant tumors, promote therapeutic resistance [2, 18]. In some cases, RHAMM can compensate for the functions of CD44 through distinct mechanisms [2]. A number of RHAMM isoforms are present in the cytoplasm or cell surface and associate with HA, or translocate to the nucleus to regulate transcription. It is widely accepted that CD44- and

11

RHAMM-HA interactions are crucial in both tumor malignancy and resistance to therapy. Major challenges for future research in the field include elucidating the mechanism of activation of CD44 signaling in response to HA and the relative importance of the CD44 variant forms in different contexts. There are limits to what can be learned about CD44-specific signaling mechanisms by studying HA oligosaccharide, because o-HA has such varied biological roles. More selective inhibitors of HA binding by CD44 might be used to provide convincing evidence that it is indeed the competition for native HA by HA oligos that accounts for their functional role. A selective small molecule inhibitor of HA binding to CD44 would provide a means to study signaling and cell localization effects in HA mediated responses, without the interfering consequences of off-target effects involving other HA receptors.

It has been proposed that selective antagonists of HA binding by CD44 may also be useful in the treatment of inflammatory diseases [40-42], and certain types of cancer [4]. Monoclonal antibodies to CD44 that block binding to HA have been used to investigate the role of CD44 in immune cell recruitment [43], where they reduce tissue swelling and inflammation by inhibiting immune cell infiltration [44]. A more recent study has also shown that anti-CD44 antibodies reduce HA-induced vascular smooth muscle cell migration and the production of inflammatory cytokines and reactive oxygen species in vascular injured mice [42]. However, a clinical trial initiated to study the utility of CD44 antibodies in cancer chemotherapy was halted due to toxicity [45, 46]. Limited work toward the development of small molecule inhibitors of HA binding has been reported. A natural product isolated from fungi reportedly possesses potent activity as an HA antagonist (24

12

μM) [47], but the molecule – a fusion of a long fatty acid and a tri-saccharide – has not been synthetically prepared and does not have a structure that would be appealing as a drug candidate.

1.4 Structure of CD44 HA binding domain

Crystallographic and solution structural studies have revealed much about the structure of the CD44 hyaluronan-binding domain (HABD). A crystal structure of the human HABD (residues 20-178; hHABD$_{20-178}$) confirms that it includes "Link module" – a conserved α/β fold consisting of two triple-stranded antiparallel β-sheets and two α-helices, stabilized by a pair of disulfide bonds [31]. Importantly, this structure reveals unique stabilizing elements that contain additional flanking sequences to provide an additional disulfide bridge and four β strands forming an enlarged HA binding domain from the Link module.

The *murine and human* HABD structures share the same protein fold. Murine HABD shares 86% sequence identity with the human protein, with 90% identity for the Link module (Figure 1-4 A). The murine structure can be superimposed upon the human atomic coordinates with a root mean square deviation of 0.45 Å for 114 paired Cα atoms. Figure 1-4 B includes a comparison to murine mHABD$_{24-174}$ structure (2JCP) [32]. Backbone differences are limited to portions of loops joining β5 to β6 (particularly residues 108-112, human numbering), β6 to β7, and β7 to β8. These are the same regions

that show conformational diversity in different human structures suggesting that these regions have intrinsic flexibility even when the primary sequence is unchanged. The difference between murine and human CD44 conformation in the vicinity of 108-112 has been noted and attributed to the insertion of Val112 (murine numbering) in the murine protein (as shown by Banerji *et al.*, 2007).

The actual HA binding interface was later confirmed crystallographically with murine HABD (residues 24-174; *m*HABD) [32]. The *m*HABD complex with a bound HA octasaccharide identifies an extended HA binding site across a shallow groove joined by the β2-α1, α2-β3 and β4-β5 loops. The binding groove comprises two basic residues (Arg45 and Arg82) and several hydrophobic side chains (Ile92, Ile100-Ala103, Tyr46, Tyr83 and Tyr109) at the front face of the Link module that make substantial hydrogen bond or van der Waal's contacts with the oligosaccharide. This was unexpected because a mutagenesis study had previously identified mostly polar and basic residues as functionally important [48]. Analysis of the binding interface in the two complexes revealed that the central recognition site is the GlcUA5–GlcNAc6 region (GlcUA at position 5 through GlcNAc at position 6). The conformational change of Arg45 and Arg82 creates a deeper cleft being involved in HA recognition. Hydrophobic interactions from Ile100 and a disulfide of Cys81 and Cys101 also pack against GlcUA5 and GlcNAc6. Other aliphatic residues, together with the tyrosine hydroxyls, make specific contributions to the binding of GlcUA7 and GlcNAc8. Molecular dynamics suggests GlcUA5–GlcNAc8 are more constrained to their crystallographic positions during simulation [49]. This observation is supported by a much smaller temperature factor for GlcUA5–GlcNAc8 in

the crystal structure [32]. Earlier studies have found that residues (Arg150, Arg154, Lys158 and Arg162) in the extended region contribute to HA binding [48, 50], but beside Arg150, these residues do not come close to the Link module in the crystal structure.

A

| | | β1 | | β2 | | α1 | | α2 |

```
                         β1                β2            α1              α2
CD44_HUMAN │1UUH   20----AQIDLNITCRFAGVFHVEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALSIGF
CD44_MURINE│2JCP   23---MNQIDLNVTCRYAGVFHVEKNGRYSISRTEAADLCQAFNSTLPTMDQMKLALSKGF
                        *****  ***  ********************* ********  **   *** **

                         β3   β4            β5        β6
CD44_HUMAN │1UUH   ETCRYGFIEGHVVIPRIHPNSICAANNTGVYIL-TSNTSQYDTYCFNASAPPEEDCT
CD44_MURINE│2JCP   ETCRYGFIEGNVVIPRIHPNAICAANHTGVYILVTSNTSHYDTYCFNASAPPEEDCT
                   ********** *********** ***** ******* ***** * ****************

                         β7            β8      α3
CD44_HUMAN │1UUH   SVTDLPNAFDGPITITIVNRDGTRYVQKGEYRTNPEDIYPSNPTDDDV
CD44_MURINE│2JCP   SVTDLPNSFDGPVTITIVNRDGTRYSKKGEYRTHQEDIDAS-------
                   ******* **** ************ ****** *** *
```

B

Figure 1-4. (A) Sequence alignment of the HABD from human and murine CD44. The Link module was highlighted in yellow. (B) Comparison of the crystal structures of the human (green, 1UUH) and murine HABD (cyan, 2JCP) showing secondary assignments and annotations.

16

Previous structural studies have confirmed that only one difference occurs between *h*HABD and *m*HABD near the HA binding site [31, 32]: the substitution of Asn101 in *h*HABD for histidine in *m*HABD. This conservative difference exists at the extreme end of the binding groove; in the *m*HABD complex with HA_8, His105 contacts the GlcNAc residue in subsite 8.

The affinity of CD44 for o-HA_8 is further enhanced by the conformational change around the $\beta2$-$\alpha1$ loop and the C-terminal lobe [32, 51]. The $\beta2$-$\alpha1$ loop centered on Arg45 adjusts in concerted movement with Glu41, Asn43, Gly44 and Arg82 to create more intimate contacts with the polysaccharide in the HA_8 complex. A shift in Arg45 has been proposed as the principle difference between the open and the closed state of CD44 induced upon HA binding, denoted "A" and "B" state, respectively [32]. The solution structure of *h*HABD in complex with o-HA suggest that HABD alters its conformation from the ordered "O" to partially disordered "PD" upon HA binding [51-53]. The PD state appears to reveal the burial of the CD44-HA recognition site and display a higher HA affinity than the O state.

From the structural data, one might conclude that CD44 is not an easily druggable target. The murine HABD-HA complex reveals an extended HA binding site with surface area exceeding 800 $Å^2$ and molecular binding stabilized by a large number of weak interactions involving at least seven consecutive saccharide units of HA. The HABD has no well-formed or deep pockets that would serve as attractive binding sites for small molecule inhibitors, and is known to undergo small but important conformational changes upon binding HA [32]. In many respects, the protein-polysaccharide complex resembles protein-protein interactions that are difficult to disrupt effectively with small molecules, but with all the structural information and molecular biology of CD44 HABD

in place, it constitutes an attractive target for the development of small-molecule

inhibitors of HA binding.

1.5 Hypotheses and summary of research

With our research, we seek to identify or create small molecule antagonists of HA-

binding to CD44 that can be used to investigate the role of the CD44-HA interactions in

signaling and chronic inflammation. It has been proposed that antagonists of these events

may be useful in the treatment of atherosclerosis [42] and several types of cancer [28]. Given

the strong precedent for CD44 HABD structural biology, we have chosen a fragment-

based approach for ligand discovery [54] that has proven applicable to the search for

inhibitors of protein-protein interactions [55, 56]. Such complexes are often driven by a large

number of weak interactions enhanced by small conformational adaptations at the

protein-protein interface. This is also a defining characteristic of the complex between the

CD44 HABD and hyaluronan. Fragment screening and crystallography permits a direct

experimental observation and characterization of conformational adjustments needed to

accommodate specific bound small molecules. Knowledge of the structure of bound

fragments can be exploited to join individual molecular features using a well-

characterized spatial arrangement that best compliments the altered geometry of a larger

protein-binding site, leading to improved ligand design. It is our <u>central hypothesis</u> that

potent and selective small molecule antagonists of HA binding to CD44 can be developed

using fragment-based screening and structure-aided design.

Because CD44 catalyzes no substrate conversion that can be monitored to detect

binding events, we have developed sensitive binding assays exploiting Surface Plasmon

Resonance (SPR) that can be used to detect and quantify the binding of small molecules

with even low affinity. This assay has been used to screen a library of ~1000 small

molecules to identify compounds that bind the CD44 HABD. To better understand the

binding of small HA-derived oligosaccharides (o-HA), we have also determined the

structure of the murine CD44 HABD in complex with a HA tetrasaccharide (HA_4). The

site of binding is compared to the binding of small-molecule HA antagonists identified

through fragment screening. The structural data provides an explanation for the

antagonist activity of these compounds, and confirms the value of these assays as tools

for the discovery of very potent and antagonists specific for the HA binding site of CD44.

A combination of fragment features has led the identification of a simple

pharmacophore capable of inducing a binding site just below the HA binding groove on

CD44 HABD. Through a series of iterative steps of analog preparation, binding

characterization and structure-determination, more potent small molecule have been

created that promote better HA antagonism *in vitro*. We present an initial characterization

of these molecules and their effect on cells, and establish an expanded hypothesis

regarding the design of even more potent HA antagonists for consideration in future

work. We propose to optimize the chemical composition of these small molecules to

increase binding affinity, and to expand the binding footprint so that they abut the HA

subsite.

These results have led to our conclusion that potent and CD44-selective small-molecule antagonists of the HA interaction can be generated by optimizing these molecules or by synthetically linking chemical entities that can occupy this fragment-binding site to short polysaccharide units that occupy the HA binding site. Our long-term goal is to create a potent and selective antagonist of HA binding on CD44. We will then evaluate the potential effect of these compounds in cell-based assays that are sensitive measures of CD44-dependent processes. Ultimately, our approach to the research – using fragment screening and crystallography to discover and optimize small molecule inhibitors – is significant. This is an approach that is broadly applicable to HA binding proteins or a number of different drug discovery efforts.

CHAPTER 2

BIOPHYSICAL METHODS TO ASSESS COMPETITIVE

BINDING TO CD44

2.1 Background

To assess the ability of small oligosaccharides (o-HA) to disrupt CD44-HA binding, others have conducted fluorescence-based competition analyses in whole cells [34]. The ability of o-HA preparations of different lengths to alter cell avidity for fluorescein-labeled HA has been studied by flow cytometry, but variable amount of CD44 on cells makes it difficult to use those cells to quantify competition. A whole cell approach is difficult for use in our application because of the lack of sensitivity to detect weakly binding fragments which typically have milli- to micromolar binding affinities. Some others have used recombinant hHABD coated in a microtiter plate to quantify the binding of biotin-labeled HA in competition with o-HAs, but the method requires much of the proteins (40 μg/well) and detection reagents (e.g., streptavidin-linked horseradish peroxidase and phenylenediamine) [31].

Surface Plasmon Resonance (SPR) provides a more direct quantification of molecular affinity for biomolecules. SPR can detect the amount of soluble proteins associated with a surface coated with substrates or in a reverse format by measuring mass-dependent changes in reflective index. The assay has been employed by others to assess affinity of various CD44 HABD mutants for a p-HA coated SPR surface [32, 51, 53]. A recombinantly expressed and refolded HABD is able to bind immobilized p-HA with an affinity similar to that of the glycosylated CD44 purified from eukaryotic cells [29]. More importantly, the assay consumes less protein and the immobilized p-HA can be reused for multiple experiments.

We will adapt the assay for use as a competition assay. We will prepare recombinant CD44 HABDs, biotinylated p-HA that can crosslink to the SPR surface, and prepare the p-HA coated surface using the SPR instrument. The amount of HABD that contacts with the surface can be monitored on the SPR in real-time and determine competition activity of o-HA or small molecules if appropriate. We expect to be able to detect a change in the affinity of the HABD for this surface in the presence of small molecules such as o-HAs that effectively inhibit HA binding by CD44.

Also, we designed another SPR assay using the same HABDs coated on the SPR surface to determine binding affinity (K_D) of these saccharides that in turn affirmed the integrity of the HABDs. This assay format is commonly used to detect low-affinity binding of small molecules [57, 58]. This method has the sensitivity to detect and quantify very weak molecular binding in a range of millimolar (e.g. fragment molecules). In addition, we had established Differential Scanning Fluorimetry (DSF) as a quick and economical assay to evaluate protein stability and integrity. We have used DSF to optimize protein storage conditions and to qualitatively detect ligand interactions, which can increase protein stability observable by an unfolding temperature (T_m) shift [59]. The utility of these assays has been validated in fragment screening described in Chapter 3, and we have used these binding data to promote inhibitor development.

To implement this plan, there is a critical need for a ready supply of recombinant CD44 hyaluronan binding domains (HABDs) for both biophysical assays that can be used to detect and quantify binding of HA and small molecules to proteins. The same HABD constructs can also be used for crystallographic structural studies. Human and murine

23

HABD have been cloned into an appropriate vector, over-expressed in *E. coli*, refolded and purified in our laboratory. CD44 refolding is a key process for producing soluble HABDs and has been described in detail elsewhere [29, 31, 32].

Motivated by a desire to obtain a crystal form of human CD44 HABD suitable for use in binding studies with small molecule inhibitors, we have undertaken a search for alternate crystal forms. We sought to improve upon the resolution of diffraction, and to obtain alternate crystal forms with a small asymmetric unit and an accessible HA binding groove for use in soaking experiments with HA oligosaccharides and other ligands. To obtain the previously known hHABD orthorhombic crystal form (PDB-ID 1UUH), Jackson and co-workers expressed and purified hHABD$_{20\text{-}178}$ [31]. With a slightly different construct, hHABD$_{18\text{-}178}$, we have identified two alternate crystal forms under different crystallographic conditions, one hexagonal and one monoclinic, that diffract to high resolution. We will describe two alternate crystal forms of hHABD that eventually proved unsuitable for the structural studies of inhibitor design. Instead, what seems valuable from these crystal forms was the identification of an unexpected peptide binding site. A crystal form obtained from murine HABD has been reproduced that diffracts to high resolution has proved suitable for use as a surrogate in ligand binding studies.

2.2 Materials and Methods

2.2.1 Materials

Human and murine CD44 cDNA were generously provided by Dr. James B.
McCarthy at the University of Minnesota (Minneapolis, MN). Rosetta 2(DE3)pLysS
competent cells were purchased from EMD Biosciences (Gibbstown, NJ). HiPrep 16/60
Sephacryl S-100 HR was obtained from GE Healthcare (Piscataway, NJ). $_L$-arginine,
reduced/oxidized glutathione, hyaluronan sodium salt from rooster comb, coupling
reagents and other general buffer reagents were from Sigma (St. Louis, MO). Hydrazide
biotin linker for making biotinylated HA was from Thermo Scientific (Rockford, IL).
Short oligosaccharide assortments of HA_4, HA_6, HA_8, HA_{10} and HA_{12} with low
endotoxin (CSR-93002) were purchased from Cosmobio (Japan). All general buffer,
crystallization reagents were from Sigma (St. Louis, MO) or major chemical companies.
Monodisperse HA oligosacchride tetramer (HA_4) was purchased from Hyalose
(Oklahoma City, OK).

2.2.2 Expression of the recombinant human and murine CD44 HA-binding domains

Human and murine CD44 HABDs were cloned, expressed and purified to homogeneity using procedures similar to those previously described [29, 32]. For bacterial expression of HABDs, the ectodomain constructs of human $CD44_{20-178}$ (hCD44) and murine $CD44_{25-174}$ (mCD44) were prepared from full-length cDNA clones using appropriate primers. A segment from the human cDNA encoding residues 20 to 178 was amplified using the forward primer, 5'-gtgccgctgcatatgG$_{56}$CGCAGATCGATTTGAATATAACCTGCCG-3' and the reverse primer, 5'- ggttctcgagctaC$_{530}$ACGTCATCATCAGTAGGGTTGCTG-3'. For bacterial expression of mHABD, the ectodomain construct $CD44_{25-174}$ with Met23 and Asn24 substituted in place of His23 and Glu24 was also prepared from a full length cDNA clone using the forward primer, 5'-ggcacatcatatgACCCAGATCGATTTGAATGTAACCTGCCGC-3' and the reverse primer, 5'-taatctcgagctaATCGATGTCTTCTTGGTGTGTTCTATACTCGC. These segments were amplified using standard forms of human and murine CD44 cDNA as templates prior to being cloned into the pMCSG7 expression vector digested with NdeI and XhoI. The resulting plasmids were transformed into the $E.$ $coli$ strain Rosetta 2(DE3)pLysS, and grown in a 10-ml LB media overnight for bacterial expression. The cells were used to inoculate a 1-L LB media up to an optical density of 0.6 before

induction by isopropyl-β-D-thiogalactosidase (IPTG) of 0.4 M. At 4 h of postinoculation, the cell pellet was obtained by centrifugation, resuspended in a 30-ml MES buffer and taken through two cycles of freeze and thaw before a 2-m sonication until the lysate was homogeneous.

2.2.3 Purification procedures for the CD44 hyaluronan-binding domain

Soluble human and murine HABDs were refolded and purified equivalently from insoluble inclusion bodies much as previously described [29]. The inclusion bodies were extracted from the cell lysate by centrifugation. They were resuspended in a Triton X-100 buffer and pressed into a Dounce Homogenizer to remove soluble cell debris. Before refolding, the washed inclusion bodies are isolated by centrifugation and the HABD was extracted with 8 M denaturant urea. The supernatant (typically 10 ml from 1 L culture) comprises mainly recombinant HABD as assessed by SDS– PAGE analysis. Denatured HABDs were diluted in a 2 L buffer containing 250 mM $_L$-arginine, 2 mM reduced and 1 mM oxidized glutathione for refolding. This was done by delivering the solution in a rate of 0.02 ml min^{-1} via a ECONO gradient pump, resulting in final protein concentration as 50 to 75 μM. The solution was slowly stirred at 4 °C for 24 hours followed by ultrafiltration through an Amicon membrane. Monomeric HABDs were separated from the oligomeric forms by a single pass of HiPrep 16/60 Sephacryl S-100 eluted with 20

mM Tris-HCl, 150 mM NaCl at a flow rate of 0.4 ml/min. Multiple fractions (3 ml/each) were collected based on the elution profile and analyzed on SDS/PAGE to trace monomeric proteins. In general, fractions containing monomeric proteins were mostly eluted between 90 and 120 minutes, then pooled and concentrated in a Centriprep concentrator (Amicon) as aliquots at 1 mg/ml and stored at −80°C. This results in yields of 30 mg and 10 mg each from 1 L culture for the human and murine construct, respectively.

2.2.4 Mass spectroscopy

ESI-MS provides a sensitive means to determine the oxidation states for medium-size proteins like HABDs. The HABD (1 mg/ml) dissolved in the protein buffer containing 20 mM Tris-HCl and 150 mM NaCl was desalted by passage through a C4 ziptip column and eluted with a solution of 5% acetonitrile 0.1% trifluoroacetic acid in water. Fractions of 100 μl were injected into LTQ Orbitrap (Thermo Scientific). The protonated forms of HABD (m/z from 10+ to 17+) were selected for mass deconvolution, using the Bayesian Protein Reconstruction Parameter: S/N = 5 and % of intensity = 0.4%. No observable sodium adducts of HABD were found. The mass deconvolution has sensitivity up to 0.2 Da. To prepare HABD under a reducing condition, the sample was added 1 mM of a reducing agent TCEP and incubated for 15 min before analysis. The same procedure was applied to a control protein lysozyme purified from chicken egg white (L4919, Sigma).

28

2.2.5 Thermal stability of the protein

Differential Scanning Fluorimetry (DSF) was performed using a bench top real-time PCR (BioRad C1000 Thermocycler). Melting curves are generated by slowly warming hHABD (at 0.05 mg/ml) through a thermal cycle (5 to 95 °C) in 1°C increments for 30 seconds in the buffer containing 20 mM Tris-HCl, 150 mM NaCl, and SYPRO Orange (5X) dye, while monitoring increased fluorescence of the dye as it binds to denatured protein. The control sample using egg white lysozyme was done similarly at 1 mg/ml in 100 mM sodium acetate. To explore buffer conditions that promote protein stability and crystallization, we performed stability screening with various buffer, pH, solvent compositions prepared in house. 2X of the stability screen solutions were mixed 1 : 1 with a 2X master mix that contains hHABD (0.1 mg/ml) in the original buffer for a reaction of 40 μl. The melting transition for the protein was observed using SYPRO Orange in 5X in 1°C increments for 30 seconds from 5 to 95 °C. 96 conditions of the stability screen solutions can be screened in parallel in less than 60 minutes.

2.2.6 Immobilization of Hyaluronan for Surface Plasmon Resonance

Competitive binding analyses were performed using SPR (Biacore, GE Healthcare) and sensor chips onto which p-HA is immobilized (Figure 2-1) [32, 51, 53]. Biotinylated HA was synthesized from hyaluronan sodium salt and hydrazide biotin through amide coupling as described [25]. Briefly, 5 mg/ml of HA in 100 mM MES buffer (pH 5.5) was first activated

with 13 μl of EDC at 25 mg/ml in the same buffer, then 20 μl of biotin-LC-hydrazide at

92.88 mg/ml (in DMSO) was added, and the reaction mixture was agitated on a shaker

for 18 hr at room temperature. The sample was dialyzed vigorously against dH_2O (2 L) at

4 °C and changed water four times before removing insoluble materials by centrifugation.

A surface coated with immobilized p-HA was prepared by simply passing the

biotinylated HA in HEPES-EP (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA,

0.005% (v/v) surfactant P-20) over a streptavidin-coated SA chip (Biacore, GE

Healthcare) at flow rate of 20 μl/min as progress was monitored in a Biacore 2000.

Surface saturation was achieved when the reflectivity stabilized at 115 RU.
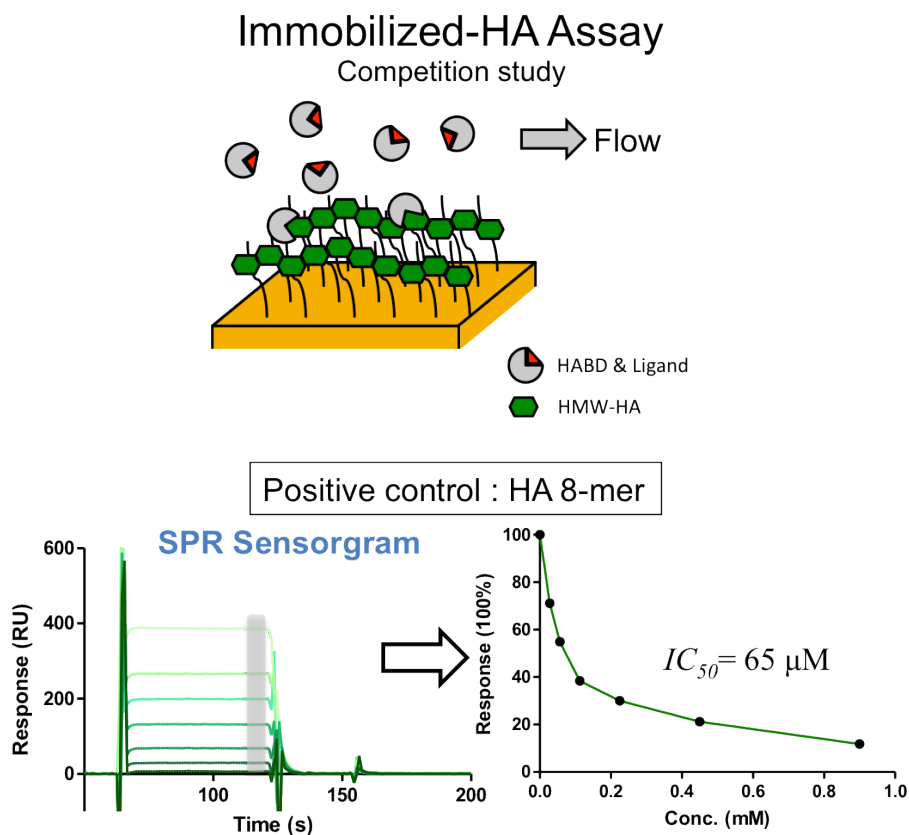
Figure 2-1. Biophysical assays utilizing SPR have been implemented in two formats. In one format polymeric HA (or high-molecular weight HA) is biotinylated and immobilized to measure the interference of HABD binding by various concentrations of small molecules. This assay is suitable for use to screen large fragments libraries efficiently. Control dose-response sensorgrams (raw data) obtained with HA_8 are plotted on the bottom left. HA_8 can disrupt *h*HABD binding to the immobilized HA. The grey area indicates the averaged response extracted from 5 seconds prior to the end of the injection to determine steady state binding. The experimental data were calibrated and then fit to a binding isotherm as a function of ligand concentrations to derive an inhibition concentration IC_{50} (inset, See Section 2.2.6).

2.2.7 Competitive Interference with HABD binding to Immobilized HA

The affinity of hHABD to the HA-coated surface was quantified in the presence of

dimethyl sulfoxide (DMSO) as solvent vehicle for use as the baseline in competition

binding studies. An equilibrium binding constant (K_D) for surface-immobilized HA was

obtained from dose response curves. Two-fold serial dilutions of eight protein solutions

(0, 1.4 – 181 μM) in HEPES-EP containing 5% DMSO were injected over the surface at

20 μl/min for 1 minute supplemented with the same buffer as a running buffer. To

quantify ligand interference with this binding, competition binding experiments were

conducted similarly using a fixed concentration of hHABD (typically, 17 μM) pre-

equilibrated with various concentrations of ligands in HEPES-EP containing 5% DMSO

at 25 °C. Ligand activity can be expressed as an IC_{50}, the concentration resulting in a

50% decrease in hHABD binding to HA on the chip surface. Each ligand series includes

a blank solution containing the protein only that provides a reference full response of

hHABD binding to the immobilized HA. A "blank" reference surface was coated with

only streptavidin. All of the samples, along with blanks from buffer and solvent

calibration solutions [60], were centrifuged and measured on a 96-well microplate (Biacore,

GE Healthcare) at 25 °C.

Reflectivity response data points were extracted from response curves in the 5 seconds

prior to the end of the injection to determine steady state binding (Figure 2-1). All the

data were solvent calibrated and double referenced with blanks using standard procedures

with BIAevaluation software v4.0.1 [60]. The equilibrium binding constant K_D of hHABD

was measured by fitting the calibrated data a curve with nonlinear regression of one-site saturation binding. Calibrated experimental curves of reflectivity response were fit to a four-parameter dose-response curve using Graphpad Prism 5.04 to calculate an IC_{50}.

2.2.8 Immobilization of hyaluronan-binding domain for Surface Plasmon Resonance

Recombinantly expressed human or murine HABD was immobilized on CM5 sensor chips (Biacore, GE Healthcare) via amine coupling for use in HABD binding assays (Figure 2-2). Chips were prepared and evaluated on a Biacore 3000 instrument. The dextran surface was activated with a 1 : 1 mixture of 0.4 M 1-ethyl-3-(3-dimethylamino-propyl) carbodiimide hydrochloride and 0.1 M N-hydroxy succinimide in the instrument. Human HABD (150 μg/ml) in 10 mM sodium acetate at pH 4.4 was flowed past a working surface (flow cell 2) for 7 minutes before blocking the remaining activated carboxymethyl groups with 1 M ethanolamine at pH 8.5 to achieve a level of 3500 to 4000 RU suitable for binding analysis.

2.2.9 Small-molecule HABD affinity by SPR

The immobilized HABD surface is used to quantify the binding of small molecules to the human and murine HABD and determine dissociation constants (K_D). A reference surface (flow cell 1) was activated and reacted with only ethanolamine. The binding

sensorgram can be observed in real-time when the small molecule flows pass both surfaces, but the signal from flow cell 2 can be subtracted from that of flow cell 1 to generate a referenced signal specific to the HABD binding. The referenced response extracted from the portion of the curve representing the 5 seconds prior to the end of the injection (binding plateau) was obtained from a concentration-dependent variations in response and used to plot a binding curve to determine steady state binding. The experimental data were calibrated and then fit to a 1 : 1 binding isotherm as a function of ligand concentrations (inset, Figure 2-2).

Immobilized-HABD Assay

Screen & Dose-Response
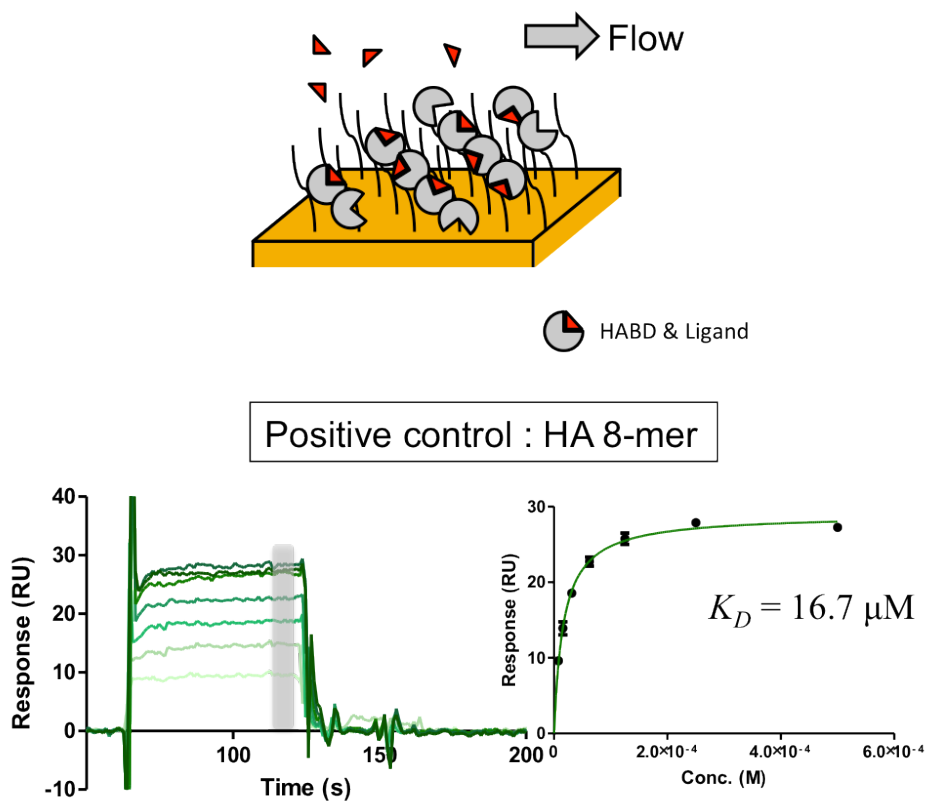
Figure 2-2. In the other SPR assay, recombinant HABD is immobilized and the affinity of small molecules for the HABD can be measured. This assay is suitable for use to screen large fragments libraries efficiently. Control dose-response sensorgrams (raw data) obtained with HA_8 are plotted on the bottom left. HA_8 can bind to the immobilized hHABD. The grey area indicates the averaged response extracted from 5 seconds prior to the end of the injection to determine steady state binding. The experimental data were calibrated and then fit to a binding isotherm as a function of ligand concentrations to derive an affinity K_D (inset, See Section 2.2.8).

2.2.10 Human CD44 hyaluronan-binding domain production and crystallization

The variant construct hHABD$_{18\text{-}178}$ was expressed in *E. coli*, extracted, refolded, and purified from insoluble inclusion bodies following procedures described for preparation of hCD44$_{20\text{-}178}$ [29] and as adapted previously (Chapter 2, part I). This construct was amplified using the forward primer, 5'-CCTCGTGcatatg G$_{50}$CTATGGCGCAGATCGATTTGAATATAACCTGCCG-3' and the reverse primer, 5'-AGGActcgagCTAC$_{530}$ACGTCATCATCAGTAGGGTTGCTG-3'. An error in primer design resulted in expression of construct CD44$_{18\text{-}178}$ with two unintended residues (Ala18-Met19) at the N-terminus, but this proved fortuitous. Amplified segments of the complementary DNA for each construct were prepared from human CD44 cDNA generously provided by James B. McCarthy at the University of Minnesota (Minneapolis, MN) using appropriate primers incorporating recognition sites (*Nde*I and *Xho*I as shown in lower case, respectively) for ligation into expression vector pMCSG7 [61]. Vector sequences were confirmed by DNA sequencing and the resulting plasmids were transformed into *E. coli* strain BL21(DE3)pLysS for bacterial expression. This resulted in yields of more than 20 mg recombinant HABDs from 1 L of culture. The sequence and formation of three disulfides in the construct hHABD$_{18\text{-}178}$ were confirmed by mass spectroscopy and crystallography. The molecular mass of the refolded protein was 17783.1 Da, consistent with the expected sizes for fully reduced proteins 17788.8 Da.

Hexagonal and monoclinic crystals of hCD44$_{18\text{-}178}$ were obtained by hanging-drop vapor diffusion using the same protein in a buffer containing 20 mM Tris-HCl and 150

mM NaCl concentrated to 8 mg/ml. To obtain hexagonal crystals, this protein was mixed in a 1 : 1 ratio with well solution containing 30% MME 5000, 200 mM Na_2SO_4 and 100 mM MES at pH 6.5 and suspended over the well solution in 2 μl drops. Crystals with symmetry consistent with space group $P6_122$ grew to $30 \times 30 \times 100$ μm^3 over a period of two weeks at 4ºC with or without 5% DMSO. Crystal growth can be accelerated to one week by introducing 0.3 μl of the well solution containing micro-crystals from previous experiments into the crystallization drops. Monoclinic crystals with $C2$ space group symmetry grew in several weeks at 4ºC over a well solution comprising 25% PEG 3350, 100 mM HEPES at pH 7.5, 100 mM NaCl and 5% DMSO. Monoclinic crystals are less numerous, grow more slowly, and get larger (up to $100 \times 100 \times 150$ μm^3).

2.2.11 Crystallization of the murine CD44 hyaluronan-binding domain

Murine HABD *apo* crystals have been reproduced using procedures previously described [32] for co-crystallization experiments. Briefly, protein was concentrated to 8 mg/ml prior to mixing in a 1 : 1 ratio with a well solution containing 30% PEG MME 5000, 200 mM Na_2SO_4 and 100 mM MES buffer pH 6.5 to obtain a monoclinic form in space group $P2_1$ in a few days. Separate plate-shaped crystals form after two to three rounds of microseeding with crushed polymorphic plates that form initially.

2.2.12 Preparation of co-crystals of mHABD with a hyaluronan fragment

HA tetrasaccharide complex co-crystals were prepared by soaking *m*HABD *apo* crystals in a droplet of well solution containing HA oligosaccharide 4-mer (HA$_4$) in final concentration 32.5 mM for 30 minutes. Crystals were flash frozen in liquid nitrogen after quick transfer into the well solution with and additional 10% glycerol and stored until required for data collection.

2.2.13 Data collection and structural determination

Diffraction data were collected from single crystals at IMCA-CAT station 17-ID-B, GMCA-CAT 23ID-D at Advanced Photon Source, Argonne, IL or 4.2.2 at Advanced Light Source, Berkeley, CA. Data were processed using XDS [62] and scaled with SCALA [63]. Summary scaling statistics are provided in Table 2-2. (Reflection statistics in parentheses are for highest resolution shell). Structures were determined by molecular replacement with the program PHASER [64], using a search model monomer A of PDB entry 1UUH [31] as for the human structures and 2JCP [32] for the HA$_4$-murine structure. Both human crystal forms have two protein molecules in the asymmetric unit, whereas the murine co-complex remains a single chain as the original form. Iterative rounds of model building and restrained refinement were carried out with COOT [65] and REFMAC5 [66]. Discreet alternate conformations were modeled only when it appeared that two states with roughly equal occupancy could best explain the electron density. Occupancies of

atoms in alternate conformations were constrained to sum to 1.0 and refined as a group using REFMAC5. Tetrasaccharide coordinates and stereochemical restraints were generated in JLigand [67]. Refined structures were validated with MolProbity [68]. Atomic coordinates and reflection data have been deposited in the Protein Data Bank [69] (accession codes 4PZ3, 4PZ4 and 4MRD).

2.3 Results

2.3.1 Optimization of the recombinant protein purification

  Slight modification to the refolding procedure has led to an increase of protein yield for both the human and murine HABD constructs. This was achieved by slowly delivering denatured HABD into the refolding buffer while keeping the final protein concentration between 50 and 75 μM. Since protein aggregates built up with an increase of time, the best result was observed when gel filtration chromatography proceeds immediately after ultrafiltration. The samples isolated from gel filtration were analyzed closely on a 12.5% polyacrylamide gel under reducing and nonreducing conditions and stained with Coomassie blue (Figure 2-3). The native fold of this monomeric protein was further confirmed as discussed below.

2.3.2 Assessing the structural integrity of the human CD44 hyaluronan-binding domain

The mass predicted by the sequence of both mHABD$_{24\text{-}174}$ and hHABD$_{20\text{-}178}$ constructs is confirmed by MS. To confirm that disulfide bonds were correctly formed, electrospray ionization mass spectroscopy (ESI-MS) was used to analyze oxidation states of a protein. To evaluate the methodology, we analyzed molecular mass of oxidized or reduced lysozyme, which is about similar size of HABDs (14.3 kDa) and contains one disulfide bond. Indeed, disulfide linkage of an ionizable protein like lysozyme can be discerned based on mass changes by ESI-MS. The oxidized condition gives an expected mass for the refolded state and the instrument provides the resolution for m/z peaks up to 0.1 Da. The molecular mass of the refolded human and murine HABD determined by ESI-MS was 17580.9 Da and 16947.6 Da, respectively, consistent with the expected masses of 17580.5 Da and 16946.9 Da (Figure 2-4). This in turn validates sequence identity of the constructs and provides information about sample purity. The mass suggested a mixture of fully-reduced and folded states when pre-incubated with reducing agents, which was further supported by the subsequent thermal shift assay (discussed below). Crystallography has further confirmed the sequence identity and the presence of disulfide bonds at expected locations.
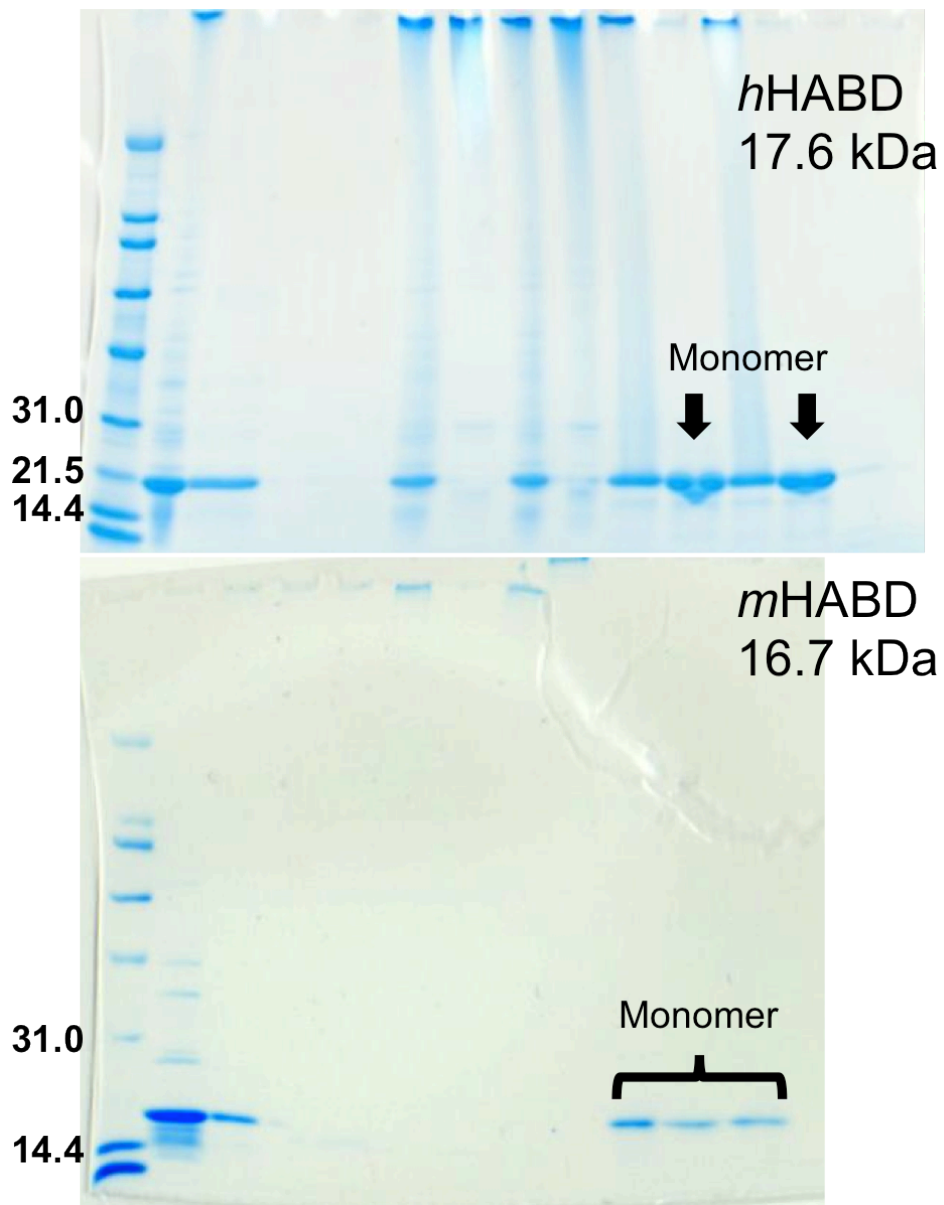
Figure 2-3. SDS-PAGE gel analysis to the eluents of HABDs collected from the size-exclusion chromatography. Fractions that contain monomeric proteins show a band that matches the expected mass. (A) *h*HABD. (B) *m*HABD.
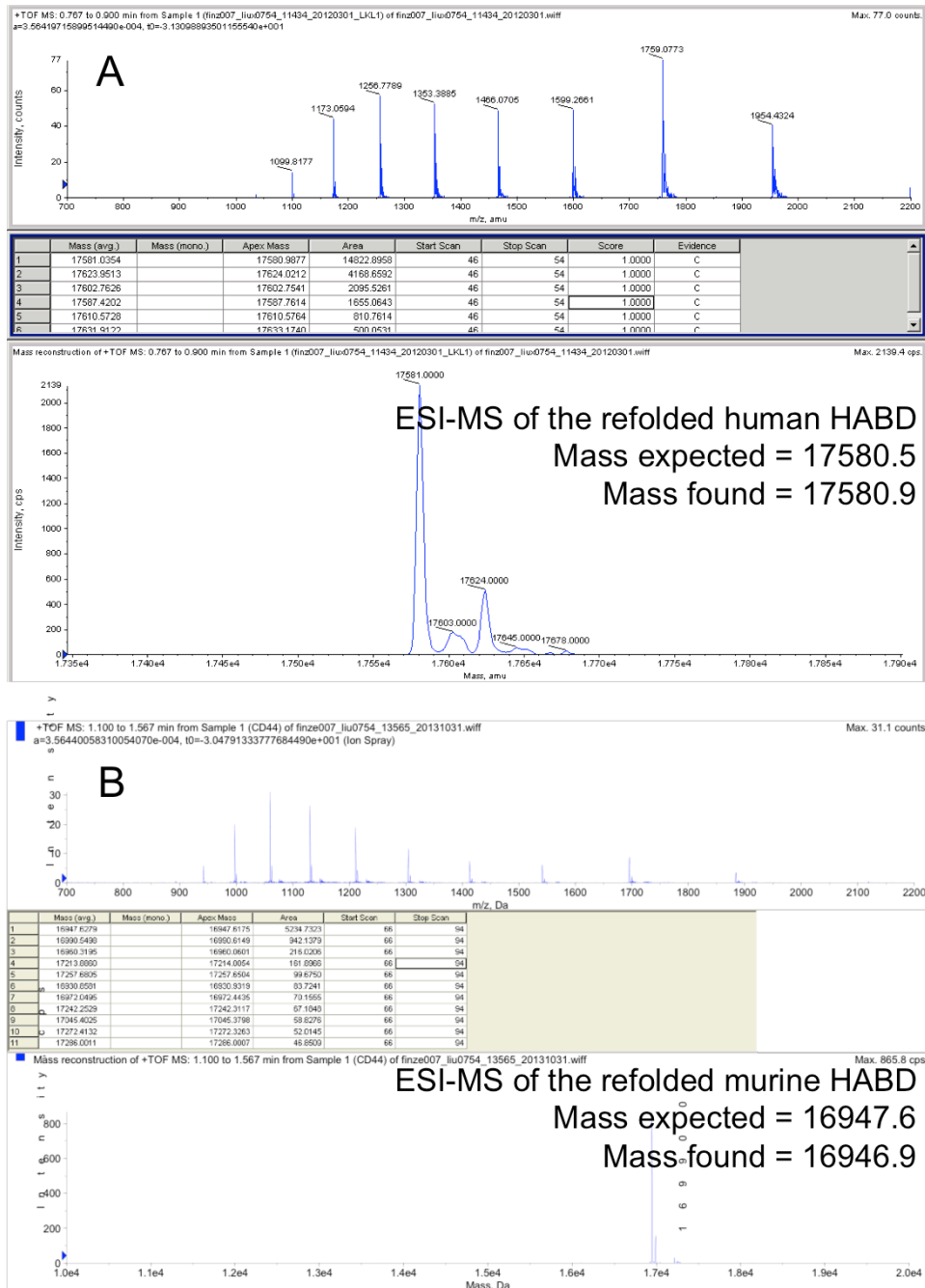
Figure 2-4. Mass spectrometry analysis to the purified monomeric HABDs (A) *h*HABD. (B) *m*HABD. Calculated and found masses are noted.

2.3.3 Thermal stability of the human CD44 hyaluronan-binding domain

DSF enables a simple and inexpensive way to assess protein stability, and to search for thermal stabilizing ligands for structural studies and drug discovery [59, 70]. An unfolding temperature (T_m) is measured by an increase in the signal of a fluorescent dye whose spectral properties change upon binding to an unfolded protein. The T_m is not so important, but relative changes correlate with the change in thermal stability.

Protein stability is often correlated with buffer composition, so we sought to identify optimum solution conditions for protein stabilization using a buffer screen [59]. A change of the conditions could shift the equilibrium between the folded and unfolded states, and thereby increase or decrease the melting temperature. Using the thermal shift assay, a buffer screen was performed and revealed that the protein is comparably stable in different buffer systems and pH values, but sensitive to reducing reagents and salts at high concentrations. The T_m of hHABD is 45 °C in the original buffer (Figure 2-5). The T_m of hHABD remained near 45 °C when the protein was in a range of concentrations of the Tris-HCl buffer (pH 8) used in the literature. Therefore, we continued to use it for protein storage [29]. The addition of 10% glycerol allows a restoration of T_m in high salt (500 mM). This result supports the advantage of using glycerol as a component for storage as well as a precipitant as used in the literature.
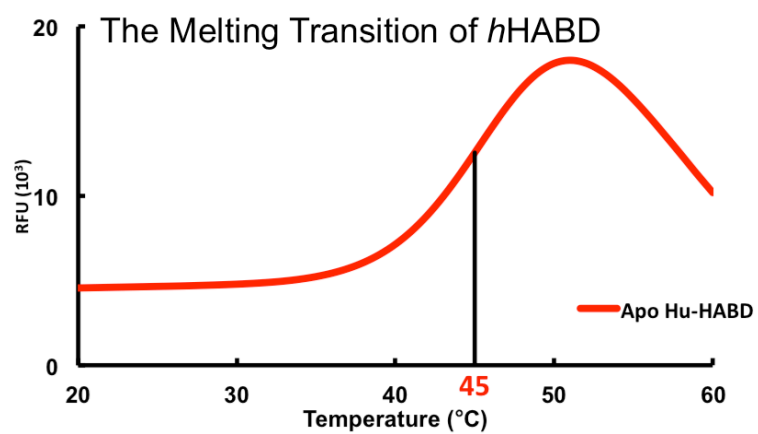
Figure 2-5. The melting curve of *h*HABD assessed by differential scanning fluorimetry. The T_m of *h*HABD is 45 °C in the original buffer of 20 mM Tris-HCl and 150 mM NaCl.

To understand the relationship between thermal stability and disulfide bond formation, we compared the thermal shift of hHABD with lysozyme. Under a reducing condition of 1 mM TCEP for 15 minutes, a disulfide-linked lysozyme showed two transition points of a melting curve, at 50 and 72 °C, respectively. Untreated lysozyme has a T_m of 72 °C in the original buffer, suggesting a mixture of oxidation states under a reducing condition. hHABD is much more sensitive to reducing agents. High fluorescence intensity appeared in the protein denaturation curve before the sample was heated up, suggesting a large portion of hHABD was fully reduced in the presence of TCEP.

2.3.4 Competitive Interference HA binding assay

For this assay, we prepared biotinylated high-molecular-weight HA that was subsequently immobilized to a commercially available streptavidin-coated chip on the Biacore machine (Figure 2-6). This surface can be reused in multiple experiments to monitor the binding of soluble HABD. Approximately 5% of carboxylate (1 out of 20 HA disaccharide) was modified with the biotin linker [71]. The length of the spacer arm that links to biotin to the HA apparently influences how well the HA conjugate can be attached to the streptavidin-coated chip. We tried and failed to couple the commercially available hyaluronate biotin salt (> 700 kDa, Sigma Aldrich) or a conjugate of hydrazide-biotin we have made with a shorter spacer arm. Ultimately, we made and coupled a conjugate with a longer linker. Hydrazide-LC-biotin is 9 Å longer than the counterpart

hydrazide-biotin and provides length needed for streptavidin interact with the conjugate

through biotin.
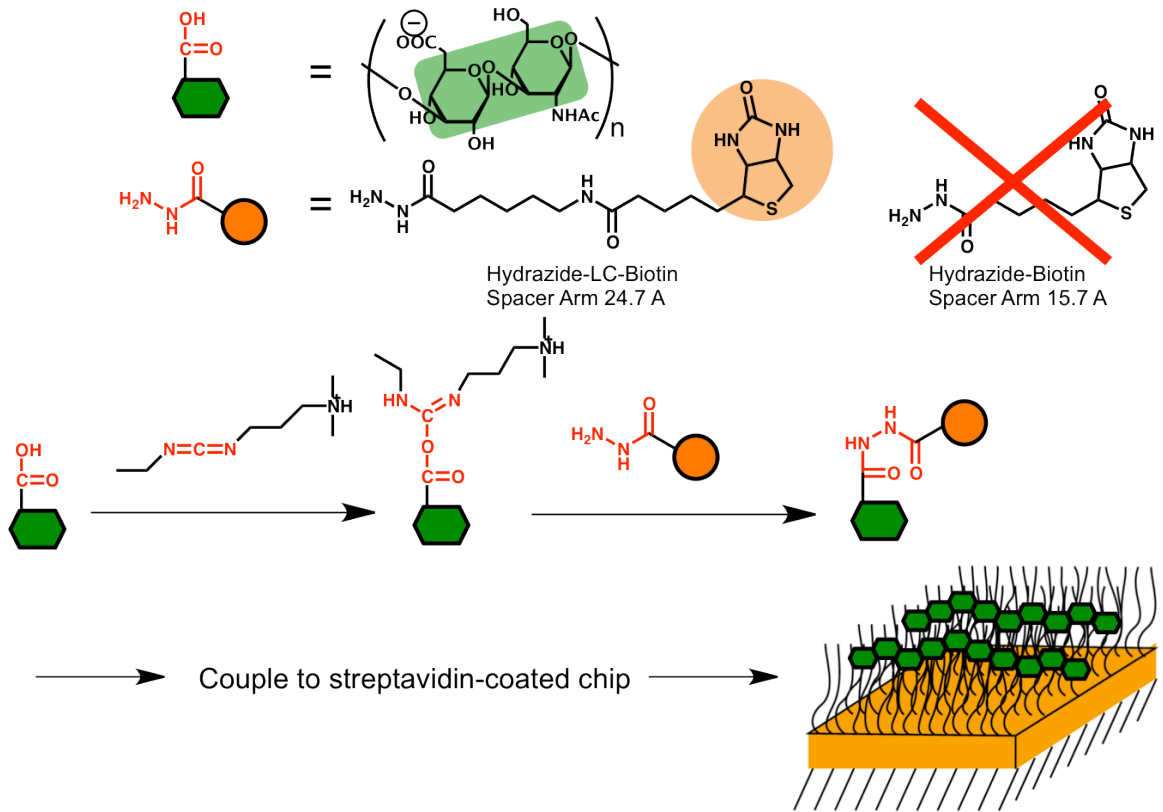
Figure 2-6. The linkage chemistry of the biotinylated HA. Hydrazide-LC-biotin was used instead of a shorter spacer of hydrazide-biotin.

2.3.5 The soluble human hyaluronan-binding domain binds to the immobilized hyaluronan

To affirm that our recombinant hHABD binds HA, we measured the K_D using the

Immobilized HA-SPR assay (section 2.2.7). With this assay, we have measured human

$CD44_{20-178}$ (hHABD) binding with a dissociation constant (K_D) of 36.6 ± 4.0 µM (Figure

2-7 A). This K_D is higher than the ~20 µM K_D reported previously [32, 51], but may differ

due to differences in the density of HA immobilized on a given chip surface, or because

DMSO is included in the vehicle buffer.

2.3.6 Hyaluronan fragments interfere with HABD binding to immobilized hyaluronan

Inhibitors of CD44 binding can be recognized by their ability to cause a dose-

dependent reduction in binding response observed from a fixed concentration of hHABD

(Figure 2-1). Since inhibitor activity is expressed as the concentration of 50% inhibition

(IC_{50}) relative to a full response of hHABD induced concurrently with vehicle (5%

DMSO) on the same chip, the apparent binding affinity to any particular prepared surface

may vary without consequence. Since the association between the CD44 HABD and HA

immobilized on the surface supposedly occurs as a consequence of HA binding across the

HA binding groove, small molecule inhibitors of that interaction should induce a

concentration-dependent reduction in HABD retention on the surface. To validate this

hypothesis, we quantified the surface-binding interference of short HA oligosaccharides (o-HA) known to impact CD44-mediated cell adhesion [34], presumably by direct competition with matrix HA binding. Competitive inhibition may be expressed as the concentration of 50% inhibition (IC_{50}) relative to a full binding response of hHABD induced concurrently with vehicle (5% DMSO) on the same chip. Resulting IC_{50} values for oligosaccharides varying in length from 4 to 12 saccharide units are summarized in Table 2-1. (Full dose response curves underlying these values are shown in Figure 2-7 B). Smaller HA fragments inhibit hHABD adsorption less effectively (higher IC_{50}), but there is a clear trend indicating that the limit in potency is nearly reached with HA_{12}. Even HA_4 gives rise to quantifiable inhibition from a reasonable dose response curve in this assay, although the IC_{50} is only in the millimolar range (1.3 mM, Table 2-1). This data demonstrates that this assay can be used to quantify the p-HA binding interference of even very weak inhibitors.
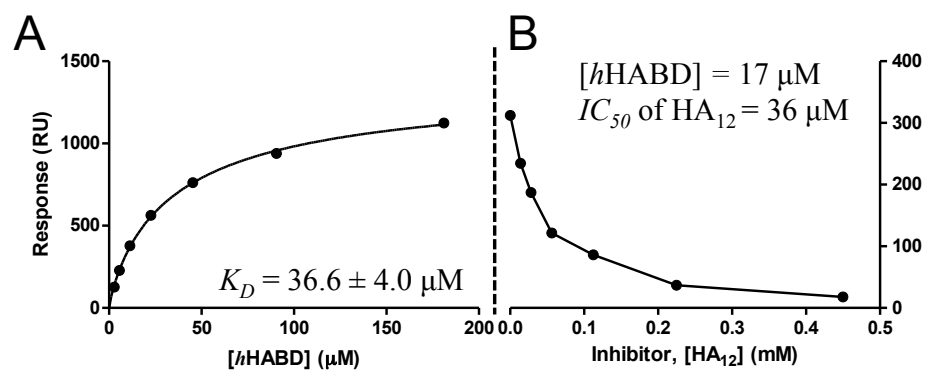
Figure 2-7. A) The affinity of the CD44 *h*HABD for the immobilized-HA surface was quantitated as 36.6 µM in HEPES-EP in the presence of 5% DMSO. A dose-dependent inhibition of this binding response is observed by small molecules that are inhibitors of p-HA binding, such as soluble oligosaccharide fragment HA_{12} **(B)**.

2.3.7 Immobilization of human and murine HABD for SPR

The immobilized-HABD assay (section 2.2.9) was implemented to enable direct

measurement of equilibrium dissociation constants (K_D) for ligand binding to CD44

hyaluronan binding domains (Figure 2-2). In order to monitor binding by SPR,

recombinant *h*HABD or *m*HABD were immobilized on dextrin-coated gold film chips by

amine coupling. To facilitate chemical coupling of protein constructs with pI near 5,

(*h*HABD pI = 4.7; *m*HABD pI = 5.5), an acidic buffer is desirable to make the protein

more positively charged and facilitate contact with the negatively charged dextrin

surface. Sodium acetate buffer with pH 4.4 or lower is required.

Coupling should not interfere with HA binding. Each HABD has multiple lysine

residues at the protein surface that can form covalent amide bonds via activated

carboxylate on the dextrin film, but none of these are particularly near the known

hyaluronan binding site.

In general, 30 to 40% of the dextrin carboxylate is coupled with HABD under this

reaction condition. The remaining activated carboxylates can be blocked by adding

ethanolamine. Typically, the coupling step consumes ~20 μg of HABD to reach a

saturating immobilization level that produces around 4000 response units (RU) in

reflectivity.

In order to estimate the effective concentration of HABD at the surface, and to assess

how much of that is accessible for ligand binding, we examined the binding of HA_8 as a

positive control. Only 30% of the immobilized protein is freely accessible to the ligand

binding, as HA_8 usually generates a response of 40 RU in saturation. The ratio is

calculated using the comparison of the amount of HA_8 that binds to the surface and a theoretical value that fully occupy all the HABD (~120 RU). The theoretical value can be calculated from the immobilization level multiplying the ratio of the molecular weight of HA_8 versus HABD [60]. This level corresponds to the surface concentration 2 mM. The loss of protein activity can be closely monitored over time using the same control ligand. The surface can be reused over multiple experiments before the need for creating a new surface. In general, the surface should be regenerated when the affinity of HA_8 is deviated from 16 µM by 3 folds.

2.3.8 Hyaluronan fragments bind to the human and murine HA-binding domain

The binding affinity of oligosaccharides (HA_4 to HA_{12}) to murine and human CD44 hyaluronan binding domains was assessed using the immobilized-HABD assay (Table 2-1). The result was correlated with those obtained by other methods [32]. Dissociation constants (K_D) calculated upon analysis of dose response curves (shown in Figure 2-8) are given in Table 2-1. Oligosaccharide fragments exhibit progressively higher binding affinity with increasing oligosaccharide size. Our measurements reflect affinities that are 3 to 4-fold higher than previously determined by isothermal titration calorimetry for HA_8 and HA_{10} [32], but which preserve the same relative ranking. The sharply lower affinity of HA_4 for either human or murine HABD (K_D is 245 µM and 428 µM, respectively) compared to larger oligomers is consistent with monovalent competition analyses conducted in whole cells to evaluate oligosaccharide influence on cell adhesion [34]. The

sensitivity of this assay for small molecules with low binding affinity validates this assay format for use in finding other molecules that bind to the HABD *in vitro*.

Table 2-1. HA-derived Oligosaccharide CD44 HABD Binding and p-HA-binding Interference

	K_D (μM)[1]		IC_{50} (mM)[2]
HA Oligosaccharide	*h*HABD	*m*HABD	*h*HABD
4	245 ± 227	428 ± 332	1.32
6	38.8 ± 4.9	58.4 ± 8.7	0.108
8	16.7 ± 1.2	32.7 ± 3.6	0.065
10	11.5 ± 0.9	21.7 ± 2.6	0.048
12	3.4 ± 0.5	6.4 ± 1.3	0.036

[1]K_D values were obtained by SPR, measuring dose-dependent binding to immobilized human and murine HABD. Values are shown as the mean ± standard error from triplicate measurements.

[2]IC_{50} values were obtained by SPR, measuring ability of increasing compound concentrations to effect the binding of flowing *h*HABD to immobilized polymeric HA (115 RU). All experiments were conducted using flowing *h*HABD concentration fixed at 17 μM. Values are shown as the mean from duplicate measurements.

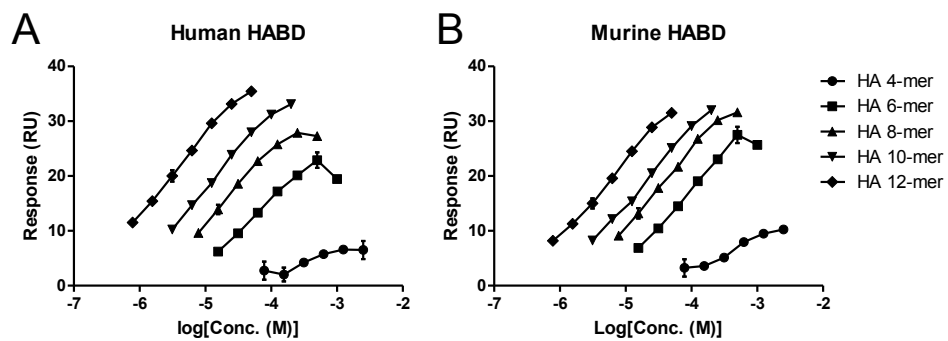**Figure 2-8**. Binding of HA-derived oligosaccharides of different lengths to immobilized *h*HABD (**A**) and *m*HABD (**B**). Values are plotted as the mean ± standard error from triplicate measurements as given in Table 2-1.

2.3.9 Human CD44 hyaluronan-binding domain crystal forms

To begin a search for alternate crystal forms, three different constructs of human

HABD were expressed, purified and evaluated in crystallization trials – $hHABD_{20-178}$,

$hHABD_{18-178}$ and $hHABD_{20-168}$ (Table 2-3). Broad screening with commercially available

incomplete factorial crystallization screens did not result in the discovery of new

conditions. Instead, targeted screening guided by the known conditions for successful

crystallization of murine CD44 constructs [32] resulted in the identification of two new

crystal forms. Hexagonal crystals of $hHABD_{18-178}$ arise from conditions using MME 5000

as precipitant similar to those used for crystallization of *apo* murine HABD. Monoclinic

crystals result from the same construct under conditions similar to those used in the

crystallization of the murine CD44-hyaluranon oligosaccharide complex. Both crystal

forms were obtained with a construct ($hHABD_{18-178}$) that extends constructs used in

earlier work by two residues at the N-terminal end, and substitutes the naturally occurring

Ser18-Leu19 of CD44 with Ala18-Met19. A shorter construct ($hHABD_{20-168}$) was

engineered to resemble the molecular extent of the $mHABD_{24-174}$ construct used

extensively in our laboratory, while construct $hHABD_{20-178}$ replicates the exact construct

successfully crystallized by Jackson et al. [31]. No usable crystals of either $hHABD_{18-178}$ or

$hHABD_{20-168}$ were obtained. Throughout this discussion, protein constructs are identified

by species and residue range (e.g., $hHABD_{20-178}$, or $mHABD_{24-174}$, while crystal structures

are identified unambiguously by the four-letter PDB-ID (e.g., 1UUH, 4PZ3 or 4PZ4).

Individual protein monomers within a crystal structure are identified by PDB-ID and

chain identifier (e.g., 1UUH-A, or 4PZ3-B).

2.3.10 Comparison of the CD44 hyaluronan-binding domain crystal structures

The four crystallographically independent snapshots of the hHABD$_{18-178}$ revealed in these crystal structures (4PZ3-A, 4PZ3-B, 4PZ4-A and 4PZ4-B) can be pooled with the other two conformations available from the non-isomorphous orthorhombic hHABD$_{20-178}$ crystals previously reported by others [31] (1UUH-A and 1UUH-B), to conduct an analysis of conformational flexibility in these domains. To look for and quantitate differences in the six available structures, we overlaid all on a common positional framework (1UUH-A) (Figure 2-9 B), computed a mean position for each atom from all six structures, and then calculated root mean square deviations (RMSD) from this mean position by residue. Results are presented graphically in Figure 2-10. The mean RMSD for all atoms was 0.53 Å. More discussion of differences and similarities in these structures is provided below.
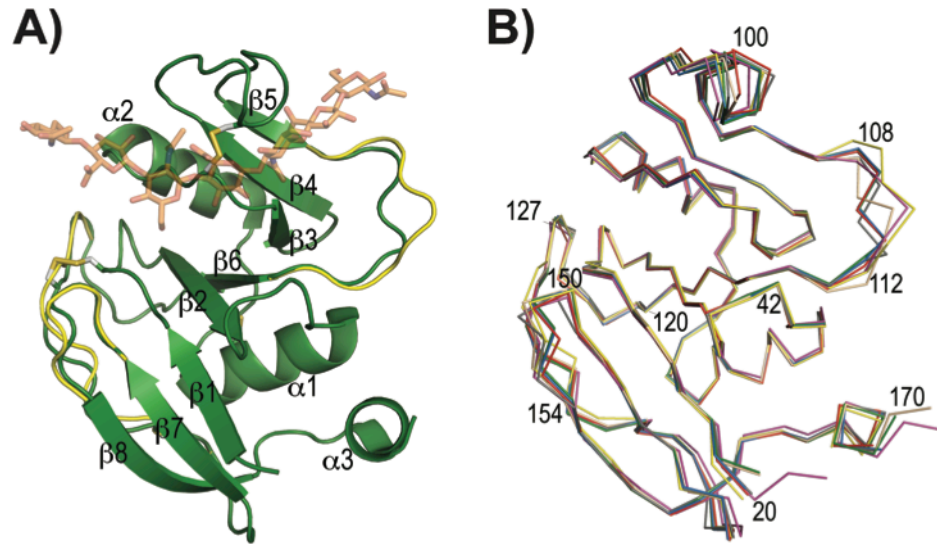
Figure 2-9. A) The structure of the $h\text{HABD}_{18\text{-}178}$ monomer (green; 4PZ3-A) showing secondary structure assignments and disulfide bond locations. The *apo* $m\text{HABD}_{24\text{-}174}$ (2JCP; yellow) is shown for comparison, but regions of the backbone that overlay the 4PZ3-A precisely are omitted for clarity. The HA binding position inferred from the complex of $m\text{HABD}_{24\text{-}174}$ with an oligosaccharide octamer (2JCR; orange) is shown only for positional reference. B) Comparison of the alpha-carbon backbone geometry in six crystallographically independent *h*HABD crystal structures: $h\text{HABD}_{20\text{-}178}$ (1UUH-A, grey; 1UUH-B, red) and $h\text{HABD}_{18\text{-}178}$ (4PZ3-A, blue; 4PZ3-B, green; 4PZ4-A, magenta; 4PZ4-B, wheat). Also shown is the backbone from the $m\text{HABD}_{24\text{-}174}$ *apo* structure (2JCP; yellow).
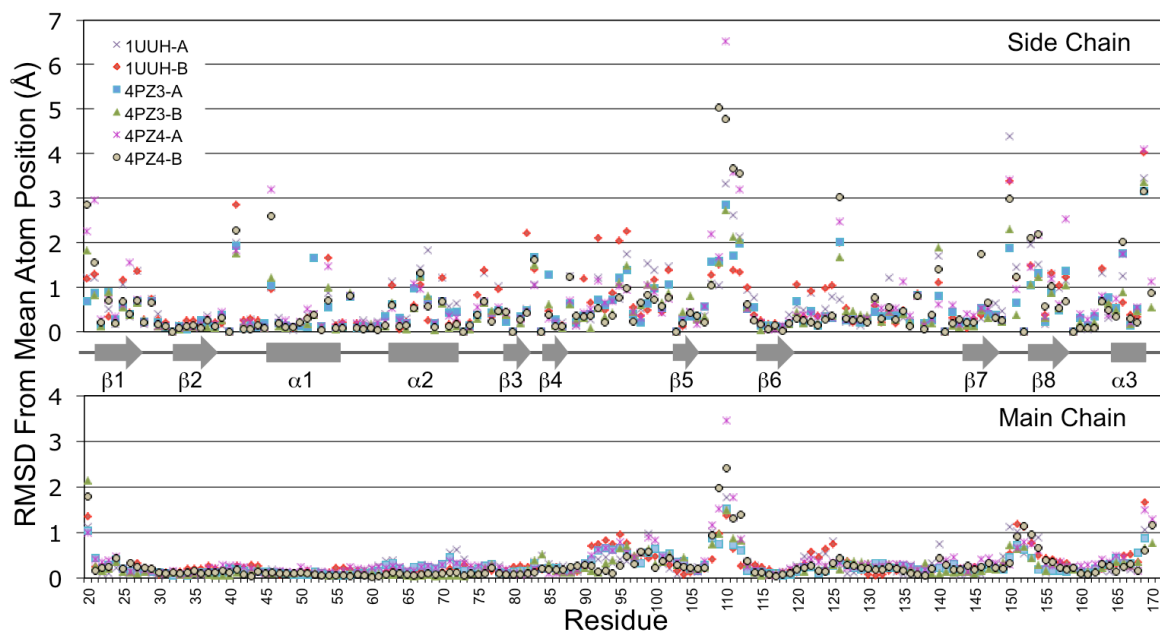
Figure 2-10. RMS differences between *h*HABD structures and a consensus mean structure. Differences by residue are separated to display main chain atoms (N, C$_\alpha$, C, O) and side chain atoms separately. The consensus was obtained by averaging atomic positions for each atom after co-superposition on 1UUH-A.

2.3.11 A peptide-binding site exists in the human CD44 hyaluronan-binding domain

Quite unexpectedly, an unidentified peptide (or likely a mixture of short peptides) lies in a hydrophobic pocket of one molecule (chain B) in the monoclinic crystal form of hHABD$_{18-178}$. This binding site is on the opposite side of the HABD from the HA binding groove (Figure 2-11 A). The peptide has been present in every monoclinic crystal we have examined, but never associated with chain A, where the pocket geometry is unchanged. The peptide likely co-purified along with the protein from the expression system and has been retained throughout refolding, gel filtration and crystallization, but we have been unable to separate it for unambiguous identification, or to replace the peptide with synthetic peptides introduced as potential surrogates. Speculation regarding the possible identity and significance of this peptide binding site is presented below.

2.3.12 Structural basis of the tetrasaccharide-murine CD44 hyaluronan-binding domain complex

In an attempt to better define an appropriate foci on CD44 for the binding of HA antagonists, we determined the crystal structure of a complex of murine HABD complex with HA$_4$ at 2.5 Å resolution. HA$_4$ was soaked into pre-formed monoclinic apo crystals, and the structure solved by molecular replacement. These crystals are of the same form previously reported in apo and HA$_8$-bound structures [32]. Overall, the mHABD structure

differs little from the previously reported complex with HA_8 (PDB-ID 2JCR). The protein structure can be superimposed upon the HA_8-bound atomic coordinates with a root mean square deviation of 0.18 Å for 129 paired Cα atoms. HA_4 appears to exclusively occupy saccharide subsites 5-8 of the larger HA_8 binding groove (Figure 2-12 A). There is no density suggesting even low occupancy in subsites 2-4. Glucuronic acid residues occupy subsites five and seven, while *N*-acetylglucosamine units occupy sub-sites six and eight. There is a small 0.6 Å shift of HA_4 toward lower numbered sub-sites, but key hydrogen bond interactions utilized by comparable residues in the HA_8 complex are maintained. The smaller o-HA appears sufficient to induce the "O conformation" [51] or the "type B" ligand-bound state previously described only in relation to HA binding of HA_8 [32]. The loop centered on Arg45 adjusts in concerted movement with Glu41, Asn43, Gly44 and Arg82 to create more intimate contacts with the polysaccharide in the HA_4 complex (Figure 2-12 B).
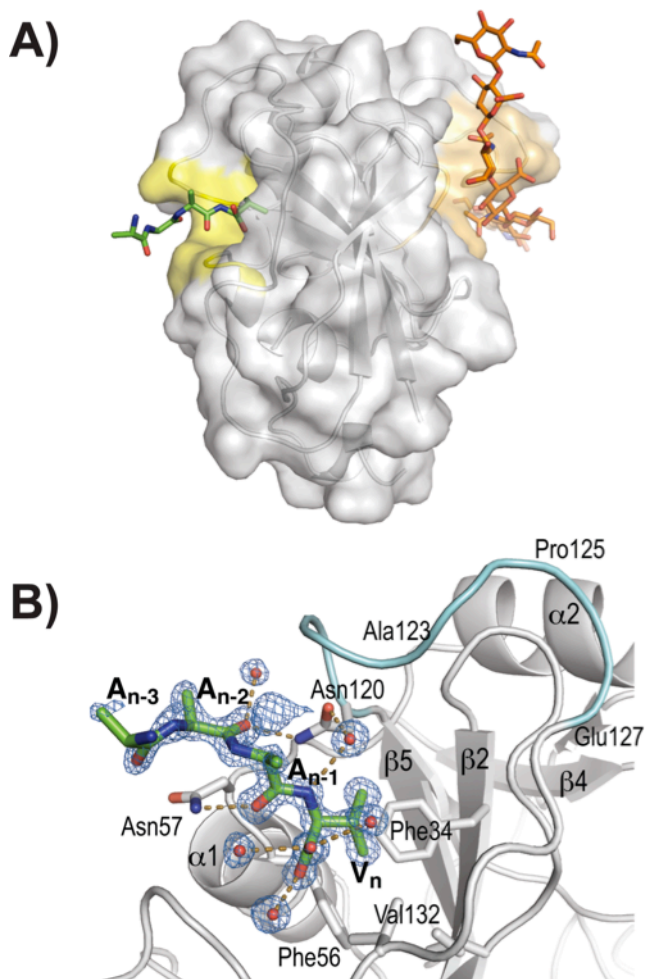
Figure 2-11. The unexpected binding site for a peptide on the back side of chain B in the monoclinic hHABD$_{18-178}$ (4PZ3-B). A) Binding site (yellow) in relation to the known HA binding site (orange). (HA position modelled by superposition of *murine* HABD structure 2JCR). B) Detail of H-bonding interactions (mostly water-mediated) and the 3σ omit ($F_O - F_C$) electron density (blue) validating the peptide geometry and model integrity. Backbone geometry for the three C-terminal residues (A_{n-2}, A_{n-1} and V_n) and the side chain of the valine are very well defined. The A6 peptide homolog on CD44 (residues 120-127) is colored cyan.
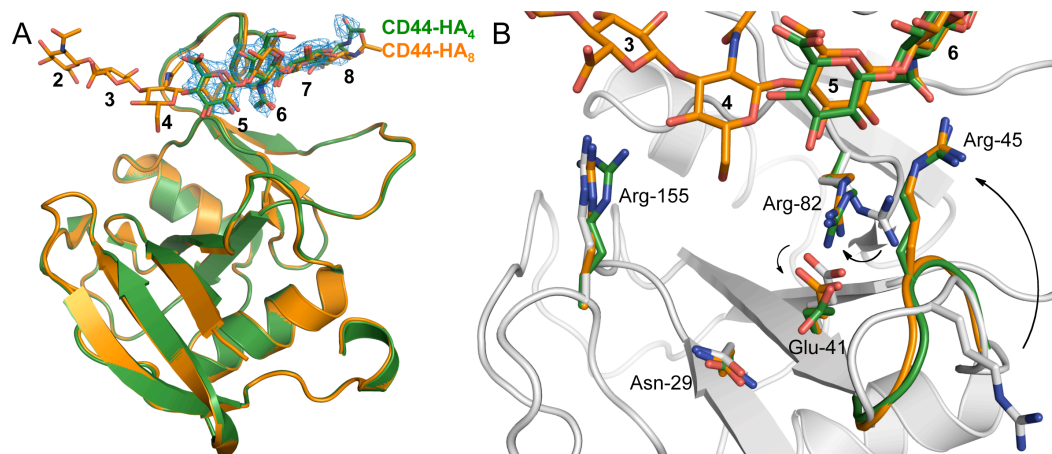
Figure 2-12. A) The crystal structure of murine CD44 HABD with bound HA_4 (green, PDB-ID 4MRD) aligned with previous described complex with HA_8 (orange, PDB-ID 2JCR) [32]. Saccharide sub-sites are numbered $2 - 8$ as in the HA_8-bound structure. Electron density ($2F_O - F_C$) of bound HA_4 is contoured to 1 σ (blue). **B)** A comparison of these structures to the *apo* structure (white; PDB-ID 2JCP). Only substructures that change significantly are shown. Arg82 and Glu41 shift in concerted motion with the loop supporting Arg45, which swings 12 Å toward the HA. The same change is induced by the binding of either HA_4 or HA_8. Arg155, and Asn29 are unafffected by o-HA binding.

2.4 Discussion

The binding of CD44 to the extracellular matrix promotes cell adhesion and tissue structural integrity [2, 41]. Small o-HA fragments can disrupt these interactions [33-35]. It is likely that many of the biological effects on cells attributed to o-HAs are simply due to the dissociation of receptors bound to the ECM that can occur upon o-HA binding. The ability of o-HA preparations of different lengths to alter cell avidity for fluorescein-labeled HA has been studied quantitatively by flow cytometry [34]. The multivalent binding of o-HA to cells is observed with oligo size larger than HA_{20}. Monovalent binding avidity is diminished upon exposure to o-HA of defined sizes including HA_6 or HA_8, and, to a significantly lesser extent, HA_4. Our measurements of o-HA binding provide the quantitative assessment of the specific binding affinity of oligosaccharides to CD44 HABD (Table 2-1), and these results are entirely consistent with measurements made with competition analyses using soluble p-HA [31, 33, 34]. The SPR surface with immobilized polymeric HA with concentrations fixed at ~1 mg/ml (115 RU) may better represent a compact matrix-like environment for use in competition assays than soluble p-HA.

A SPR surface onto which the extracellular hyaluronan binding domain of human and murine CD44 has been immobilized was used to measure absolute binding affinity to HA oligosaccharides of defined sizes. The K_D values presented here (Table 2-1) compare favorably to comparable measurements obtained by ITC and NMR [32]. The affinity of these oligosaccharides for the CD44 HA binding domain parallels their ability to interfere with polymeric HA binding to soluble HABD or CD44 expressing cells [34] (Table 2-1). Our SPR assays support monovalent binding of these sugars to the immobilized HABD

(data not shown), consistent with previous findings that monovalent binding occurred with saccharides up to 18 in length [34]. More importantly, our data confirm the integrity of the assays, as well as affirm the integrity of the protein, whose production is a major part of the project.

We have consistently measured lower binding affinities of oligosaccharides to immobilized *m*HABD than to *h*HABD (Table 2-1), even as the same relative affinities of different o-HA molecules are preserved. A possible explanation for this lies in the substitution of Asn101 in *h*HABD for histidine in *m*HABD. This conservative difference exists at the extreme end of the binding groove; in the *m*HABD complex with HA_8, His105 contacts the GlcNAc residue in subsite 8. This substitution could account for the lower affinity of any of the oligosaccharides we have studied for *m*HABD, since they all occupy this subsite upon binding.

Structures of both human forms were determined by molecular replacement, and refined against high-resolution diffraction data (1.60 Å and 1.08 Å, respectively). Summary refinement statistics for 4PZ4 and 4PZ3 are provided in Table 2-2. Both structures include two protein chains in the crystallographic asymmetric unit. The higher resolution of either structure significantly improves upon that achieved with the previously reported orthorhombic crystal form (2.20 Å; PDB-ID 1UUH) [31] (Figure 2-13). The structures recapitulate earlier findings of overall fold and secondary structure reported for both the human [31] and murine [32] CD44 HA binding domains (Figure 2-9 A). All four chains are well-ordered from Ala20 to Tyr169. The two additional residues at the N-terminus (Ala18-Met19) are disordered in all but one chain of the hexagonal form

(chain A). It is not at all clear how these residues might alter crystallization properties as they are not involved in specific intermolecular contacts in either of the new crystal forms (4PZ3 or 4PZ4), or the original orthorhombic form (1UUH). C-terminal residues are disordered beyond Ser-171 in all structures. In the hexagonal form, 13 residues are modeled with side chains in two discrete conformations. In the higher-resolution monoclinic form, 24 residues are so modeled.
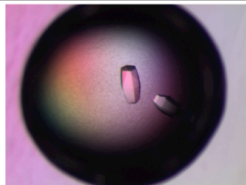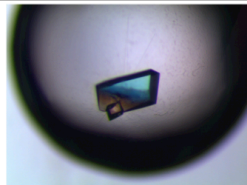
Published Crystal Forms of HABD		
	Apo Hu-HABD[20-178]	Apo/Holo mu-HABD[25-174]
Space group	$P\,2_1\,2_1\,2_1$	$P\,2_1$
Cell dimensions		
a, b, c (Å)	48.9, 77.3, 87.7	30.7, 82.1, 32.3
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 117.9, 90.0
Resolution (Å)	2.20	1.30
Molecule/Asym Unit	2	1

Novel Crystal Forms of Hu-HABD[18-178]		
Space group	$P\,6_1\,2\,2$	$C\,1\,2\,1$
Cell dimensions		
a, b, c (Å)	70.8, 70.8, 288.0	72.7, 59.1, 75.7
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 112.9, 90.0
Resolution (Å)	1.60	1.30
Molecule/Asym Unit	2	2

Figure 2-13. Summary of the current crystal forms of murine and human CD44 HABD

In both new crystal forms, the two protein chains are positioned face-to-face in the crystallographic asymmetric unit about a non-crystallographic dyad axis (Figure 2-14 A) to form almost identical dimers. In this packing arrangement, access to the HA binding groove of each monomer is blocked by the presence of the opposing monomer. Residues Cys77, Arg78 and Tyr79 sit squarely in the HA_5 and HA_6 subsites, and Ile96-Ala99 protrude from the long loop joining β4 and β5 to obstruct the HA_6 subsite (Figure 2-14 B, right). Gly40-Arg41 of loop β2-α1 occupies the HA_4 subsite (Figure 2-14 B, left), and likely prevents rearrangement of the inducible binding site accessible to small molecule inhibitors of HA binding in *murine* HABD crystals. We have been unable to form co-crystals with small molecule inhibitors that bind to the same site by either co-crystallization or soaking using either of these crystal forms, likely because access to the groove is occluded by intermolecular packing. No similar dimer exists in the orthorhombic crystals (1UUH); in those crystals, the HA binding site is also blocked by packing, but differently.

Large differences (greater than 3 times the mean RMSD) in the backbone conformation occur only at the N- and C-termini and in residues 108-112 of the β5-β6 loop (Figure 2-9 B). Apart from some side chains along the C-terminal helix, large differences are restricted to the side chains of Arg41, Arg150, and the loop segment with diverse backbones (108-112). While these are all unliganded *apo* structures (besides chain B of 4PZ3 to be discussed below), these are side chains known to shift in response to HA or inhibitor binding [32, 53]. Some contiguous portions of the backbone (e.g., residues 94-104, 122-127 or 150-154) have moderately large RMSDs (elevated, but less than 3 times the
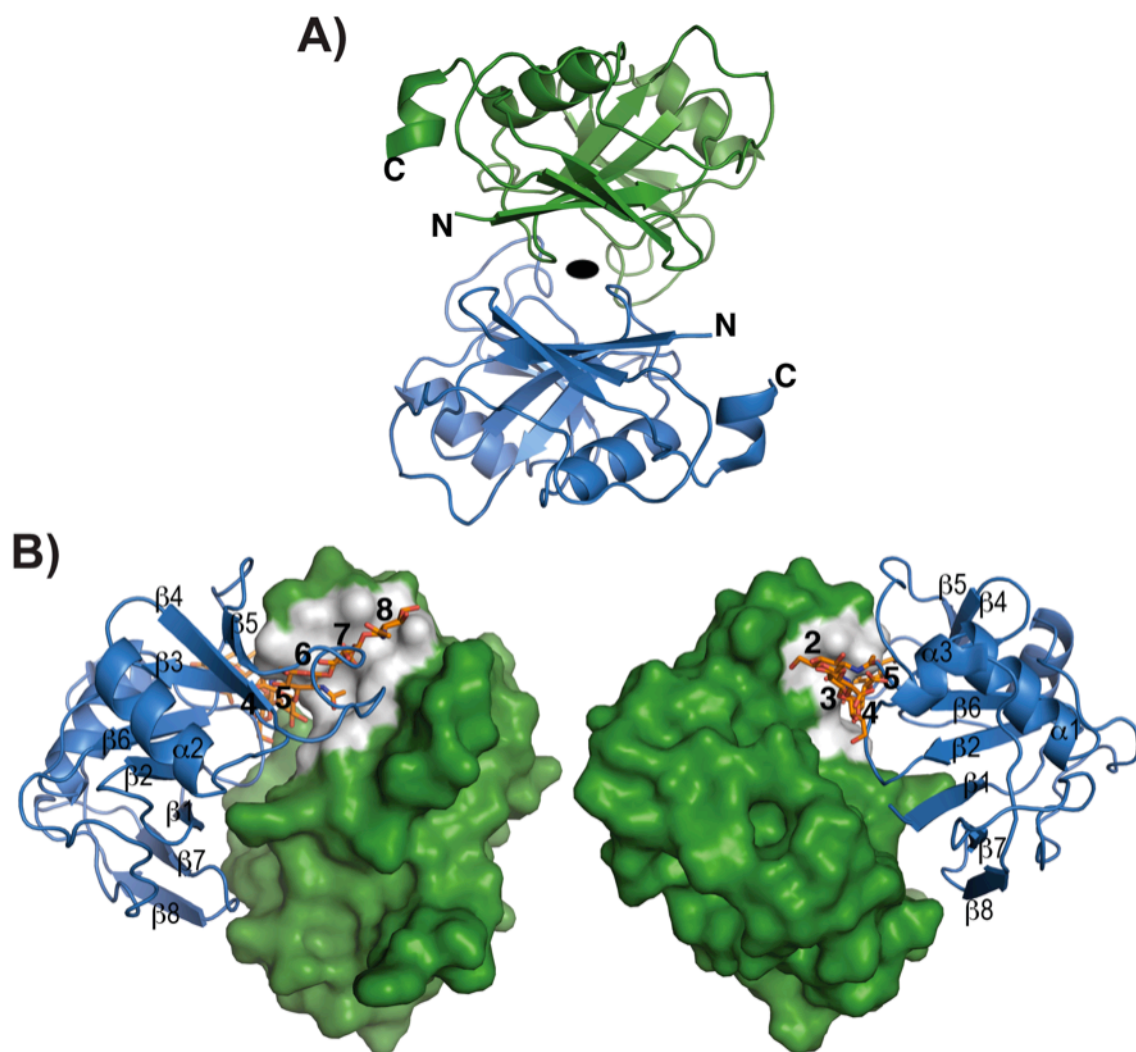
Figure 2-14. A) A similar non-crystallographic two-fold axis of symmetry relates two molecules in the asymmetric unit of both monoclinic and hexagonal crystal forms of hHABD$_{18-178}$. 4PZ3-A (green) and 4PZ3-B (blue) are shown as illustrative of the dimer in both forms. B) Two different views of the HA binding groove show how intermolecular contacts made within the dimer prevent HA oligosaccharides and small molecules from binding. HA (orange) is included for positional reference only using its location upon superposition of the mHABD$_{24-174}$ structure with bound HA (2JCR) onto 4PZ3-A (green). HA saccharide subsites (white) are numbered for reference in the text.

mean). These differences are not the result of any specific backbone torsional changes, but instead reflect small but concerted shifts in positions of each segment.

Others have characterized 120-127 as "potentially conformationally dynamic" [72], but the structure comparison does not support this. Moreover, the bracketing disulfide bonds on either side of this peptide leave little opportunity for flexibility. In general, the disulfide-stabilized link domain is very rigid and adopts the same conformation when different ligands are bound or in different environments (Figure 2-9).

Figure 2-9 B also includes a comparison to murine $mHABD_{24-174}$ structure (2JCP) [32]. Backbone differences are limited to portions of loops joining β5 to β6 (particularly residues 108-112, human numbering), β6 to β7, and β7 to β8. These are the same regions that show conformational diversity in different human structures suggesting that these regions have intrinsic flexibility even when the primary sequence is unchanged. The difference between murine and human CD44 conformation in the vicinity of 108-112 has been noted and attributed to the insertion of Val112 (murine numbering) in the murine protein (as shown by Banerji *et al.*, 2007). It is worth noting that there is no difference in backbone conformation in residues 40-42, which undergoes a significant change upon ligand binding [51, 53]. We see no evidence that this is an intrinsically flexible loop when no HA is bound.

The two CD44 HABD constructs share 86.2% sequence identity across 159 residues, and previous structural studies have confirmed that only small differences occur near the HA binding site [31, 32]. The yield from *E. coli* expression of *h*HABD is higher than *m*HABD, so it has been used in the competition studies with immobilized-HA that

require more protein. All subsequent co-crystallography has utilized *m*HABD. The monoclinic crystal form previously described [32] provides open access to the binding site upon the soaking of small molecules, even molecules as large as HA_4. While we have discovered alternate crystal forms, we as others [32], have not been able to prepare co-crystals with *h*HABD. The search for a human crystal form suitable for use in HA binding studies must continue.

We show with oligosaccharides that there is good correlation between binding interference as measured with immobilized p-HA on SPR *in vitro* (Chapter 2. Part I), and binding to CD44 measured by other means [31]. Others have shown that small oligosaccharide HA fragments that compete directly for HA also modulate CD44-mediated adhesion and motility in cells [33, 35]. Banerji *et al*. have shown that HA binding occurs across a shallow groove that spans the width of the HABD [32]. Our HA_4 complex structure further isolates subsites 5-8 as those preferred for smaller o-HA fragments that bind CD44 (Figure 2-12).

The peptide binding site is a pocket inside the concave curl of the extended HABD beta sheet and at the C-terminal end of helix α1. It is rimmed by portions of the long meandering loop joining β6 to β7 (residues 120-136). Our comparison of *h*HABD structures shows this to be one of the more rigid portions of the structure (Figure 2-10). Ile26, Arg29, Phe34, Phe56, Val132, surround a hydrophobic subsite for the C-terminal residue of the peptide, which we confidently identify as a valine based on high-resolution electron density (Figure 2-11 B). Threonine is an unlikely candidate as there is no possible partner for a side chain hydroxyl group in the hydrophobic subsite. The fact that

this is the C-terminus of the peptide is also convincingly shown by the presence of well-resolved water molecules bound to the carboxylate. Most hydrogen bonds stabilizing the peptide binding are water-mediated, although H-bonds from peptide backbone carbonyl oxygens to the side chain amides of Asn57 and Asn120 also exist. Other amino acids in the peptide cannot be identified with certainty based on electron density alone, and we cannot exclude the possibility that more than one species is bound; the peptide has been modeled as A_{n-3}-A-A-V_n, with all peptide bonds in a *trans* conformation. Longer side chains than alanine might fit the broken density, particularly in place of A_{n-1} (Figure 2-11 B).

This pocket and all residues bordering it are also conserved in murine CD44 HABD. A small molecule identified by fragment screening has been confirmed to bind in this location by co-crystallization with mHABD$_{24-174}$ (PDB-ID 4MRH), but binding by peptides has not been observed in any other crystal form, including the hexagonal form prepared from the same protein preparation. Peptide binding does not seem to be prohibited by crystal packing in any of these forms, so we cannot offer any reasoned explanation for the lack of binding at other essentially equivalent sites.

The newly identified pocket may represent a binding site for some yet-to-be-identified proteinaceous partner of CD44, which is thought to be involved in several different signaling pathways and regulatory protein assemblies [28]. A number of studies have been conducted to identify peptides that specifically interfere with CD44-mediated cell motility or tumor migration [72-74] but the site of action of these peptides is not known. We considered the possibility that the A_{n-3}-A-A-V_n fragment might represent the C-terminal

end of a known protein-binding partner. Several hexapeptides selected using an alignment of sequences to known CD44 partners (e.g. fibronectin, MMP-9 and CD44) were evaluated. Some contain Ser-Val instead of Ala-Val as the last two residues that were built in the initial model, but the density of the β-carbon where Ser was modeled turned out to be uncertain in higher resolution structures. Nevertheless, none were shown to bind hHABD (Table 2-4).

Peptide "A6" (acetyl-KPSSPPEE-NH$_2$) has antitumor activity [75, 76] and inhibits CD44-medited migration and metastasis [72]. It has been suggested that A6 may bind to CD44 and facilitate dissociation of CD44 homodimers that repress biological activity and HA recognition [72]. No direct binding against hHABD$_{20-178}$ by A6 as measured in the SPR HABD binding assay (LKL-Pep-2). This does not rule out the possibility that binding occurs elsewhere on CD44. A6 shares sequence homology with CD44 residues 120-127 (NASAPPEE), so it has also been proposed that it may act as a decoy to antagonize the association of CD44 with another protein whose identity is not yet known [72]. These residues surround half the outer rim of the binding site of the peptide we have detected crystallographically (Figure 7B), lending support to a hypothesis that this region of the HABD surface is the site for a biologically significant protein-protein interaction. The subsite surrounding the peptide found in the monoclinic crystal form may be the epicenter of this interaction.

Table 2-2. Crystallographic Data and Refinement Statistics of *h*HABD [18-178]

Data collection and processing

	*h*HABD (Monoclinic)	*h*HABD (Hexagonal)	*m*HABD-HA$_4$
PDB ID code	4PZ3	4PZ4	4MRD
Diffraction source	APS 17-ID	APS 23-ID-B	ALS 4.2.2
Space group	*C* 1 2 1	*P* 6$_1$ 2 2	*P* 1 2$_1$ 1
a, *b*, *c* (Å)	72.50, 60.29, 75.13	70.27, 70.27, 286.46	30.62, 82.28, 32.52
α, β, γ (°)	90, 113.40, 90	90, 90, 120	90, 117.92, 90
Resolution range (Å)	44.67-1.08	286.46-1.598	41.14-2.55
(highest shell)	(1.14-1.08)	(1.604-1.598)	(2.61-2.55)
Total No. of reflections	376553 (34299)	720488 (5519)	17738 (1092)
No. of unique reflections	122395 (15623)	56356 (455)	4710 (332)
Completeness (%)	97.3 (85.5)	99.3 (80.4)	98.7 (94.6)
Redundancy	3.1 (2.2)	12.8 (12.1)	3.8 (3.3)
$I/\sigma(I)$	21.2 (6.7)	26.4 (5.3)	11.1 (3.4)
R_{merge}	0.031 (0.154)	0.063 (0.490)	0.123 (0.391)
$R_{\mathrm{r.i.m.}}$ ‡	0.038 (0.208)	0.066 (0.544)	0.143 (0.462)
Wilson *B* factor (Å^2)	12	28	31

‡ Estimated $R_{\mathrm{r.i.m.}} = R_{\mathrm{merge}}[N/(N-1)]^{1/2}$, where N = data multiplicity.

Structure solution and refinement

Resolution range (Å)	68.96-1.08	60.860-1.600	41.14-2.55
(highest shell)	(1.11-1.08)	(1.640-1.598)	(2.62-2.55)
No. of reflections, working set	116222 (6803)	53422 (3738)	4355 (253)
No. of reflections, test set	6171 (339)	2864 (180)	237 (12)
Final R_{cryst}	0.166 (0.187)	0.197 (0.260)	0.187 (0.224)
Final R_{free}	0.177 (0.190)	0.229 (0.284)	0.245 (0.393)
Cruickshank DPI (Å)	0.03	0.08	
No. of non-H atoms	2596	2772	1229
No. of solvent waters	372	199	10
R.m.s. deviations			
Bonds (Å)	0.012	0.013	0.012
Angles (°)	1.5	1.4	1.4
Average *B* factors (Å^2)	11.6	18.4	14.0
Ramachandran plot			
Most favoured (%)	94.0	95.8	92.5
Allowed (%)	5.2	3.8	6.1
Outliers (%)	0.8	0.4	1.4

Table 2-3. Sequence Comparison of Constructs Used to Generate Different Crystal Forms

Construct	Sequence	Space Group	Reference
hHABD$_{20\text{-}178}$	20- AQIDL......PEDIYPSNPTDDDV-178	$P2_12_12_1$	Teriete *et al.*, 2004
hHABD$_{18\text{-}178}$	18-**AM**AQIDL......PEDIYPSNPTDDDV-178	$P6_122$; $C2$	
hHABD$_{20\text{-}168}$	20- AQIDL......PEDI -168		
mHABD$_{24\text{-}174}$	23- **MN**QIDL......QEDID -174	$P2_1$	Banerji *et al.*, 2007

Non-native residues are shown in bold.

Table 2-4. Binding Analysis of Purified Peptides against *h*HABD

No.	Sequence	ΔT_m (°C)	K_D by SPR (mM) *h*HABD/*m*HABD		Reference
			PBS	**HBS-EP**	
LKL-Pep-1	Ac-AASV	+0.5	> 20/5.5	> 20/> 20	From the peptide complex
LKL-Pep-2	Ac-KPSSPPEE-NH$_2$	−2	> 20/> 20	> 20/> 20	[72]
LKL-Pep-3	Ac-NASAPPEE-NH$_2$	−8	> 20/> 20	> 20/> 20	[72]
LKL-Pep-4	HTSVQT	+0.5	insoluble	> 20/> 20	Fibronectin; PDB: 1FBR
LKL-Pep-5	KNSVGR	+0.5	> 20/> 20	> 20/> 20	Fibronectin; PDB: 1FBR
LKL-Pep-6	KESVPI	−0.5	> 20/3.4	> 20/> 20	Fibronectin; PDB: 1FBR
LKL-Pep-7	HSSVPE	−9.5	> 20/no fit	> 20/> 20	MMP9; PDB: 1L6J
LKL-Pep-8	STSVTD	+0.5	4.3/no fit	> 20/0.67	CD44; PDB: 1UUH
LKL-Pep-9	Ac-YDSeYeSe				[71]
LKL-Pep-10	Ac-FNLPLPSRPLLR				[77]

CHAPTER 3

DISCOVERY OF INHIBITORS OF HA BINDING TO CD44

3.1 Background

Upon review of the available structural data, one might conclude that CD44 is not an easily druggable target. The murine HABD-HA complex reveals an extended HA binding site with surface area exceeding 800 Å^2 and molecular binding stabilized by a large number of weak interactions involving at least seven consecutive saccharide units of HA. The HABD has no well-formed or deep pockets that would serve as attractive binding sites for small molecule inhibitors, and is known to undergo small but important conformational changes upon binding HA. In many respects, the protein-polysaccharide complex resembles protein-protein interactions that are notoriously difficult to disrupt effectively with small molecules.

Given the strong precedent for CD44 HABD structural biology, we have chosen a fragment-based approach to identify possible sites for small molecule binding [54]. Fragment screening, when coupled with the crystallographic characterization of low affinity hits, permits a direct empirical observation of conformational adjustments needed to accommodate specific bound small molecules, and has proven applicable in the structure-based design of inhibitors of protein-protein interactions [55, 56]. Knowledge of the structure of complexes with bound molecules can be exploited to combine individual molecular features of different molecular fragments using a well-characterized spatial arrangement that best compliments the altered geometry of a larger protein-binding site, leading to improved ligand design.

Biophysical screening techniques such as NMR, X-ray crystallography, and ITC have been used by others to directly detect weak protein-ligand interactions and are compatible with moderate to high throughput technology [78]. However, direct screening generally requires large amounts of the target proteins to screen a sizable library of potentially active compounds. Previously, we had introduced the use of DSF as a quick and economical assay to evaluate CD44 HABD folding and stability (Chapter 2). Thermal shift assays have also been used to qualitatively detect ligand interactions in other systems [59, 70, 79, 80]. The primary advantages of screening by DSF are the sensitivity to detect weak binding, speed (96-well plate format), and relatively small amount of protein required. A disadvantage may be that thermal shift assays can produce false negatives. Another disadvantage is that nothing is learned about the capacity of compounds to inhibit protein function; only the existence of an interaction is indicated by the presence of a thermal shift.

Other efficient screening campaigns have utilized surface plasmon resonance [57, 60]. The SPR assay using immobilized polymeric HA established by Banerji *et al.* measuring CD44-HA interactions [32] (and implemented as described in Chapter 2), could be used to identify compounds that inhibit the association of CD44 HABD and hyaluronan, but the amount of protein needed to evaluate the interference of hundreds or thousands of compounds is prohibitive. An alternative format where the HABD is immobilized on the surface and the binding by small molecules is detected, would require only enough protein to cover the surface of chips used in the instrument. As with DSF, no information about the relevance to inhibition is extracted from the binding affinity alone, but an SPR

assay could also be used to validate primary hits through more rigorous dose-response experiments and determine an accurate binding affinity (K_D).

Our choice of chemical library was the Maybridge Ro3 Diversity Library of 1000 small molecular "fragments". This is a much smaller chemical library than that used in conventional high throughput screening. These compounds are small (average MW ~ 150 Da), "rule of 3" compliant [81] and structurally diverse to span a lot of chemical space. Many of these molecules are generally flat, containing one or two aromatic rings that share little similarity to the carbohydrates that are the natural ligands. To add spatial diversity to the Maybridge library, we have included an additional fifty Ro3 fragments from Infarmatik. These compounds are somewhat larger (average MW ~ 300 Da), and most of them are saturated heterocyclic compounds or contain spiral hydrocarbons.

We have employed biophysical screening assays that rely on DSF and SPR to identify small molecules from the libraries that bind with low affinity to the CD44 HABD. These hits were then validated and characterized by crystallographic analysis, resulting in the identification of four unique sites of small molecule binding, including what might be a functionally relevant inducible binding site to which a distinct pharmacophore binds. Co-crystal structures of CD44 with these hits provide detailed information regarding previously unknown conformational changes and molecular interactions that stabilize binding that provide a structural guide for the design of more potent analogs.

3.2 Materials and Methods

3.2.1 Materials

Maybridge Ro3 diversity fragment library of approximately 1000 compounds ($\geq$ 95% purity) was purchased from Thermo Fisher Scientific (Pittsburgh, PA). Forty of these compounds not soluble at a concentration of 200 mM were excluded from the screen. An additional 50 compounds ($\geq$ 95% purity) were purchased from Infarmatik (Budapest, Hungary). Compounds from both libraries were dissolved in 100% DMSO to prepare stock solutions at 200 mM that were sealed in 1.4-ml MATRIX storage tubes (Thermo Scientific) with removable septa and stored at $-$ 20ºC.

Crude HA fragments containing a mixture of HA_6 and HA_8 were a gift of Dr. Xuefei Huang from Michigan State University. Short oligosaccharide assortments of HA_4, HA_6, HA_8, HA_{10} and HA_{12} with low endotoxin were purchased from Cosmobio (Japan).

General crystallization and buffer reagents were from Sigma (St. Louis, MO) or Hampton Research (Aliso Viejo, CA). Individual hits from the Maybridge Ro3 diversity fragment library ($\geq$ 95% purity) were re-purchased from Thermo Fisher Scientific (Pittsburgh, PA) for further characterization. Purity was determined by analytical HPLC or NMR techniques.

3.2.2 Fragment screening using Differential Scanning Fluorimetry

A fragment screen with *h*HABD and the combined Maybridge Ro3 and Infarmatik library was conducted using a standard BioRad C1000 Thermocycler real-time PCR instrument. Fragment screening was done with individual fragments of 5 mM in 2.5% DMSO. The melting transition for the protein was monitored by fluorescence using a SYPRO Orange dye. A 20-μl master mix that contains *h*HABD (0.1 mg/ml) and a fluorophore dye SYPRO Orange (10X) was mixed 1 : 1 with a 20-μl compound solution (10 mM of each individual compound in 100% DMSO) in the protein buffer comprising 20 mM Tris-HCl and 150 mM NaCl, for the total well volume of 40 μl. The scan was conducted with 1°C steps and 30-second resting time. To increase the throughput, the scanning temperature began from 20 to 80 °C instead of 5 to 95 °C. The resulting fluorescence curve was fit to a non-linear regression using a Boltzmann model [79] on a curve-fitting program Bio-Rad CFX software 2.0. The melting temperature shift induced by each individual compound was then compared to that of the vehicle control with DMSO present on each plate (an average of 16 replicates).

3.2.3 Fragment screening using Surface Plasmon Resonance

The same library of compounds was screened using the immobilized-HABD SPR assay on a Biacore 3000 using chips bearing immobilized *h*HABD as described in section 2.2.9. Each compound was diluted from 200 mM stock in 1.05X PBS buffer with detergent P-

20, to create 5 mM fragment solutions in 5% DMSO. Any compounds not soluble in this concentration were excluded from the screen. Fragments plated in 96-well plates were centrifuged, re-plated with supernatants and screened in duplicate at this concentration. The compound solution was injected over the HABD-coated surface for 30-second followed by a 60-second dissociation time and the flow sequence was repeated for each compound. This was pumped with PBS buffer containing 0.005% P-20 and 5% DMSO at a flow rate of 30 μl/min. A positive control comprised of crude HA_8 at 500 μM was included every tenth sample throughout screening to calibrate changes of surface activity. A new surface was refreshed at intervals with monitoring the surface activity as discussed in Section 2.3.7. The SPR response generated from a ligand can be monitored in parallel using up to three different channels (or flow cells). Murine HABD can be immobilized as done with human HABD in a separate flow cells (usually flow cell 3).

Primary hits from SPR were selected from those giving signals above the detection cut-off (9.85 RU). The cut-off was calculated from a response of averaged negative controls (buffer only, 200 replicates for the fragment library) plus three times of standard deviations [57]. Compounds that generated responses above this threshold were designated as "primary hits". Compounds producing very high signals larger than three times that of HA_8 ($\sim$ 100 RU) were excluded, as this suggesting a possibility of promiscuous binding.

3.2.4 Affinity determination in Fragment screening

Binding specificity was validated with a more rigorous dose-response experiment for human HABD using 2-fold serial dilutions of compounds (from 78 µM to 10 or 20 mM) for a 60-second association and dissociation period in triplicate. Reflectivity response data points were extracted from response curves in the 5 seconds prior to the end of the injection to determine steady state binding (Figure 2-2). All the data were solvent subtracted and referenced with blanks using standard procedures with BIAevaluation software v4.0.1 [60]. The equilibrium binding constant K_D of hHABD was measured by fitting the calibrated data curve with nonlinear regression assuming one-site saturation binding. The fit of a dose-response curve of each individual hit was used to judge if the binding is specific; compounds fit with a $\chi^2 < 4$ with a binding plateau, were designated a "confirmed hit". The affinity can be determined similarly with either hHABD or mHABD.

3.2.5 Fragment co-complex preparation of the human CD44 hyaluronan-binding domain

Co-complex crystallization with *h*HABD can only be achieved with soaking at very high ligand concentrations. Compounds were re-purchased so that pure compounds can be used to create 1 to 4 M stock solutions in DMSO. Preformed hexagonal *h*HABD crystals were transferred into a well solution droplet with a final ligand concentration between 50 mM and 200 mM and soaked for 3 h to overnight at 4°C. Crystals were cryo-protected by adding 10% glycerol before flash freezing in liquid nitrogen. These crystals in complex with fragments were reproducible under these soaking conditions.

3.2.6 Fragment co-complex preparation of the murine CD44 hyaluronan-binding domain

Apo *m*HABD crystals were prepared as described in Section 2.6.3. Small molecule co-complexes can be prepared by soaking *apo* crystals in a well solution to which 50 mM to 200 mM compound has been added for 3 to 12 hours at room temperature. Crystals were flash frozen after brief transfer to 10% glycerol in the presence of equal concentrations of compounds. Several crystals were always prepared and the best chosen for data collection.

3.2.7 *In vitro* SPR assays for inhibitor characterization

Small molecule binding affinities (K_D) were determined as described in Section 3.2.4. Competition studies were analyzed similarly as different lengths of HA oligosaccharides and fit into a 1 : 1 binding isotherm (See Section 2.2.7).

3.3 Results

3.3.1 Primary hits from Differential Scanning Fluorimetry

The DSF screen is efficient in terms of time and materials. The amount of proteins used for an entire 96-well plate was less than 200 μg. Eighty conditions of the fragment library can be screened in parallel in less than 60 minutes. The fragment library of 1000 compounds plus the Infarmatik focused library (50 compounds) were completed in three days in duplicate. Although ligand binding can cause an upward (stabilizing) or downward (destabilizing) shift in the melting temperature from protein-only samples Figure 3-1 A, we mainly looked for compounds that led to protein stabilization and defined our selection threshold as twice the standard deviation (1°C) in increase the T_m over all compounds screened. The T_m shifts for each of 1000 compounds is summarized in Figure 3-1 B. Sixteen primary hits were found to cause a T_m shift of ≥ 2°C under the conditions of this screen (1.6% hit rate, Table 3-1). These compounds did not show much

auto-fluorescence or interactions with the SYPRO dye. Throughout the ensuing

discussion, DSF-derived hits will be referenced by number with a "D" prefix as defined

in Table 3-1.
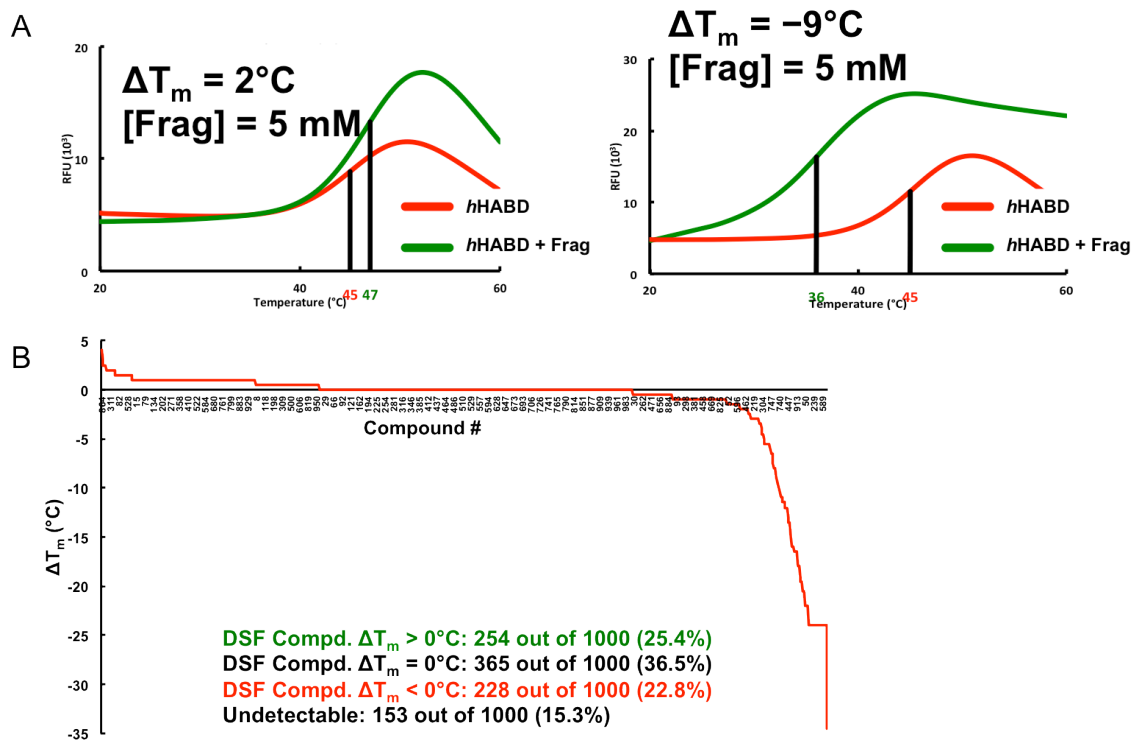
Figure 3-1. (A) Representative thermal shift curves of a stabilizing (left, BTB14321) and destabilizing (right, KM08552) fragments. (B) DSF fragment screen against the Maybridge library using *h*HABD. The thermal shift magnitude was organized in a descending order in the histogram.

3.3.2 Primary and confirmed hits from Surface Plasmon Resonance

   The Maybridge Ro3 Diversity Library was screened to identify compounds that bind to
surface-immobilized hHABD at a fixed initial concentration (5 mM). The SPR assay
permitted efficient screening; only two sensor chips and less than 100 μg of hHABD
were consumed in the screening of 1000 compounds. A cut-off response threshold (9.85
RU) was selected [57] to achieve an initial 4.1% hit rate (Figure 3-2 B). Compounds giving
a response above the threshold in duplicate measurements with a well-formed binding
curve, baselines and minimal nonspecific binding [82] were examined for dose-dependent
responses that fit a 1 : 1 binding isotherm. Twenty-one compounds listed in Table 3-2
met these criteria and were "confirmed hits" for more detailed characterization. Nine
compounds have ligand efficiency (LE) of 0.32 or greater. A Z-score of 0.65 was
computed from both positive and negative controls interspersed throughout the screen [57],
suggesting good quality of the screen. Ten out of fifty Infarmatik fragments passed the
cut-off and were selected for ligand titration. The data is summarized in Table 3-3. Some
of the hits are very potent against hHABD as measured by SPR, while none of the hits
have more than 2°C T_m shift. Throughout the ensuing discussion, SPR-derived hits will
be referenced by number with a "S" prefix as defined in Table 3-2.

3.3.3 Comparison of the two independent primary screens

We compared the fragment hits identified with the DSF method to those obtained by SPR at the same concentration (5 mM). Using the criteria established, out of 1000 compounds we selected 21 confirmed hits using SPR compared to the 16 hits using DSF. Combining the results from both assays gives a total of 34 unique and 3 common binders. Eight of the DSF hits were later shown to give an SPR dose response: five compounds were eventually be co-crystallized (Table 3-1). Compounds can produce a variety of different responses by SPR (Figure 3-2 A). Only four of the SPR-confirmed compounds give a T_m shift that meets the cutoff criteria of larger than 2 degrees (S1, S3, S4, S5) (Table 3-2). Twenty-one SPR confirmed hits can be readily quantified by their melting temperature shift, in which seventeen of these failed to give a T_m shift larger than 2°C (Table 3-2). There is no direct correlation between the affinity calculated from SPR and the magnitude of T_m shifts. Some SPR hits can be equivalent to DSF negative, suggesting a possible destabilizing interaction. Our results clearly indicate that either SPR or DSF are capable of identifying specific and unique binders, but that neither method detects all compounds that bind.
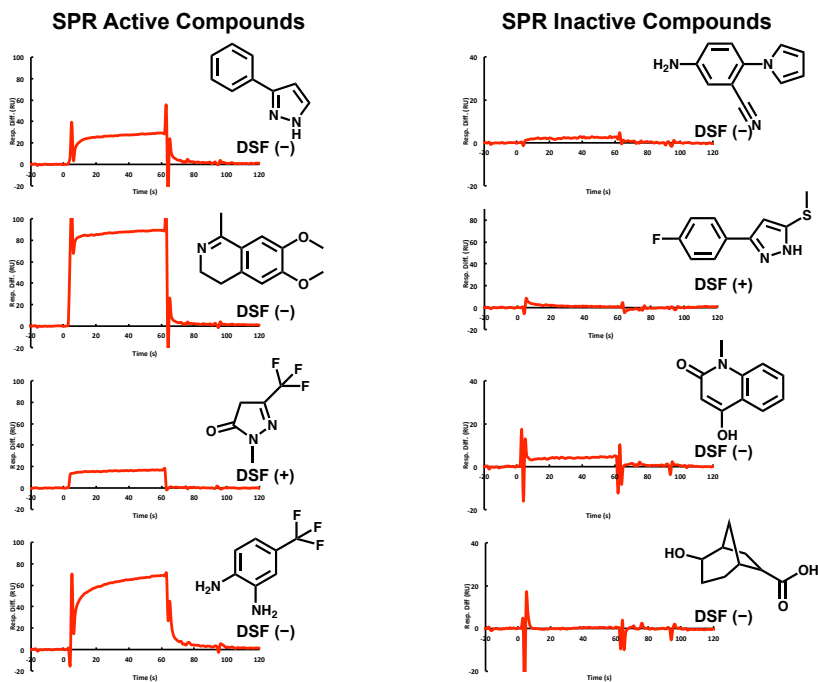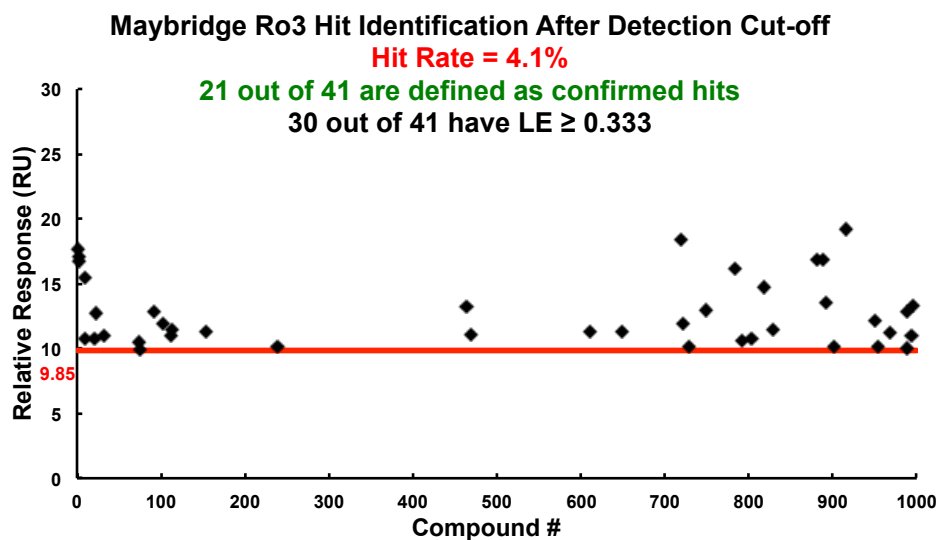
Figure 3-2. *In vitro* SPR assay is sensitive to small molecule in contact with immobilized HABD (A). This surface was thus used to first screen the Maybridge fragment library and to characterize binding affinity (B).

3.3.4 Fragment binding sites in the human CD44 hyaluronan-binding domain

We first attempted to prepare co-crystal structures with human HABD including all 34 primary screening hits identified from both screening methods, and five compounds were successfully co-crystallized. These compounds occupy two distinct sites (Figure 3-3 A), including a site far from the HA binding site. Compounds D1, D2 and D5 bind similarly in a pocket (Site 1) at the bottom of the protein with a common interaction with Asp160. It is a shallow pocket joined by both terminal ends of the construct that consists of mostly the hydrophobic and negatively charged residues. The di-amine of D1 donates hydrogen bonds to the carboxylate of Asp160 and the complex is stabilized by a water molecule (Figure 3-4 A). The tryptamine of D2 interacts with Asp160 through an H-bond and the benzene is in close proximity with a nearby tyrosine hydroxyl of Tyr169 (Figure 3-4 B). D5 binds slightly better ($K_D = 5.0$ mM) than that of D1 or D2 ($K_D = 8.5$ and 15.0 mM, respectively); the carboxylate forms a weak hydrogen bond with the Asp160 carboxylic group, likely because Asp160 forms a salt bridge with a nearby Arg162 that increases its basicity (Figure 3-4 C). The site seems unlikely to be relevant in HA binding.
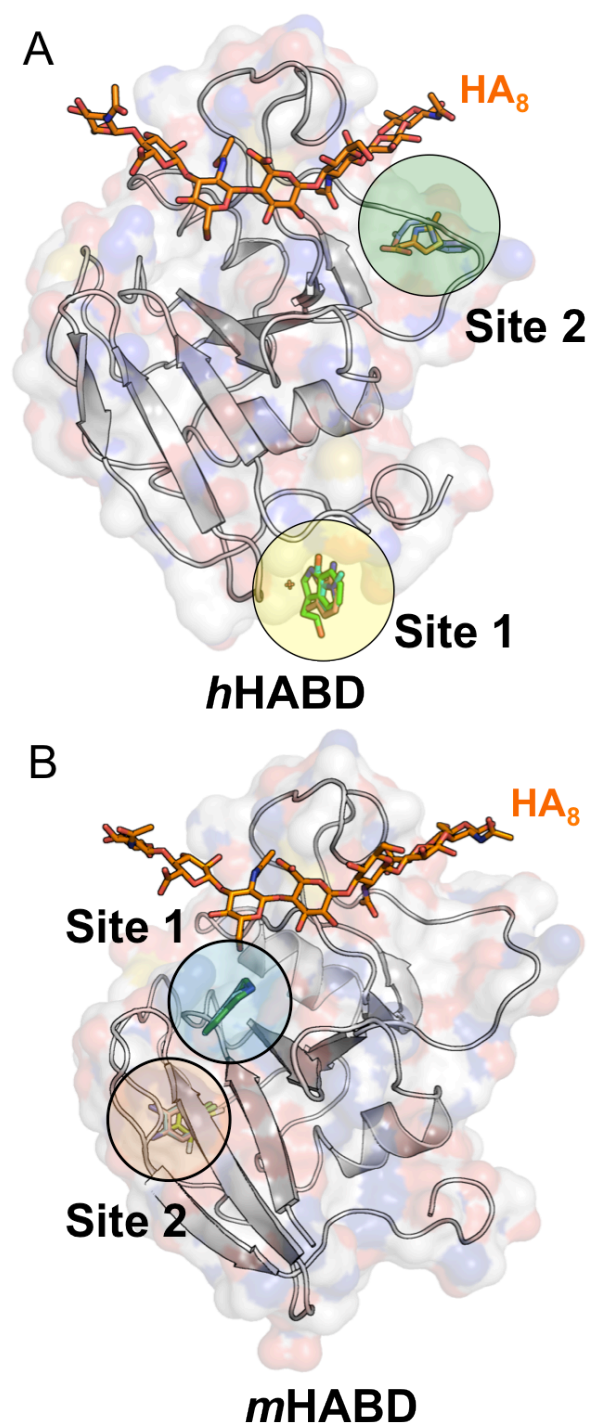
Figure 3-3. Crystal structures of individual compounds allow identification of two target binding sites in both *h*HABD (A) and *m*HABD (B).
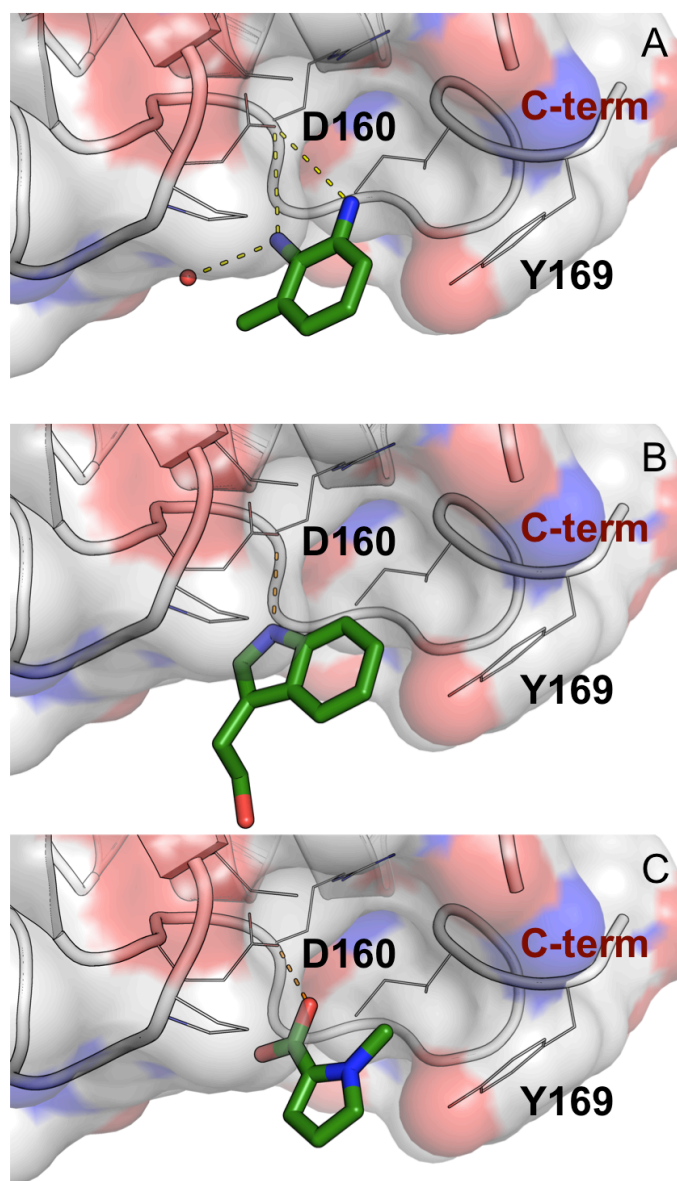
Figure 3-4. Co-crystal structures of *h*HABD with primary fragment hits selected from DSF. Hydrogen bonds to ligands are identified with dashed lines. Bound water molecules are shown as small red spheres. Compounds that bind to the Site 1 interact with D160 and residues close to C-terminus, including D1 (A), D2 (B), and D5 (C).

Fragments D17 and D18 bind to a pocket between two crystallographic symmetry molecules but interact favorably with one monomer (Figure 3-3 A, Site 2). The crevice is flanked by the region of β5 and the loop Leu106-Ser109, and is separated from the HA binding groove (the GlcNAc-8 subsite) by the loop (white, Figure 3-4 D–E). The other monomer resides closely to the pocket by α1 and the C-terminal loop (wheat, Figure 3-4 D–E). Both fragments contain a carboxylate group, but the thiazole fragment D17 binds more deeply and forms three H-bonds to the mainchain and a Ser109 hydroxyl (Figure 3-4 D). The benzylthiol carboxylate of D18 accepts only one H-bond from His86 (Figure 3-4 E). Binding to the relatively solvent-exposed site was validated by unambiguous electron density of the ligands.

3.3.5 Murine fragment binding sites

All 34 primary screening hits including 20 SPR primary hits that did not give a dose response were selected for co-crystallization trials with murine crystals. Five compounds identified through this initial screening using *h*HABD were successfully co-crystallized with *murine* HABD, including compound D1 that was initially found in the *h*HABD Site 1 but that also binds at a different site of *m*HABD. Ten fragment hits of the Infarmatik library were subjects of repeated soaking and co-crystallization experiments motivated by the high potency as determined by SPR, but binding of none of them could be confirmed by crystallography.

Two compounds (compound S1 (or D1) & S2; Table 3-2) occupy the same site adjacent to the HA binding groove (Site 1, Figure 3-3 B). One exocyclic amine of aniline S1 makes a hydrogen bond with the main chain oxygen of Val30 and two bound water molecules, while the other amine interacts with another buried water (Figure 3-5 A). The 1,2,3,4-tetrahydroisoquinoline (THIQ) of compound S2 binds with the aromatic ring in the same location beneath Arg155. This structure reveals a preferred half-chair conformation for the saturated six-membered ring that positions the endocyclic amine directly above the carboxylate of Glu41 to which it may donate a hydrogen bond (Figure 3-5 B). The induced apo conformations of Arg155 and Glu41 are occupied in roughly equal proportions in this complex, suggesting that occupancy of compound S2 is also less than 100%.
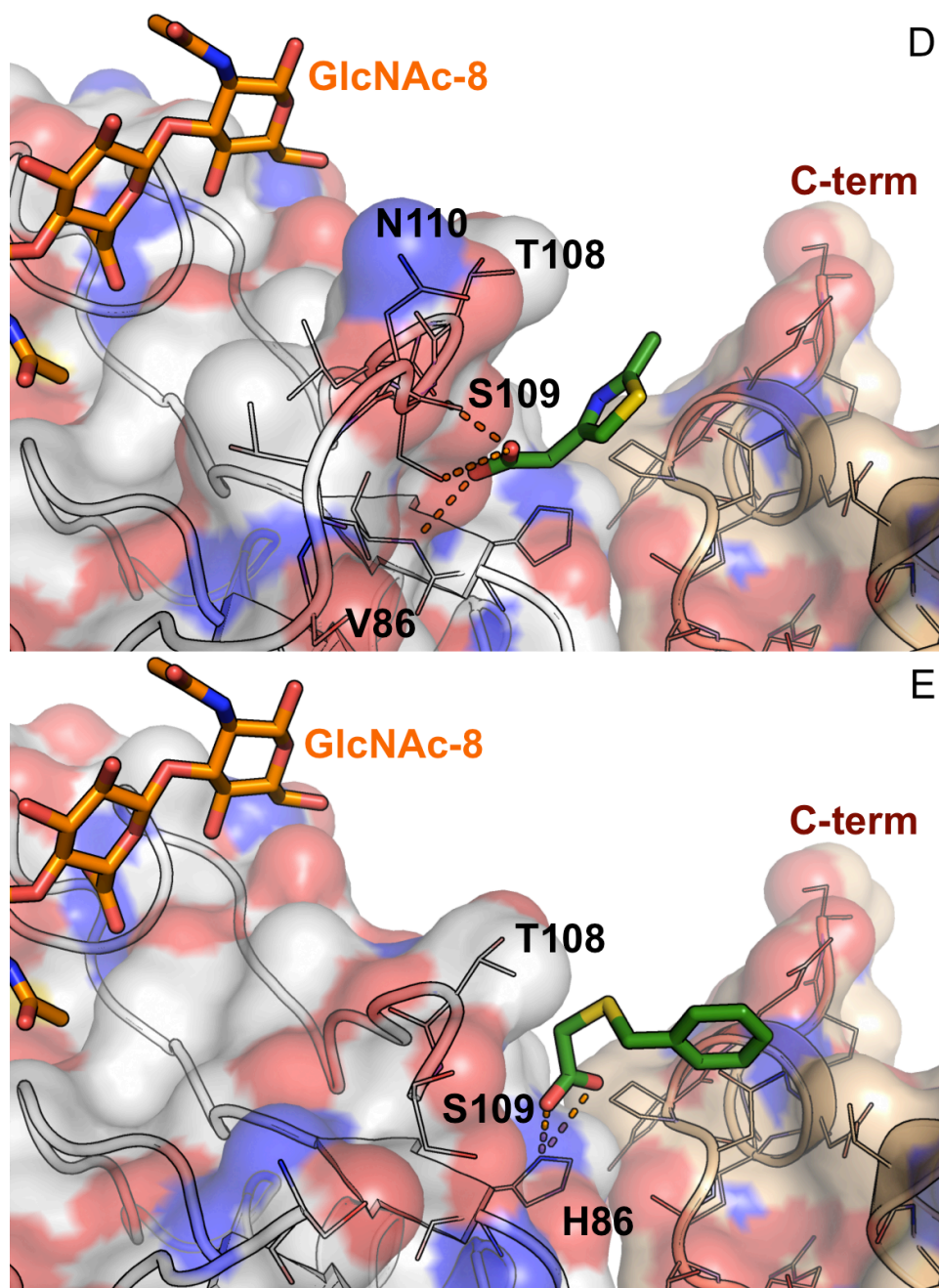
Figure 3-4. (Continued) Co-crystal structures of *h*HABD with primary fragment hits selected from DSF. Two compounds, D17 (D) and D18 (E) bind to Site 2 at the interface between two symmetry-related monomers (one white, one wheat).
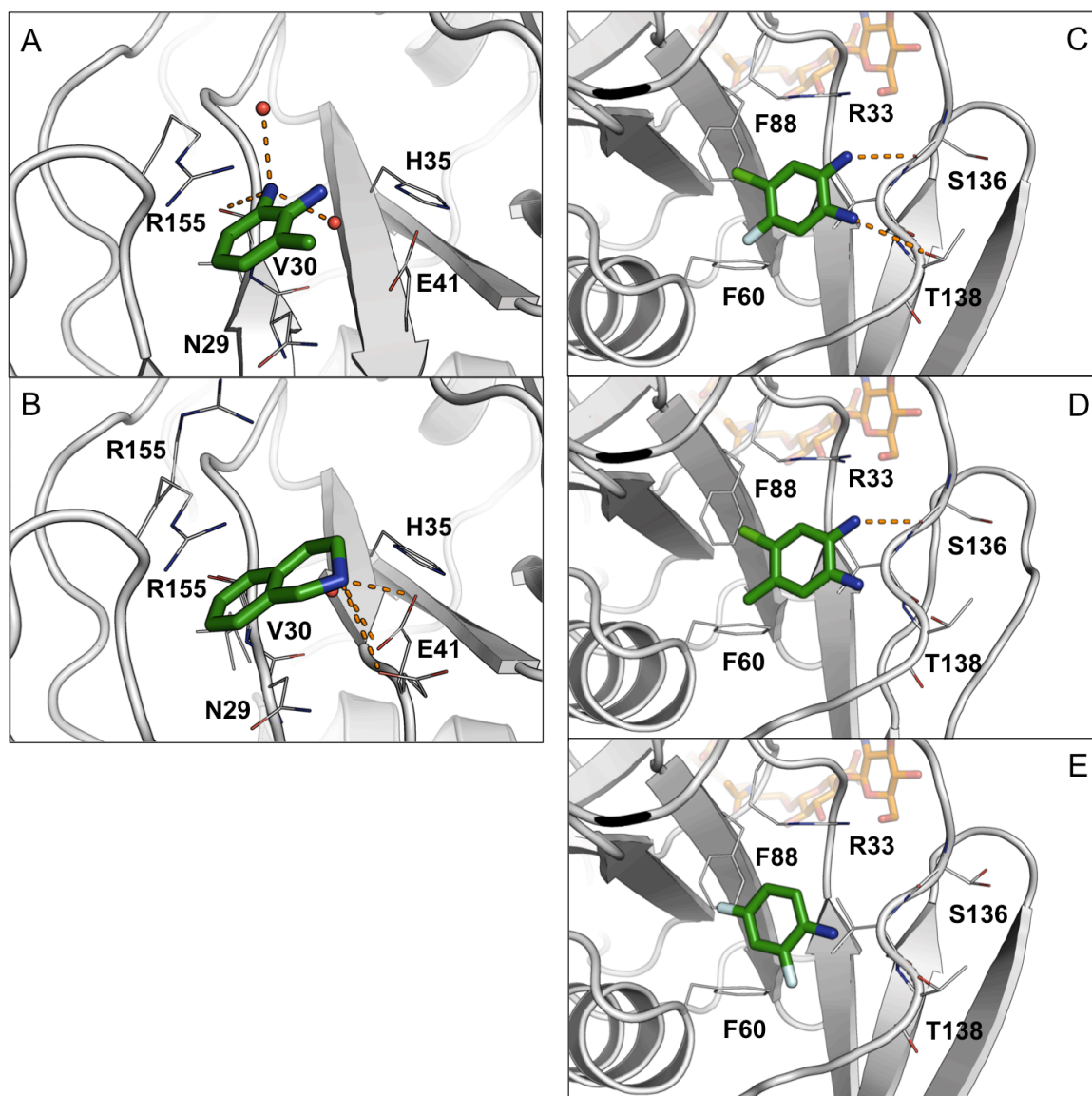
Figure 3-5. The ligand binding environment from a series of crystal structures with bound small molecules that occupy the two binding sites in *m*HABD. Hydrogen bonds to ligands are identified with dashed lines. Bound water molecules are shown as small red spheres. A) Compound S1 (PDB-ID 4MRE). B) Compound S2 (THIQ; PDB-ID 4MRF). Side chains of Arg-155 and Glu-41 are found in two discreet conformations of roughly ½ occupancy. Both molecule S1 and S2 share the Site 1 that permits stacking of the aromatic ring between Arg-155 and Asn-29. Compound S3 (PDB-ID 4MRG) (C), Compound S4 (E), and Compound S5 (F) bind to the same location of Site 2.

Three molecules initially identified by SPR bind on the backside of the HABD far from the HA binding groove (Figure 3-3 B, Site 2). Binding occurs without any significant changes to the *apo m*HABD conformation in a shallow site formed adjacent to Arg33, Phe38, Phe60, Asn125, Ser136 and Val137, adjacent to the β7- β8 loop. These molecules are amphiphilic and structurally similar with three or four substitutions on a benzene ring. Compound S3 and S4 are only one atom different and almost superimpose upon each other in the crystal structure. They interact with a relatively hydrophobic core of the pocket via a chlorine, fluorine or a methyl group and hydrogen bond with a carbonyl oxygen of the more solvent accessible Ser136 (Figure 3-5 C–D). S5 binds to the same location with only a slight reorientation of the benzene ring but is unable to make the hydrogen bond (Figure 3-5 E). The additional H-bond that compound S3 has gained may help to explain the tighter binding observed by the crystal structure.

3.4 Discussion

To discover possible binding sites for small molecules that might interfere with HA binding to CD44 we conducted fragment screening with label-free biophysical assays utilizing DSF and SPR. Binding of several compounds was then validated through co-crystallization. This work affords comparison of the two primary screening approaches.

Fragment screening methods

The simplicity, low cost and potential effectiveness has made DSF screening widely applicable for fragment hit identification [59, 83]. The screen was completed in three days in duplicate using only about 5 mg of proteins. Our definition of hit threshold gives a reasonable number of hits to work with – 16 primary hits. Half of the primary hits were confirmed by SPR, including five compounds that were eventually co-crystallized.

DSF seems to confirm fewer hits than as seen for other screening techniques (e.g. NMR or SPR) and increases the rate of false negatives [83-86]. The parallel screen of the same compounds using a Biacore assay identified 21 hits, including 17 unique compounds that were missed with the DSF screen because the T_m shift was negative or below the threshold cutoff ($\geq +2°C$) (Table 3-2). Eleven of these compounds induce no thermal shift but binding was validated with reasonable dose response curves on SPR. Similarly, eight DSF primary hits were missed in the primary SPR hit selection but confirmed by a dose response validation. This is because the response magnitude from a single concentration screen does not necessarily tell binding specificity. More unique hits may be recovered by screening at different concentrations (or dose response) or a parallel screen using different techniques.

There is little apparent relationship between the magnitude of the T_m shift and the K_D measured by SPR. The compound that produced the greatest stabilizing shift (D12; T_m = +4 °C) could not be confirmed by SPR, while many fragment compounds that induce no T_m shift (e.g. S2), or even wildly destabilizing T_m shifts (e.g., S14, S19, S20) can be. One standard for accuracy may be the structural confirmation of screening hits, although it is

well-established that co-crystallization can fail even in the case of very tight binding compounds. In this case, however, nine weakly binding unique hits can be co-crystallized. Eight of these compounds showed thermal shifts of $\geq 2°C$ and passed the secondary confirmation by SPR. The only exception is compound S2, which shows a zero degree thermal shift. It proved, however, to be an important compound. Consensus evidence of binding from multiple biophysical techniques lends reliability and credibility to the hits and provides justification for effort in crystallography.

Lead compound evaluation

We have been able to confirm binding of five fragments to *human* HABD by co-crystallization (Figure 3-5), with unambiguous electron density revealing fragment binding. The three compounds that gave a positive response in both the DSF and SPR assays (D1, D2, and D5) bind to the back side of the HABD far from the HA binding site (Figure 3-4 A–C, Site 1). Site 1 is partially formed by N- and C-terminal residues of the truncated *h*HABD construct that might adopt quite different conformations from the wild-type CD44. Two fragments that markedly destabilize *h*HABD bind in Site 2 (Figure 3-4 D–E). Either Site 2 or the 107-109 loop is in direct contact with the HA binding groove. Moreover, binding at this site is further stabilized by the presence of a neighboring *h*HABD molecule in crystal packing. Given the complete absence of any evidence of specific binding of D17 and D18 by SPR, these molecules may bind only as a crystallographic artifact, and not in solution. Both Sites 1 and 2 of *h*HABD are likely functionally unimportant with respect to the inhibition of HA binding, and neither of

them possess druggable characteristics.

Site 1 of *m*HABD is one we hoped to target with fragment screening. Prior to screening, a computational analysis [87] was applied to the *apo h*HABD structure (PDB-ID 1UUH) [31] to detect and rank possible small molecule binding sites along the HA binding groove. The method identified a cavity between Asn29 and Arg155 (subsequently identified as Site 1) based on surface curvature, surface area and residue hydrophobicity. In *murine* mutation studies, an R155A mutation had previously been shown to reduce the affinity of *m*HABD for HA, but not to abolish binding [32, 50]. Greater importance has been attributed to the conformation of Arg45, because the R45A mutant has much smaller affinity for HA [32]. A shift in Arg45 has been proposed as the principle difference between the open and the closed state of CD44 induced upon HA binding [32]. Both solution structural studies [32, 53] and molecular dynamics calculations [49] have supported this hypothesis. Small molecules that interfere with the shift of Arg45 may interfere with p-HA binding. Residues around Site 1 are in direct contact with the loop of Arg45 that may affect the effectiveness of HA binding.

Both S1 and S2 bind to Site 1 and induce alternate conformations of Arg155, and Asn29 in order to create π-stacking above and below the aromatic ring of the small molecule (Figure 3-6). More importantly, upon binding to both fragments, the Arg45 loop in the two structures resembles the open state of CD44 that has less surface contact with HA, suggesting both molecules may interfere with the shift of Arg45.
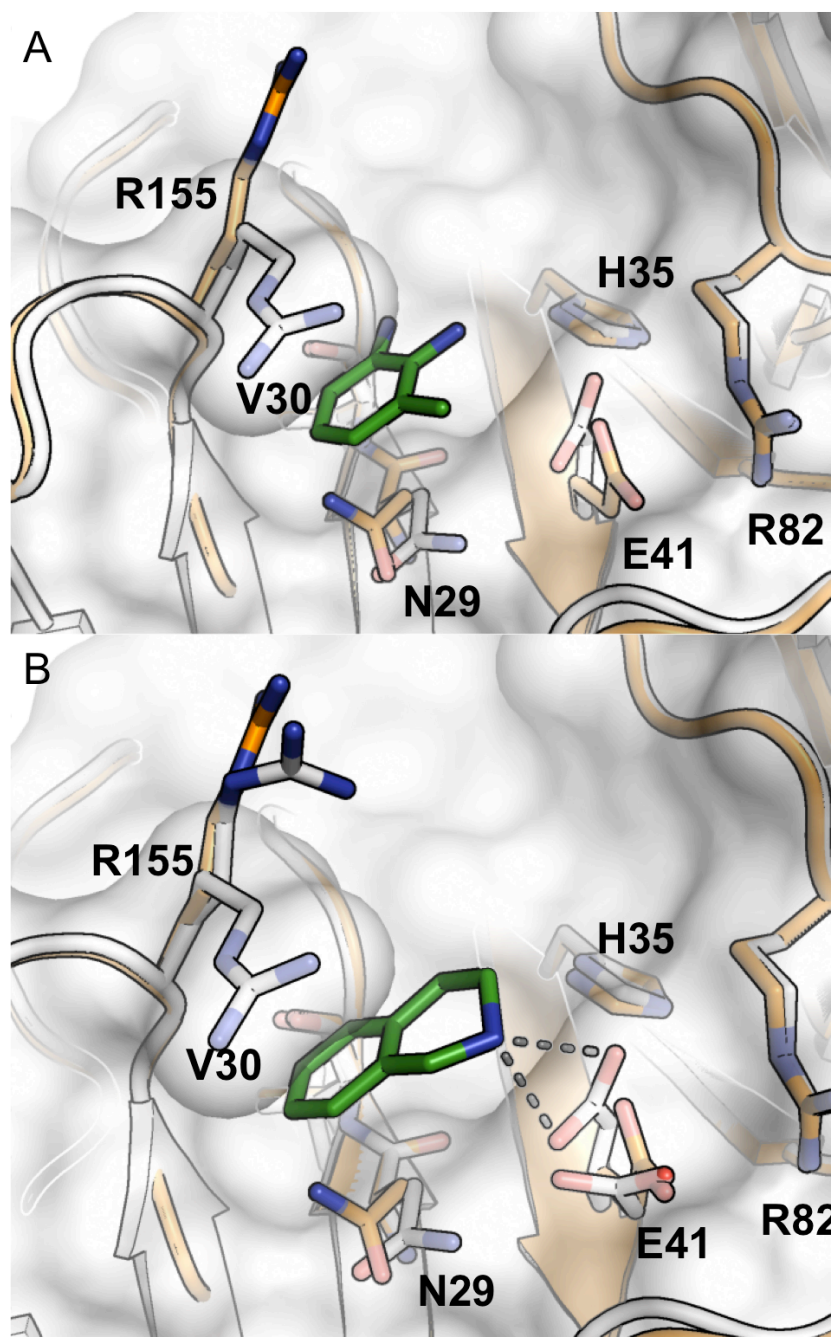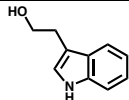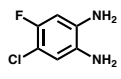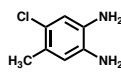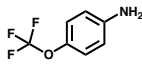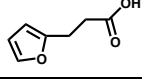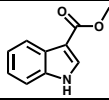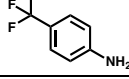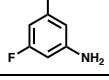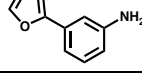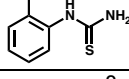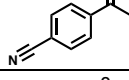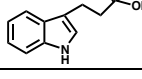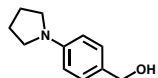
Figure 3-6. (A) Site 1 of Compound S1 (green; 4MRE) is shown in relation to the apo structure (orange, 2JCP). (B) Conformational differences between *apo* and S2-bound states in the crystal structure (4MRF).

Modeling could not have predicted the conformational changes that lead to the creation of the binding site for S1 and S2. In this case, fragment screening accomplished exactly what was hoped – it led to the empirical identification of conformational changes that would not have otherwise been known. The binding affinity of S1 (K_D = 6.9 mM) and S2 (K_D = 8.5 mM) reflect weak binding that its not significantly better than that observed with other fragments. The 1,2,3,4-tetrahydroisoquinoline (THIQ) that is compound S2 does possess a barely detectable ability to interfere with the binding of *h*HABD to immobilized HA assay (IC_{50} = 27 mM). Neither of these molecules is a good inhibitor, but any larger molecule that binds at this site, or that has the capacity to stabilize a protein conformation that is consistent with this binding mode, can potentially act as an antagonist that interferes with CD44 binding to polymeric HA or the HA-coated surface. These two structures provide direct insights into the rational design of more potent and larger antagonists described in Chapter 4.
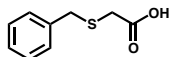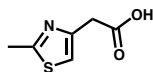
It should be noted that HABD binding alone is not sufficient to inhibit binding to immobilized HA. Compound S3 and S4 bind *h*HABD with higher affinity than either compound S1 or S2 (Table 3-2). Nevertheless, these two compounds have no effect on HABD binding to immobilized HA (Table 3-4). Co-crystallography shows that compound S3, S4 and S5 bind on the back side of the *m*HABD (Site 2) that is one of the more rigid portions of the structure. The opportunistic binding by peptides has also been observed in the same location with *h*HABD (Chapter 2). We think the site may be used by CD44 to assemble with other proteins.

Table 3-1. Maybridge Ro3 Fragment Hits Selected by DSF Screening

Structure	Maybridge	ΔT_m (°C)	MW	K_D (mM) for hHABD	LE	Co-crystal	Entry
	BTB07447	2	122.2	8.5	0.31	YES	D1
	BTB14321	2	161.2	15.0	0.21	YES	D2
	BTB03690	2	160.6	5.0	0.26	YES	D3
	TL00770	2	156.6	5.4	0.31	YES	D4
	CC02701	2	125.1	5.0	0.35	YES	D5
	SB00732	2	137.2	15.5	0.27	NO	D6
	TL00398	2	177.1	9.5	0.23	NO	D7
	AC33010	2	140.1	5.0	0.31	NO	D8
	AC30928	2	175.2	N.A.	N.A.	NO	D9
	SB00719	3	161.1	N.A.	N.A.	NO	D10
	SB01840	3	129.1	N.A.	N.A.	NO	D11
	SB01855	4	205.0	N.A.	N.A.	NO	D12
	SPB06629	2	182.2	N.A.	N.A.	NO	D13
	TL00209	2	145.2	N.A.	N.A.	NO	D14
	BTB13431	2	189.2	N.A.	N.A.	NO	D15

	CC30109	2	177.3	N.A.	N.A.	NO	D16

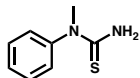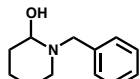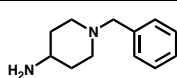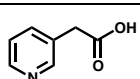Destabilizing fragment hits that are false-positives by crystallography
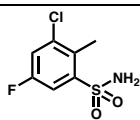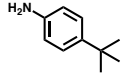
	CD04055	−24	182.2	N.A.	N.A.	YES	D17
	CD12020	−10.5	157.2	N.A.	N.A.	YES	D18

K_D values were obtained by SPR, measuring dose-dependent binding to immobilized human HABD.
N.A. = Not available

Table 3-2. Confirmed Maybridge Ro3 Compounds Selected by SPR-based Screening

Structure	Maybridge	ΔT_m (°C)	MW	K_D (mM) for hHABD	LE	Co-crystal	Entry
	BTB07447	2	122.2	8.5	0.31	YES	S1
	BTB09216	0	133.2	6.9	0.29	YES	S2
	BTB03690	2	160.6	5.0	0.26	YES	S3
	TL00770	2	156.6	5.4	0.31	YES	S4
	SB00555	2	129.1	7.0	0.35	YES	S5
	SP00460	1	200.3	2.8	0.25	NO	S6
	SEW03743	0	174.2	4.0	0.33	NO	S7
	SEW04169	1	187.2	1.6	0.27	NO	S8
	BTB06790	1	166.1	2.8	0.32	NO	S9
	AC10403	1	115.2	11.0	0.33	NO	S10
	AC33015	1	126.1	2.0	0.41	NO	S11
	CC54613	0	142.2	9.0	0.31	NO	S12
	BTB10270	0	150.2	6.0	0.28	NO	S13
	TL00805	N.D.	133.2	3.4	0.34	NO	S14
	TL00917	0	155.6	3.8	0.33	NO	S15

Structure	Compound		KD				
	TL00838	0	166.2	3.0	0.35	NO	S16
	AC13425	0	191.3	7.5	0.21	NO	S17
	SB01177	0	190.3	5.6	0.22	NO	S18
	AC13196	-24	137.1	4.0	0.33	NO	S19
	TL01004	N.D.	223.7	9.0	0.21	NO	S20
	SB01038	1	149.2	7.0	0.27	NO	S21
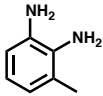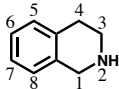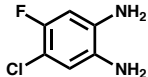
K_D values were obtained by SPR, measuring dose-dependent binding to immobilized human HABD.

Thermal transitions of S14 and S20 were not observed due to the early unfolding, therefore the Tm changes were not detectable (N.D.).

Table 3-3. Confirmed Infarmatik Ro3 Fragment Hits Selected by SPR-based Fragment Screening

Reference	Structure	MW	ΔT_m (°C)	K_D (mM) for *h*HABD	Co-crystal
EL-0254		275.2	1.5	> 20	No
EL-0105		215.3	1	4.2	No
EL-0099		238.7	0.5	0.06 – 0.75	No
EL-0022		255.3	0	> 20	No
EL-0177		287.3	0	1.73	No
EL-0059		166.2	−0.5	0.04 – 2.68	No
EL-0264		273.3	−1	0.07 – 2.89	No
EL-0017		295.2	−6	6.4	No
EL-0251		304.2	−11.5	0.004	No
EL-0207		272.3	−23.5	0.03 – 6.75	No

Table 3-4. Small Molecule CD44 HABD Binding and p-HA-binding Interference

Structure[1]	No	MW	LE[2]	K_D (mM)		IC_{50} (mM)[3]
				*h*HABD	*m*HABD	*h*HABD
	S1	122.2	0.31	8.5 ± 1.2	6.4 ± 1.0	No inhibition
	S2	133.2	0.29	6.9 ± 1.5	11.2 ± 4.5	27.0
	S3	156.6	0.31	5.4 ± 0.6	12.6 ± 9.4	No inhibition

[1]The numbering of atom position is provided for compound **2** to help clarify the identification of analogs in the discussion. By convention, the endocyclic *N* of the isoquinoline is atom 2.
[2]Ligand efficiency, (– RT lnK_D)/(number of non-hydrogen atoms), was calculated based on the affinity of ligands to *h*HABD in *kcal mol⁻¹*.
[3]K_D and IC_{50} values were measured similarly as done with oligosaccharides by SPR. Most competition experiments were conducted using flowing *h*HABD concentration of 23 μM.

CHAPTER 4

MORE POTENT INHIBITORS OF HA BINDING

4.1 Background

In the previous section, we described a small molecule binding site adjacent to the HA binding groove that could be an attractive target for the binding of more potent HA inhibitors. The key to enhancing the activity of these molecules will be to be able expand the footprint of binding outward so that they interfere and compete directly to prohibit the association with polymeric HA. In this chapter, attempts to identify more potent molecules than the original fragment leads will be presented. We will describe the integration of features from compounds **1** and **2** (Table 6) that led us to discover the scaffold of 5- and 8-amino-tetrahydroisoquinolines (5- and *8-amino*-THIQ; compound **3** and **7**; Scheme 4-1) that induce the same changes in the conformation of binding site residues noted with a bound fragment (Sect 3.3.5). Structural studies of commercially available analogs of **2** helped reveal molecular features that, when added in combination with the THIQ pharmacophore, led to the identification of molecules that compete for HA binding to CD44 in the SPR competition assay.

Following an analysis of co-crystal structures with weak HA inhibitors, additional analogs were procured and evaluated to see how and how well they bind. During this early part of a lead optimization campaign, compounds were optimized to increase molecular affinity toward *h*HABD as evaluated with dissociation constants (K_D) as determined with the *in vitro* Biacore (SPR) binding assays (See Section 2.3.9), or the inhibition of CD44 associated with an HA-coated surface (e.g., IC_{50} in HA competition) using the immobilized-HA competition SPR assay (Section 2.3.7). Eventually, however,

cell-based methods will be required.

Select compounds have been co-crystallized, and have been used to facilitate iterative compound optimization using structure-driven medicinal chemistry. The best of the compounds we described have CD44-HA antagonist activity that is modest (1 mM), but nearly comparable to that possessed by very short o-HA fragments known to effect cell migration [6]. A boost in potency of these compounds to that of HA_8 (~20 μM) should be sufficient to impart biological activity [31, 34].
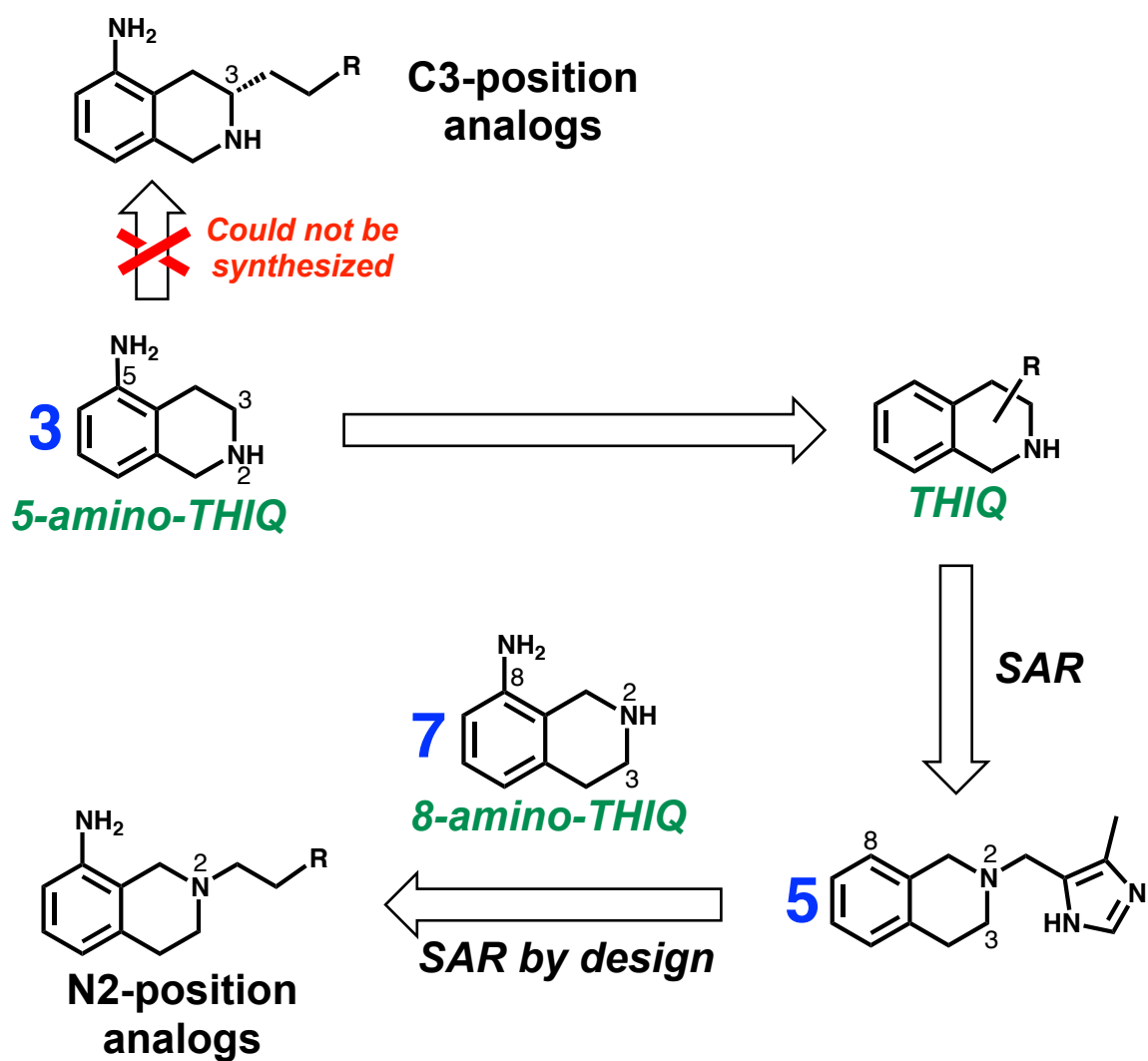
Scheme 4-1. Flow Chart of Focused Library Design and THIQ Scaffold Evolution.

Attempts to expand the molecular footprint for binding further into the HA binding groove have met with only limited success, but some of the complexes do point the way toward the design of more potent HA inhibitors. Possible designs for more potent inhibitors will be discussed. Such exploration of structural activity relationships will help to define the design criteria needed for effective binding and future drug optimization. We expect to refine the design in several iterations, requiring the preparation a series of analogs. Design criteria will evolve as more structures and SAR emerges from the first rounds of this optimization process.

4.2 Materials and Methods

4.2.1 Materials

Analogs based on *5-amino-THIQ* (compound **3**) obtained from various sources, including Cambridge (> 90% purity; San Diego, CA), Enamine (> 90% purity; Kiev, Ukraine), Sigma Aldrich (≥ 98% purity; St. Louis, MO), ASW (≥ 99% purity, New Brunswick, NJ) and NCI Diversity Set Library (> 90% purity; Frederick, MD), as noted in Table 4-2. Aliquots of improved analogs based on *8-amino-THIQ* (compound **7**) were custom-made by Glycsyn Inc. (Suzhou, China; ≥ 95% by HPLC, NMR and LC-MS), Enamine (Kiev, Ukraine; ≥ 95% by HPLC, NMR and LC-MS), and in-house at ITDD and Dept. of Med Chem (UMN Med Chem, MN; ≥ 95% by NMR) as noted. The purity

of purchased compounds was demonstrated by commercial vendors.

The synthesis of analogs based on *8-amino-THIQ*, as well as of saccharide-like moieties of **14**, **15a** and **15b**, was completed Lev Lis in the laboratory of Vadim Gurvich (Institute for Therapeutic Discovery & Development). The purity for custom-made compounds synthesized in the ITDD was confirmed by HPLC, NMR and LC-MS.

4.2.2 General methods

Computational modeling

There are hundreds of commercially available analogs of **2**. To prioritize molecules to be studied, computational program RACHEL (Tripos, St. Louis, MO) was used to survey simple chemical groups on the THIQ suitable for making interactions with nearby binding site residues. Initially, we utilized the co-crystal structure of compound **3** (PDB-ID 4MRG) as an anchoring scaffold to search for chemical groups suitable to be installed to extend toward a Cys81 in the HA site. The program generated thousands of analogs of **3** *in silico* using a 3D NCI structural database, predicted how they may bind and scored individual molecules based on the molecular contacts. Extensions modeled with reasonable binding conformation and/or scoring were used to select those structurally similar molecules that could be purchased from various vendors. These chosen analogs (~ 200) were collected into SDF files for further docking assessment.

A docking program SURFLEX (Tripos, St. Louis, MO) was used to further narrow the list of compounds to pursue. The co-crystal structure of *m*HABD with compound **3** (PDB-ID 4MRG) was used to define the receptor and the location of the active site. The receptor and ligand-containing files were prepared as recommended in the manual (e.g. ligand extraction, energy minimization), and the Tripos force field was used to generate the docking result. Molecular contacts and ligand geometry of the individual docking pose were closely inspected to result in a list of 45 compounds that were eventually purchased. This includes many compounds that have side chains installed to the endocyclic nitrogen N2 as partly shown in Table 4-2 (See Table 4-1 for THIQ ring nomenclature).

Analog design

We have partnered with commercial suppliers Enamine and Glycsyn, who provided a series of 2-position analogs of *8-amino*-THIQ (compound **7**; Scheme 4-1) upon request. Proposed analogs were first evaluated with restrained docking of SURFLEX as described above by fixing the THIQ scaffold in the original position as observed in the co-crystal structure of **5** (PDB ID 4NP2). Analogs binding with reasonable conformation and/or scoring were selected for synthesis, and the CRO would evaluate synthetic feasibility. Compounds ultimately received and evaluated are summarized in Table 4-4.

In vitro binding affinities against human and murine HABD were determined as described in Chapter 2 using the immobilized HABD SPR assay (Section 2.2.9). HA competition studies to determine IC$_{50}$s were conducted using the immobilized HA SPR assay with *h*HABD as described in Section 2.2.7. The same assay was implemented to determine if THIQ compounds inhibit binding by other soluble HA binding proteins (e.g., LYVE-1 and TSG-6). All crystals for complexes with murine HABD were prepared as for fragments (described in Chapter 3). Diffraction data were collected as described in chapter 3, using synchrotron radiation. Structures were determined by molecular replacement using the search model 4MRH and refined as described for fragment complexes in Chapter 3. Figures were made with PyMol (Schrodinger). Structures have been prepared for deposition in the shared space of the Finzel Laboratory at the University of Minnesota Supercomputing Institute.

4.3 Results
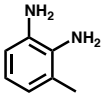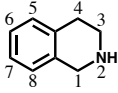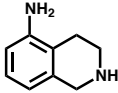
During the course of this lead optimization, 62 compounds were investigated for CD44 binding. High-resolution co-crystal structures of 34 ligands were determined in complex with *m*HABD. A catalog of this information is included as Appendices A and B to this thesis.

4.3.1 The 5-amino-THIQ Pharmacophore

Preliminary fragment screening resulted in identification of a binding site for flat small molecules **1** and **2** beneath the sidechain of Arg155 (Figure 4-1 A). A direct comparison of the structures of the protein complexes with these two initial hits led us to evaluate compound **3**, which combines the tetrahydroisoquinoline of **2** (Figure 4-1 B) with an exocyclic aryl amine placed on carbon 5 to emulate compound **1** (Figure 4-1 A). A co-crystal structure at 1.7 Å resolution has been determined to verify that this compound binds with an H-bond to Val30 via the exocyclic amine, an H-bond to Glu41 via the endocyclic amine, and planar stacking between Arg155 and Asn29 as anticipated from the fragment precursors (Figure 4-1 C). This compound displayed an improvement in affinity for hHABD (K_D = 0.9 mM) in relation to compound **2** and a similar improvement is seen in the K_D of binding with mHABD (Table 4-1); the addition of the exo-cyclic amine to C5 improves the binding affinity of the THIQ scaffold by 7-fold. Unlike in the initial complex with compound **2**, in this complex, the binding site appears to be fully occupied by compound **3**; there is no evidence of the partial occupancy of the *apo* protein conformation. Perhaps more importantly, compound **3** also shows an improved capacity to compete with polymeric HA for binding at CD44 (IC_{50} = 6.7 mM). This is a 4-fold improvement over the minimal activity of the original THIQ fragment (compound **2**), affirming the hypothesis that the binding site may be functionally relevant in the

inhibition of CD44-hyaluronan activities. These results establish the *5-amino*-THIQ

(compound **3**) as a preferred scaffold for additional lead optimization.

Table 4-1. THIQ Binding and p-HA-binding Interference

Structure[1]	No	LE[2]	K_D (mM)		IC_{50} (mM)[3]	Structure
			*h*HABD	*m*HABD	*h*HABD	*(PDB ID)*
	1	0.31	8.5 ± 1.2	6.4 ± 1.0	> 50	4MRE
	2	0.29	6.9 ± 1.5	11.2 ± 4.5	27.0	4MRF
	3	0.40	0.9 ± 0.2	1.1 ± 0.2	6.7	4MRG

[1]The numbering of atom position is provided for compound **2** to help clarify the identification of analogs in the discussion. By convention, the endocyclic *N* of the isoquinoline is atom 2.

[2]Ligand efficiency, $(-RT \ln K_D)$/(number of non-hydrogen atoms), was calculated based on the affinity of ligands to *h*HABD in *kcal mol^{-1}*.

[3]K_D and IC_{50} values were measured similarly as done with oligosaccharides by SPR. Most competition experiments were conducted using flowing *h*HABD concentration of 23 μM.
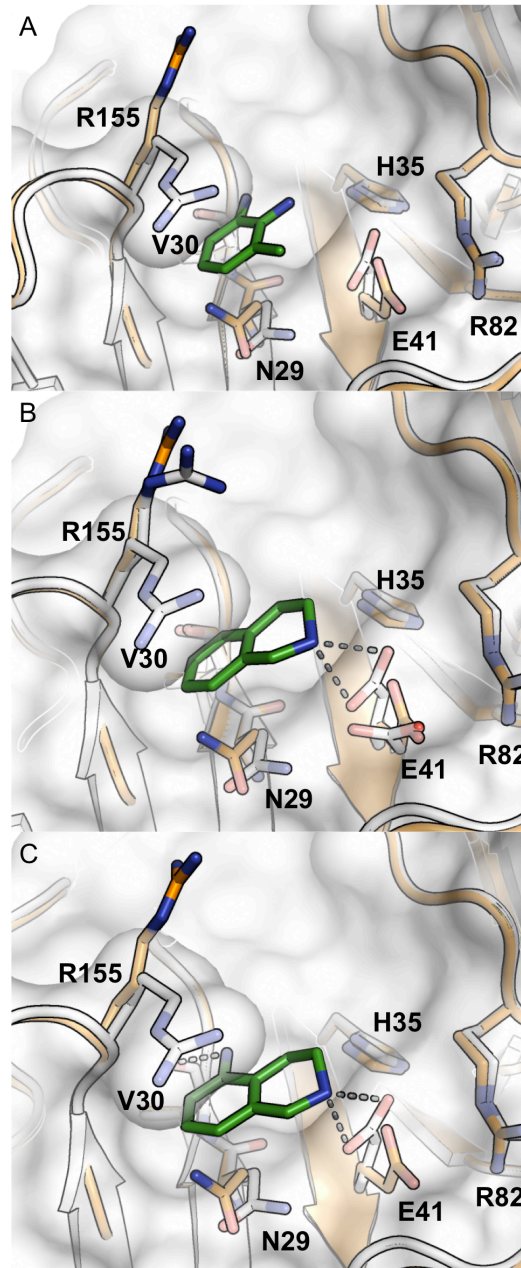
Figure 4-1. (A) Binding site of Compound **1** (white; 4MRE) is shown in relation to the *apo* structure (orange, 2JCP). (B) Conformational differences between *apo* and **2**-bound states in the crystal structure (white, 4MRF). (C) Conformational differences between *apo* and **3**-bound states in the crystal structure (white, 4MRG).
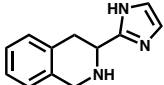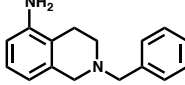
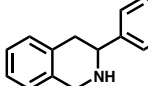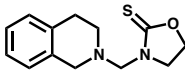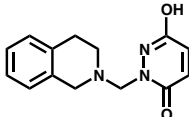## 4.3.2 Substitution of the THIQ scaffold
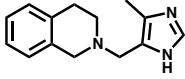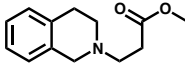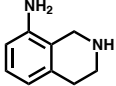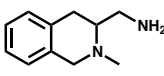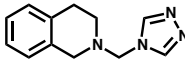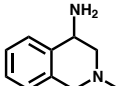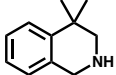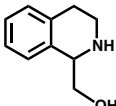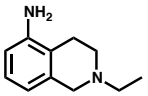
We hypothesized that analogs of compound **3** that bind in a similar fashion, but that extend further into the HA binding groove to occupy space taken by bound carbohydrate residues will be better inhibitors. Molecules are envisioned that can preserve all the important contact features of *amino*-THIQ binding, but also make additional specific protein-ligand interactions. Modifications from the THIQ N2 or C3 would extend the molecules toward this space most directly (Scheme 4-1). Analogs with the exocyclic amine group installed on C5 would have been preferred, but few molecules that include both the quinoline nitrogen and the exocyclic amine are commercially available. We therefore sought to explore the SAR with simpler commercially available THIQ analogs first, believing that the exocyclic amine could be installed onto preferred analogs later to enhance potency.
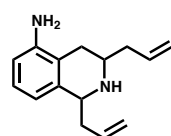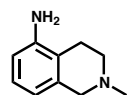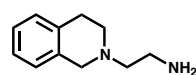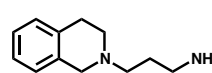
Commercially available analogs containing a THIQ that could be modeled in the active site preserving parent heterocycle orientation were purchased and evaluated for binding. In all, forty-five commercially available analogs containing a THIQ were acquired (Appendix A). Fourteen of these compounds have been co-crystallized with *m*HABD (Table 4-2). This set includes several examples with R groups extending from N2 (**5**, **6**, **7e**, **7l**, **7m**), and from C3 (**7a**, **7b**, **7k**), but also some with small substituents on C1 or C4. Only three (**7**, **7i**, **7j**) include an exocyclic amine; these generally have improved affinity as expected.

Crystal structures of these analogs reveal that the aromatic ring stacks between Arg155

and Asn29 as hoped. As seen for compound **3**, a half-chair conformation is needed for the saturated heterocycle of the THIQ if the hydrogen on the endocyclic nitrogen is to point downward to make a hydrogen bond with Glu41 (Figure 4-1 C). Given this binding mode, any substitution from N2 or C3 of the THIQ ring must be installed with stereochemistry consistent with equatorial extension from the ring. In all the complexes in which full occupancy of the ligand is observed in crystal structures, only the equatorial conformation is observed (Table 4-2).

Table 4-2. Co-crystallized Analogs of 1,2,3,4-tetrahydroisoquinoline (2)

Structure	No	K_D (mM) for hHABD	LE for hHABD	ΔT_m (°C)	Ligand Occupancy
	5	4.61	0.19	0.5	Full
	6	4.42	0.20	0	Full
	7	1.26	0.36		Full
	7a	3.07	0.23	−1	Partial
	7b	N.D.	N.D.	−1.5	Full
	7e	2.80	0.22	−6	Partial
	7f	5.00	0.26		Full
	7g	6.24	0.25	−1	Full
	7h	3.29	0.28		Full
	7i	1.31	0.30	−1	Full
	7j	13.2	0.14	0	Partial
	7k	> 20	n.a.	−1	Partial
	7l	0.85	0.25		Partial
	7m	> 20	n.a.		Partial

	Compound	K_D			
(structure)	**7n**	3.01	0.20	−1	Partial
(structure)	**7o**	4.90	0.26	−0.5	Full
(structure)	**7p**	> 20	n.a.	−0.5	Partial
(structure)	**7q**	> 20	n.a.		Full
(structure)	**7r**	> 20	n.a.		Full

K_D values were obtained by SPR, measuring dose-dependent binding to immobilized human HABD.

4.3.3 The origin of the 8-amino-THIQ scaffold

The nitrogen 2-linked imidazole of **5** and ester of **6** bind HABD with extra H-bond interactions and leave the functional groups pointing directly toward HA subsites (Figure 4-2). Both show unambiguous ligand electron density in co-crystal structures suggesting full binding site occupancy. Compound **5** binds with methyl-imidazole extending across the CD44 surface toward the HA binding groove and saccharide subsite 5 (Figure 4-2 A). A torsional preference conferred by the ortho methyl group brings about a desirable 90° twist of the imidazole from the plane of the THIQ, allowing for the protonated NH of the imidazole to donate an additional hydrogen bond to Glu41 below. The methyl group points upward to lie within 3.0 Å of the position of glucuronic acid atoms occupying the GlcUA5 sub-site in the HA_4-bound complex. Compound **6** binds with the alkyl ester extended toward sachharide subsite 4 where it makes an additional H-bond with the amide of Cys81 (Figure 4-2 B). The affinity of compounds **5** (K_D = 4.6 mM) and **6** (K_D = 4.4 mM) for *h*HABD is only marginally improved over that seen for compound **2** (K_D = 6.9 mM).

Surprisingly, a comparison of the crystal structure with compound **5** to the original complex with the parent *5-amino*-THIQ (compound **3**) reveals that the THIQ is flipped over in the binding site (Figure 4-2 C). Small conformational changes to Glu41, coupled with an inversion of the saturated ring pucker and a small rotation of the THIQ scaffold within the binding pocket, make it possible for the molecule to bind with the endocyclic

nitrogen in roughly the same position to H-bond with Glu41. The additional H-bond involving the imidazole is to the second carboxylate oxygen.

Compound **6** also binds with the ring flipped over, but the THIQ rotates more slightly in the plane of a sandwich between Arg155 and Asn29 (Figure 4-2 D). Consequently, the endocyclic NH is repositioned to lie roughly where C3 lies in the complex with the parent *5-amino*-THIQ. The carboxylate of Glu41 is rotated ~ 60° about chi-2 in this complex so that the hydrogen bond is preserved.

The flipped binding mode of compound **6** led us to consider the 1,2,3,4-tetrahydroisoquinoline (*8-amino*-THIQ; compound **7**) as an alternate pharmacophore. This compound was acquired from ASW MedChem and found to bind with a $K_D =$ 1.2 mM. Crystallographic data reveals that **7** binds with the THIQ flipped over as now expected, yet with heteroatoms coinciding closely to the same position as found in the complex with compound **3** (Figure 4-2 E). It is worth noting that THIQ alone (compound **2**) does not bind in this orientation. The amino group placed on the C8 position promotes the flipped binding mode and improves potency by donating a H-bond to the carbonyl oxygen of Val30 similar to that seen by the *5-amino* of compound **3**. The two THIQ orientations described above likely have comparable binding energies. The small difference in binding affinity of the compound **3** *5-amino*-THIQ (0.9 mM) and compound **7** *8-amino*-THIQ (1.2 mM) supports this assertion. In the course of study of analogs that differ only in alkyl extensions from either the 2 or 3-position of the THIQ core, we have observed binding of molecules in either orientation (not all data shown), confirming the ease with which compounds can easily flip if needed to optimize other interactions.
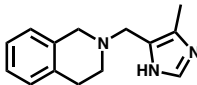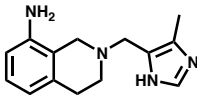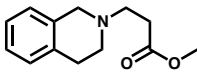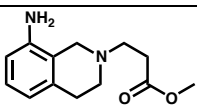
The *8-amino*-THIQ scaffold provides some distinct advantages in the design and preparation of next-generation CD44-HA inhibitors. Substitution from C3 of the initial *5-amino*-THIQ (compound **3**) requires the introduction of a chiral center that limits synthetic opportunities and commercial availability of the analogs. The endocyclic nitrogen of the *8-amino*-THIQ, on the other hand, is synthetically amenable for chemical alkylation and would serve as an equivalent scaffold branch-point with respect to compounds in the *5-amino*-THIQ series (Scheme 4-1).

As was the case with compound **3**, the potency of the *N*-alkylated THIQs can be predictably improved by adding the exocyclic amine to the 8 position (Table 4-3). Compounds **5a** and **6a** confirm a degree of additivity on compounds from the first round of SAR. Compound **5a** is one of the most potent THIQs identified ($K_D = 400$ μM), with a greater than 10-fold improvement over compound **5**. The affinity of **6a** improves 6-fold over the parent compound **6**.

There are still gaps that remain in the explored SAR, since it is now clear that available compounds do not sample all possibilities that might be of interest. We had hoped to explore the activity of a focused library of C3-substituted analogs of the *5-amino*-THIQs (Scheme 4-1) – compounds analogous to compounds **5b**, **9**, **10**, **11**, and **12** with the correct stereochemistry. A contract with Enamine to prepare several such analogs could not be executed, however, because of unsuccessful coupling reactions and difficulties with the separation of chiral products. We anticipate that these analogs would have bound with higher affinity than the counterparts in the *8-amino*-THIQ series because of the stronger H-bond of a secondary amine (N2). The failure to acquire these compounds

prompted a study of a broader collection of analogs of N2.

Table 4-3. First *8-amino*-THIQ Analogs

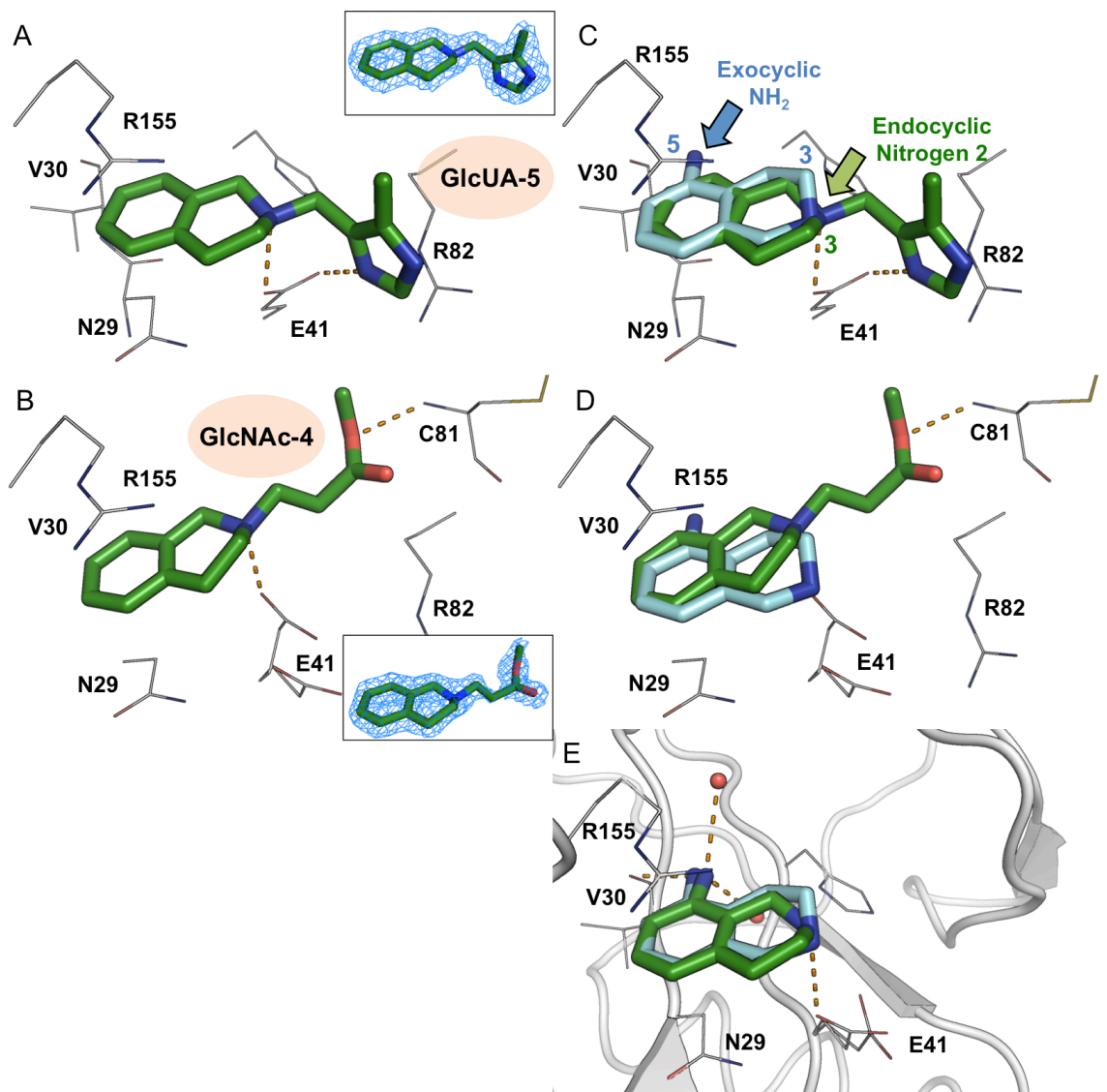| Structure	No	LE	K_D (mM)		IC_{50} (mM)	Source
			*h*HABD	*m*HABD	*h*HABD	
	5	0.19	4.6 ± 0.3	5.5 ± 0.6	14.1	Cambridge
	5a	0.26	0.4 ± 0.1	0.5 ± 0.1	3.5	Glycsyn
	6	0.20	4.4 ± 0.4	5.1± 0.5	14.9	Enamine
	6a	0.25	0.8 ± 0.8		3.9	Glycsyn

Figure 4-2. (A) Complex with Compound **5**. (B) Alignment of the crystal structure to that of *m*HABD co-complexed with *5-amino*-THIQ **3**. (C) Complex with Compound **6**. (D) Alignment of the crystal structure to that of *m*HABD complex with *5-amino*-THIQ **3**. (E) Alignment of crystal structures of *m*HABD co-complexed with *8-amino*-THIQ (Compound **7**, green) and *5-amino*-THIQ (Compound **3**, cyan). The $2F_O–F_C$ density for the ligand is shown in the inset.

4.3.4 Exploration of analogs that contribute additional H-bonds

We first focused on the empirical observation that an imidazole of compounds **5** or **5a** benefit from an additional H-bond to Glu41, and sought to exploit this as a design criterion. The THIQ of **5a** binds in the same orientation as the simpler compound **7** (Figure 4-3 A). The improved affinity of **5a** can be attributed to additional H-bonds made by the imidazole NH with the other Glu41 oxygen, a structural feature confirmed in complex co-crystal structures with both **5** and **5a** (Figure 4-3 B–C).
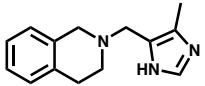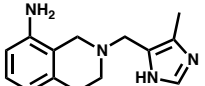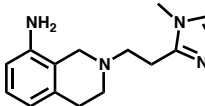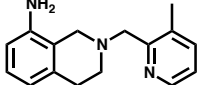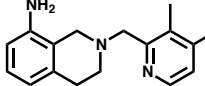
The limited flexibility of the methylene bridge to the imidazole heterocycle likely promotes this conformation. The presence of the ortho methyl substituent on the aryl ring also seemed important, as it likely forces the 90-degree twist of the aryl ring that allows the imidazole NH to be positioned to donate the H-bond to Glu41. The apparent success of these molecular elements prompted an evaluation of homologs that could make similar interactions. A methyl pyridine (compound **8**) embodies a specific design inspired by a desire to make a compound with similar interactions as **5a**. The binding affinity of compound **8** ($K_D = 0.5$ mM) is slightly weaker than that of the imidazole of **5a** ($K_D = 0.4$ mM). The pyridine nitrogen has a pKa less than 6; In theory, the nitrogen may not be as a good an H-bond donor as imidazole. The IC_{50} is much better, however ($IC_{50} = 0.95$), and almost comparable with that of HA tetrasaccharide. The crystal structure cannot readily explain this. This structure reveals the aromatic six-membered ring sits directly above the *apo* conformation of the Glu41 carboxylate to which it may donate hydrogen bonds (Figure 4-4 A). The *apo* and ligand-induced conformations of Asn29 are occupied in roughly equal proportions in this complex, suggesting that compound **8** occupies the site

with less than full occupancy in this structure (Figure 4-4 A, inset). It is likely that the boost of potency is dictated by shape complementary of the pyridine.

We hypothesized that an increase in the pKa of the pyridine nitrogen should strengthen the H-bond with Glu41 and further increase binding affnity. In theory, an electron-donating group installed on an aromatic ring would facilitate electron delocalization and influence basicity of the ring heteroatoms. To increase basicity of the pyridine nitrogen of **8**, compound **8a** was designed with an amino group as an electron donor placed on the *para* position in relation to the pyridine nitrogen (Table 4-4). This elevates the calculated pKa of the Glu41-associated nitrogen from 5.8 to 9.5. The crystal structure shows that compound **8a** aligns more closely to compound **5a** than to compound **8** and the co-complex structure reveals full occupancy of the compound (Figure 4-4 B). Unexpectedly, both the potency and the IC_{50} of compound **8a** are worse than the counterpart (compound **8**).

To probe whether a longer alkyl chain would allow the imidazole head to reach further from the THIQ pocket, we devised an imidazole homolog of **5a** with an additional carbon in the linker. The cocrystal structure with **5b** reveals that the ring conformation of THIQ is bulged upward, so that the compound loses the key H-bond with the endocyclic nitrogen (Figure 4-3 D). The imidazole head group remains associated with Glu41, and fails to relocate further toward the saccharide. The added configurational strain evident in this binding mode likely accounts for lower affinity (K_D = 1.3 mM) but not the IC_{50} (3.3 mM) compared to **5a** (Table 4-4).

132

Table 4-4. Small Molecule CD44 HABD Binding and p-HA-binding Interference

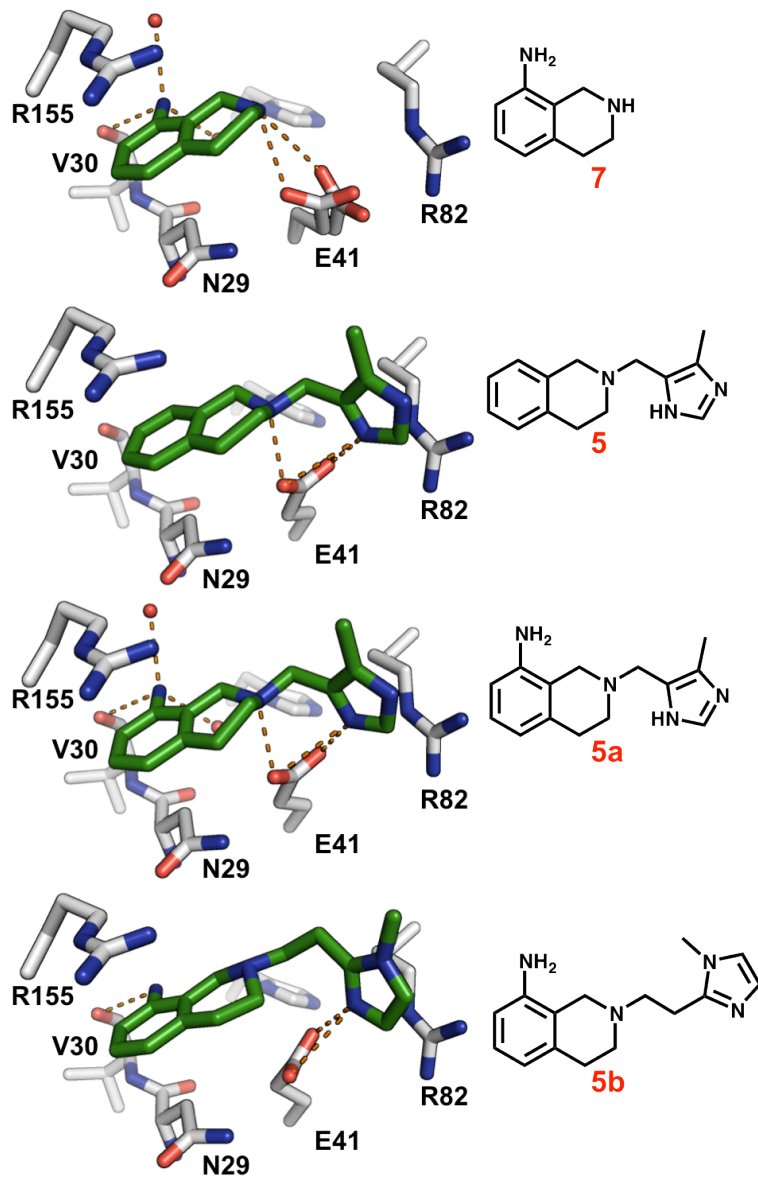| Structure	No	LE	K_D (mM)		IC_{50} (mM)	Source
			*h*HABD	*m*HABD	*h*HABD	
	5	0.19	4.6 ± 0.3	5.5 ± 0.6	14.1	Cambridge
	5a	0.26	0.4 ± 0.1	0.5 ± 0.1	3.5	Glycsyn
	5b	0.21	1.3 ± 0.6		3.3	Enamine
	8	0.24	0.5 ± 0.1		0.95	Glycsyn
	8a	0.20	1.3 ± 0.5		3.4	Glycsyn

Figure 4-3. Binding site of Compound **7**, **5, 5a** and **5b** (green) demonstrates fragment evolution (A) Arg155 and Asn29 rotate to create a sandwich above and below the ligand aromatic ring. Glu41 pivots away from Arg82 to accept an H-bond from the THIQ secondary amine, and to form an H-bond with the reoriented Asn29. An H-bond also is formed to Val30.
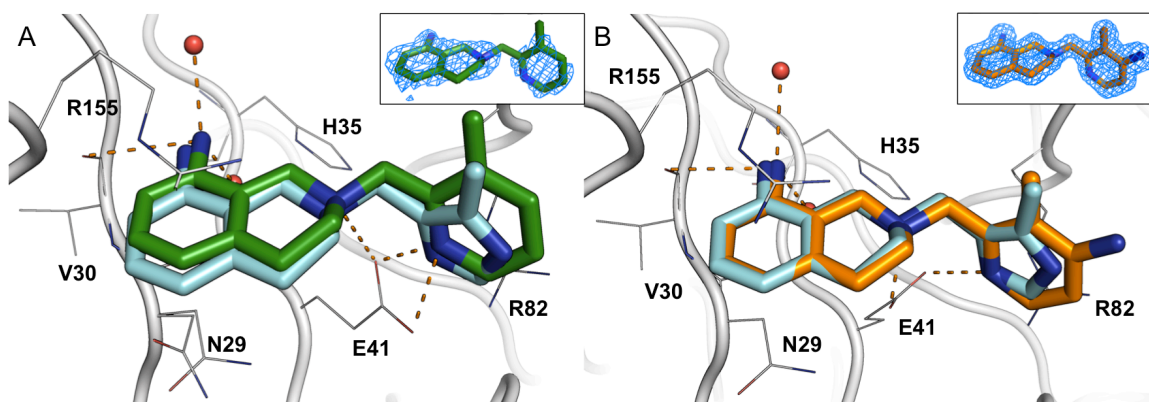
Figure 4-4. Binding site of Compound **8** (green) or 8a-bound complex (orange) is shown in relation to the Compound **5a** (cyan, 4NP3). The $2F_O$–F_C density for the ligand is shown in the inset.

4.3.5 N-Extended 8-amino-THIQ analogs

With the structure of **6**, it was revealed that at least a two carbon spacer is required for the flexible attachment of functional groups making contact with the HA groove. After additional docking analysis and modeling, the SAR was further explored by acquisition of a focused library of compounds that extend from the endocyclic N2 (Scheme 4-1). Ten additional compounds were ultimately acquired, co-crystallized and evaluated *in vitro* for competitive binding (as summarized in Table 4-5).

In all complexes, the *8-amino*-THIQ occupies the expected fixed position in each individual ligand complex. Most analogs extend outward toward the GlcUA5 subsite, while some curl up to lie in a large hydrophobic space beneath GlcUA3 and GlcNAc4 (compound **6a** and **9**). Compounds were assessed for their ability to interfere with the binding of CD44 HABD using the immobilized HA SPR assay. Specific features of representative co-complex structures are highlighted in the discussion to follow.

Ether and ester find a pocket beneath GlcNAc4

The structures with **6a** and **9** reveal that compounds curl up to lie in a large hydrophobic space beneath GlcUA3 and GlcNAc4. These two compounds add no specific new hydrogen bond interactions of note, but this space has been identified as a possible small molecule binding site along the HA binding groove (see Chapter 3,

Discussion) and has potential accommodate structural elements to increase inhibitor potency and specificity. A mixture of hydrophobic and positive charged surface of this large pocket likely attracts hydrogen bond acceptors such as ether and ester of compound **6a** and **9** to fill the space (Figure 4-5). A 90-degree bend can be affected using a combination of favorably staggered torsional rotations, but in the opposite direction of the imidazole of **5**. With the addition of the exocyclic amine to the 8-position of compound **6** to make **6a**, a repositioning of the ester occurs (Figure 4-5 A). The ether-containing arm of **9** almost superimposes onto **6a** (Figure 4-5 B), and has comparable binding affinity.
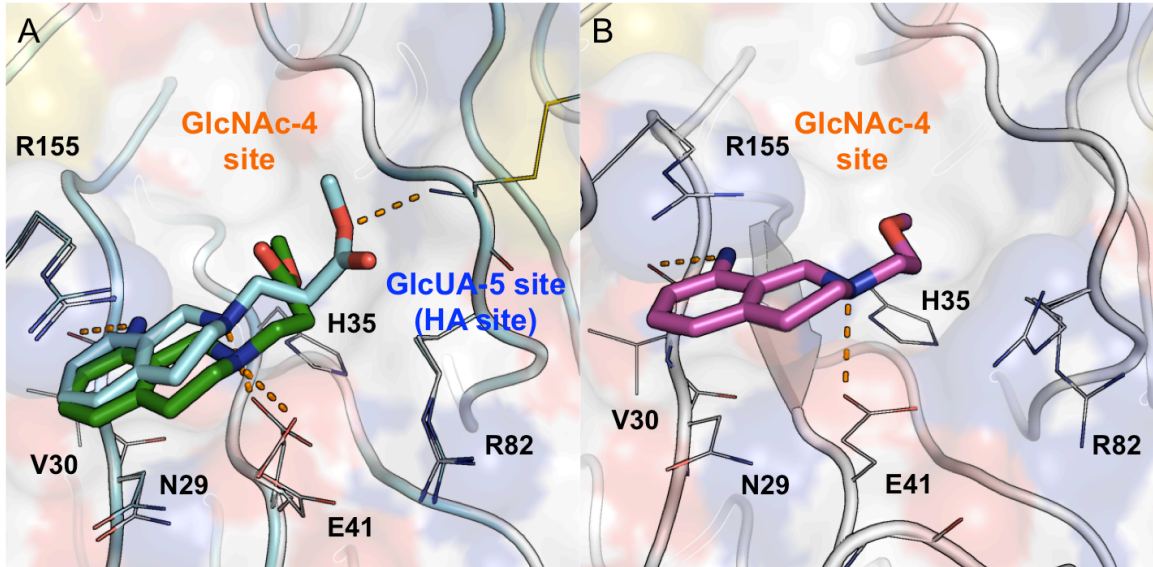
Figure 4-5. Binding site of Compound **6a** (A, green) or **9**-bound complex (B, magenta) is shown in relation to the Compound **6** (A, cyan).
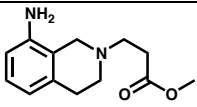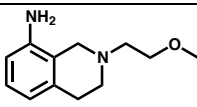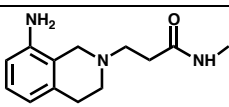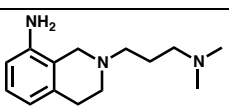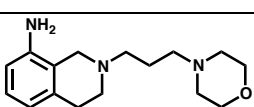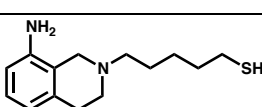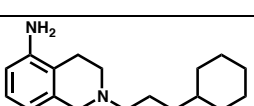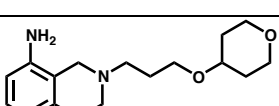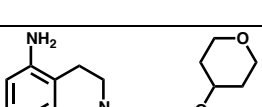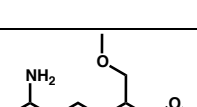
A number of longer extensions from N2 reach further toward the saccharide 5 subsite to lay over the position occupied by GlcUA5 in the HA_4 complex. The orientation of the methylamide of **10** is clearly resolved in the structure, and extends with good geometry to precisely overlay the C3 hydroxyl of GlcUA5 of bound HA_4 (Figure 4-6 A). The n-propyl morpholino of **12** also extends the linker into the GlcUA5 subsite, with the sequence of three methylene carbons almost superimposed to the methylamide of **10** (Figure 4-6 B). The morpholino of **12** resides one bond short of the GlcUA5 subsite. Even though these compounds extend further toward the HA binding groove than the shorter imidazole counterpart **5a**, these compounds do not antagonize HA binding as well, possibly due to their greater flexibility. The competition activity of **10** or **12** is slightly better than the counterparts of ether or ester that extend into the GlcNAc4 cavity, suggesting that both potency and HA groove extension are required to show improved capacity to compete with polymeric HA for binding of CD44.

Pyran compounds **15a** and **15b** that incorporate three methylenes and one oxygen atom were designed to allow the pyran group to occupy the position of GlcUA5 observed in the o-HA_4 complex, while tethered to the fixed subsite of the *8-amino*-THIQ. Indeed, the structure of **15a** shows that the pyran precursor reaches into the GlcUA5 subsite as expected (Figure 4-6 C). The pyran does not superimpose directly onto the GlcUA5, but it occupies a similar position. The n-propyl linker lacks functional groups, so does not make specific interactions with nearby residues of the HA groove. The induced and *apo*

conformations of Asn29 and Glu41 are occupied in roughly equal proportions in this complex, suggesting that compound **15a** occupancy is also less than 100%. As expected, a linker of the same length placed on the *5-amino*-THIQ (compound **15b**) does not provide enough extension for the pyran to reach the GlcUA5 ring position (Figure 4-6 D). The ability of **15b** to antagonize binding to HA is slightly weaker than **15a**. In the complex of **15b**, side chains of Arg155, Asn29 and Glu41 are found in a single conformation of full occupancy, but there is only partial density for the pyran suggesting low conformational homogeneity in the head group.

Table 4-5. Extended *8-amino*-THIQ Analogs

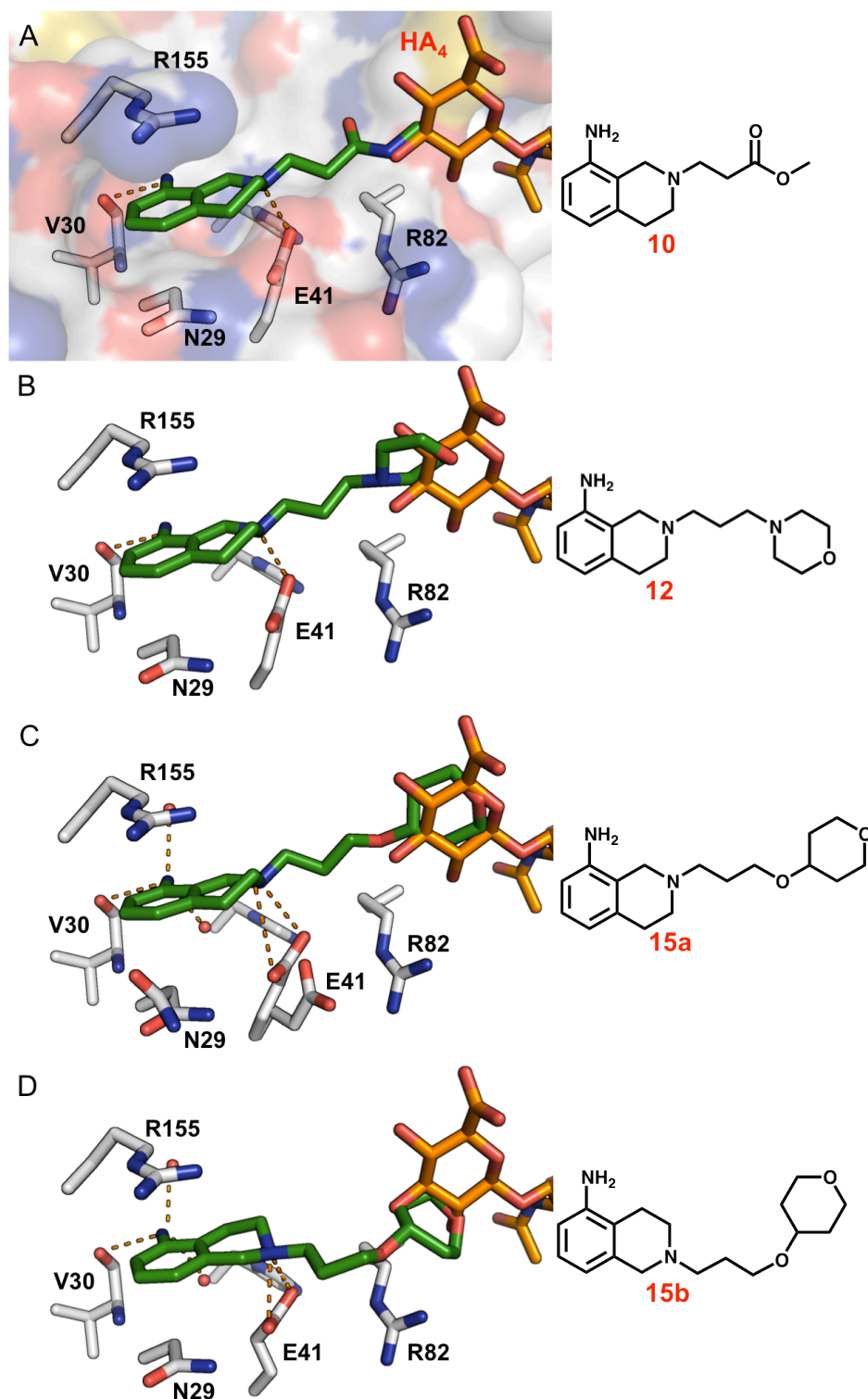| Structure	No	LE	K_D (mM) hHABD	K_D (mM) mHABD	IC_{50} (mM) hHABD	Source
	6a	0.25	0.8 ± 0.8		3.9	Glycsyn
	9	0.29	0.7 ± 1.6		4.9	Enamine
	10	0.26	0.6 ± 0.4		3.1	Enamine
	11	0.26	0.5 ± 0.4		5.0	Enamine
	12	0.22	0.7 ± 0.3		3.8	Enamine
	13	0.24	1.1 ± 0.2		1.1	Glycsyn
	14					ITDD
	15a	0.20	0.9 ± 0.2		3.4	ITDD
	15b	0.19	1.2 ± 0.6		4.3	ITDD
	B1	0.24	0.7 ± 0.3		2.7	Enamine

Figure 4-6. Crystal structures of compound **10** (A), **12** (B), **15a** (C) and **15b** (D) in relation to the HA$_4$-bound CD44 (orange).

4.3.6 Benzodiazepine as a promising alternative scaffold for future optimization

Similarities between 1,2,3,4-tetrahydroisoquinoline (THIQ) and 2,3,4,5-tetrahydro-1H-benzodiazepine (BZD) were recognized led to a consideration of the BDZ as a possible alternative scaffold for lead optimization. Molecular docking confirmed that a BZD might adopt a geometry that can engage in similar interactions with induced conformations of Arg155, Asn29 and Glu41 (Figure 4-7 A). The crystal structure with **16** reveals that the ring system adopts a twisted conformation for the seven-membered ring that positions the endocyclic nitrogen N4 directly above the carboxylate of Glu41 for hydrogen bonding (Figure 4-7 A) similar to the "unflipped" binding mode of compound **2**. This compound displays a slightly higher binding affinity for hHABD ($K_D = 5.3$ mM) in relation to compound **2**. In the crystal structure with **16**, however, side chains of Asn29 and Glu41 are found in two discrete conformations with roughly one-half occupancy, with the second conformation consistent with an unoccupied binding site. From this we infer that the ligand site is only partially occupied, but that there are further opportunities for improvement.

This prompted the synthesis of BZD analogs in collaboration with Courtney Aldrich and Joseph Buonomo (*UMN, Med Chem*). A series of synthetically accessible analogs of **16** were prepared that paralleled substitutions made to the THIQs, with the intent to engineer extensions into the HA binding groove. When possible, these included installation of an exocyclic amine at the heterocycle 9-position (*9-amino*-BZD; e.g. **16a**) that we hope would enhance potency with a hydrogen bond to Thr31 and define a
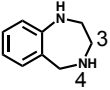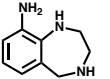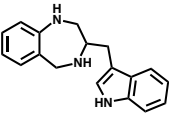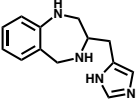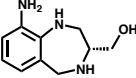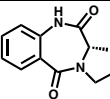
preferred orientation for the ring system. As was done in early exploration of SAR in the

THIQ series, compounds were prepared without an exocyclic amine to simplify

synthesis. Prepared compounds and measured binding affinities are give in Table 4-6.

The additivity provided by the exocyclic nitrogen is not as good in this series. The

addition of exocyclic amine at the 9-position only leads to marginal potency gains (**16a**,

K_D = 3.5 mM). Compounds **16b** and **16c** meant to emulate the imidazole *5-amino*-THIQ

of 5 possessed no detectable binding affinity. Some compounds incorporating a 3-alkyl

group have improved binding affinity. We prepared chirally pure **16d** as a precursor to a

focused library through a hydroxyl group. Notably, the installation of the methylhydroxyl

of **16d** to C3 with (R)-stereochemistry leads to modest 1.5-fold potency gains (K_D = 1.9

mM). A model of how **16e** may bind supports our conceptual design (Figure 4-7 B). A

substitution on the hydroxyl with a benzyl group improves the affinity further (**16f**, K_D =

0.8 mM), whereas a racemic mixture of **16f** lost potency significantly, suggesting the

importance of the specific stereochemistry of C3 substituents. The orientation of the *9-*

amino-BZD ring is likely fixed upon binding by the exocyclic amine in relation to Val30.

Attempts to prepare a co-crystal structure with **16a** were not successful.

The crystal structure with the tricyclic analog of **16g** reveals that the BZD is flipped
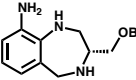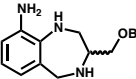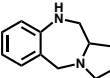
over in the binding site, and that the tertiary endocyclic amine lost the hydrogen bond

interaction with Glu41 because of an unfavorable chair conformation (Figure 4-7 C).

Despite the unambiguous structure and clear mode of binding, **16g** shows no detectable

affinity to *h*HABD, as the diol analog of **16h** (Table 4-6). A series of *6-amino*-BZD

analogs might bind favorably, however. Presumably, an aryl amine on the "flipped" BZD

ring would fix the orientation of the ring system, and the installation of alkyl groups with

correct chirality at the 3-position might be worth investigating.

Table 4-6. CD44 HABD Binding of Benzodiazepine-based Small Molecules

Structure	No	MW	K_D (mM) hHABD	Co-crystal
	16	148.2	5.3	YES
	16a	163.2	3.5	No
	16b	277.4	> 20	No
	16c	228.3	> 20	No
	16d	193.3	1.9 ± 0.1	No
	16e	283.4	0.8 ± 0.0	No
	16f	283.4	3.7 ± 0.6	No
	16g	188.3	No binding	YES
	16h	216.2	> 20	No
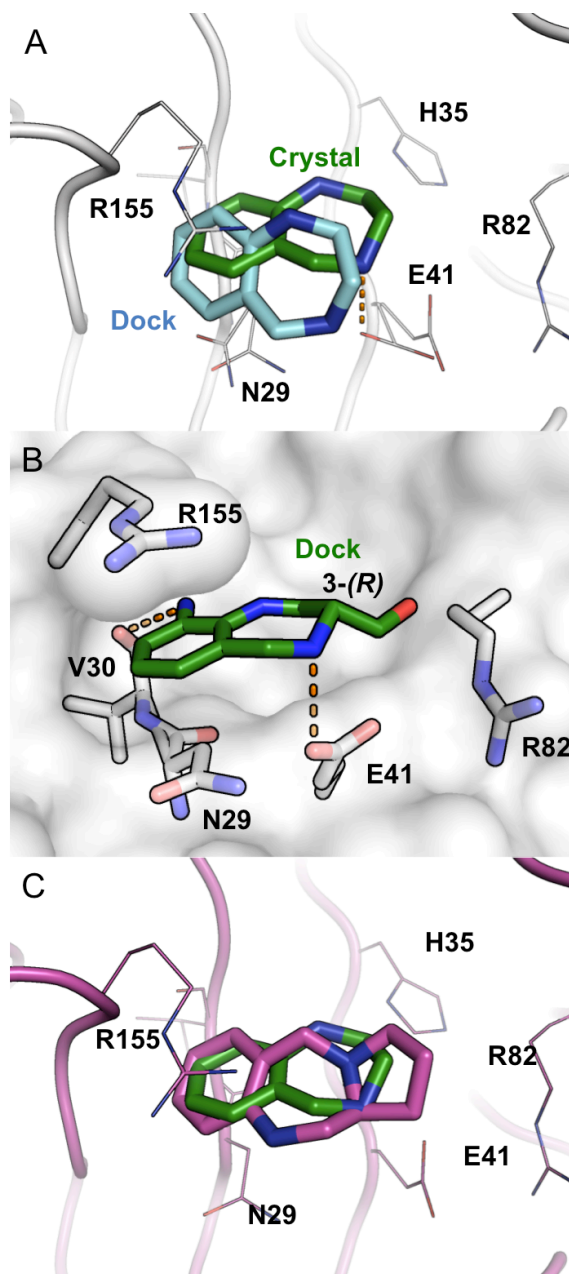
Figure 4-7. Crystal structure of diazepine-based compound **16** (A, green). The relative position of a docked binding mode of **16** is shown in A as cyan. A binding mode of **16d** was modeled to justify the intended stereochemistry imposed on the C3 atom (B). Co-crystal structure of **16h** (C, magenta) was aligned to the relative position of **16** in the crystal structure (green).

4.3.7 Compound 5a does not inhibit HA binding by other soluble HA binding proteins

We have quantified the ability of **5a** to interfere with the binding of these proteins to an HA-coated surface using SPR. We used recombinant human LYVE-1 and TSG-6 (MyBioSource, San Diego) that are close homologs to CD44 hHABD. Both LYVE-1 and TSG-6 can associate with the HA-coated surface as hHABD, and the interaction can be interrupted when the protein was pre-incubated with o-HA$_8$ (Figure 4-8). Apparently, compound **5a** does not inhibit the association of HA with either LYVE-1 or TSG-6 as it does with CD44.
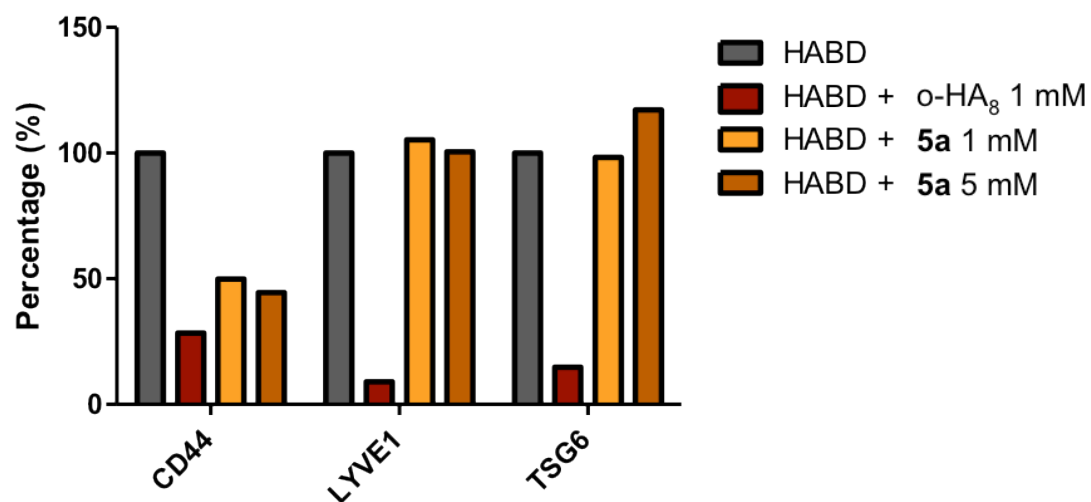
Figure 4-8. The interference of HABD binding by o-HA$_8$ and **5a** was measured using the SPR competition assay. A constant amount of HABD was passed over a HA-coated surface with or without pre-incubation with inhibitors o-HA$_8$ and **5a**. The concentration of each individual HABD was used as follow: [CD44] = 10 μM, [LYVE-1] = 20 μM, [TSG-6] = 5 μM.

4.4 Discussion

The scaffold of *8-amino*-THIQ constitutes a promising lead for further structure-driven optimization. Unlike the oligosaccharides that lie across large and exposed binding site, the THIQ molecules induce formation of a small pocket between Arg155 and Asn29 that allows them to bind with much higher ligand efficiency than oligosaccharide fragments. A local sequence alignment of close homologs of CD44 suggests the pocket is unique, and that it can be exploited to develop inhibitors specific to CD44 (Figure 4-9). His39 and Glu41 of CD44 are conserved amino acids in TSG-6, but not Arg155 or Thr31 [88]. Based on a comparison of structures (PDB-ID 2PF5), the THIQ scaffold is not expected to bind TSG-6 with any observable affinity. LYVE-1 Link domain homology to CD44 is high (51%), but among residues primarily responsible for the THIQ binding site, only Glu41 is conserved [24]. N29S, H39L, R155Q sequence differences likely mean that THIQ compounds are also ineffective inhibiting that binding of HA by LYVE-1. Indeed, our HA competition study show that there is no cross-inhibition to these proteins by one of the lead compounds, **5a** (Figure 4-8).

We have already shown that a combination of compound design elements can achieve some measure of additivity desirable for future potency optimization. However, the SAR of the *8-amino*-THIQ analogs we have examined so far is relatively flat. Additional moieties did not create much separation between strong hits and weak hits. In addition, ligand occupancy in the crystal structures does not always correlate directly with binding affinity. Some analogs with partial crystallographic occupancy yield micro-molar affinity

(e.g. **8**, **12**, **7l**). Occupancy and conformational homogeneity might be improved (or are improved) with the installation of the exocyclic amine.

Despite relatively flat SAR, the structural information has outlined critical features of the binding site and the molecular shape of a ligand that can be accommodated in it. It is possible to imagine derivatives that extend farther into the HA binding groove and that are large enough to prevent simultaneous binding of HA by steric exclusion. We think this may guarantee antagonist activity. We focused on larger molecules (e.g. **5a**, **5b**, **8**, **8a**, **10**, **12**, **15a**, **15b**) that can block HA binding by direct competition. Some of them that contribute additional H-bonds to Glu41 confer rigidity to the binding mode and display improved potency (e.g. **5a**, **8**, **8a**).

```
CD44_HUMAN  |1UUH       20----AQIDLNITCRFAGVFH-VEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALSIGF
CD44_MURINE|2JCP        23---MNQIDLNVTCRYAGVFH-VEKNGRYSISRTEAADLCQAFNSTLPTMDQMKLALSKGF
LYVE1_HUMAN            24SLRAEELSIQVSCRIMGIT-LVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALKASF
TSG-6_HUMAN|2PF5       1---------------GVYHREARSGKYKLTYAEAKAVCEFEGGHLATYKQLEAARKIGF
                                      *:      :... .:. :**    *.       *      *:: *  . .*
CD44_HUMAN  |1UUH       ETCRYGFI-EGHVVIPRIHPNSICAANNTGVYIL-T--SNTSQYDTYCFNASAPPEEDCT
CD44_MURINE|2JCP        ETCRYGFI-EGNVVIPRIHPNAICAANHTGVYILVT--SNTSHYDTYCFNASAPPEEDCT
LYVE1_HUMAN            ETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPV--SRQFAAYCYNSSDTWTNSCI
TSG-6_HUMAN|2PF5       HVCAAGWMAKGRVGYPIVKPGPNCGFGKTGIIDYGIRLNRSERWDAYCYNPHAK------
                                 ..*  *::.* *    : *   *.   .*:        : :: :**:*

CD44_HUMAN  |1UUH       SVTDLPNAFDGPITITIVN-------RDGTRYVQKGEYRTNPEDIYPSNPTDDDV
CD44_MURINE|2JCP        SVTDLPNSFDGPVTITIVN-------RDGTRYSKKGEYRTHQEDIDAS-------
LYVE1_HUMAN            PEIITT---KDPIFNTQTATQTTEFIVSDSTYSVASPYSTIPAPTTTP-P-----
TSG-6_HUMAN|2PF5       -----------------------------------------------------
```

Figure 4-9. Sequence alignment of the Link domain from human and murine CD44, LYVE-1 and

TSG-6. Residues important in forming the THIQ binding site of CD44 are highlighted in yellow,

which are absent in either LYVE-1 or TSG-6.

An effective HA inhibitor will likely need to possess a binding affinity in the sub-micromolar range. The partially aromatic ring system of a promising lead compound *8-amino*-THIQ, with its endocyclic secondary amine and the exocyclic amine are needed to achieve one milli-molar binding affinity. Additional potency might be achieved by combining molecular features illuminated by co-crystal structures. The imidazole or pyridine derivatives exhibit both improved binding affinity and better CD44-HA antagonist activity. We expect that more potent analogs will come from these series. The pyrrole analog of **5a** for example, should provide improved hydrogen bonding due to higher polarity of the N-H bond. So far, directed synthesis of this compound was unsuccessful due to severe decomposition of the end products.

Another direction that might be taken is toward a "branched" type inhibitor inspired by two complexes where the original substituents find two distinct cavities beyond the THIQ binding site (Figure 4-10 A). The close alignment of these two molecules reveals a reasonable branch position at the α-carbon attached to the endocyclic nitrogen, from which separate arms might point toward both the GlcNAc4 and GlcUA5 subsites. Specific stereochemistry at the α-carbon would be required to allow both arms to occupy preferred positions in the GlcNAc4 and GlcUA5 subsites. We have examined symmetrical substituents that simplify the synthetic chemistry, but a single molecule incorporating asymmetric branches could achieve an additive increase in potency (Figure 4-10 B).

Compounds **6a** and **9** lie in a large hydrophobic space beneath GlcUA3/GlcNAc4 that represents an attractive target for inhibitor design (Figure 4-11 A). Replacement of ester

153

oxygen of **6a** with an amide nitrogen (such as in compound **10**) should cause the arm to shift similarly from the GlcUA3/GlcNAc4 subsite toward GlcUA5. The installation of a larger chemical group beyond the original ester, such as a morpholino head group of **6b**, might extend further and make better inhibitors. A sampling of docked conformations supports the hypothesis that **6b** may occupy the space well and improve binding affinity (Figure 4-11 B). Subtle, shape based effects might influence whether a particular group stays up, or comes down.
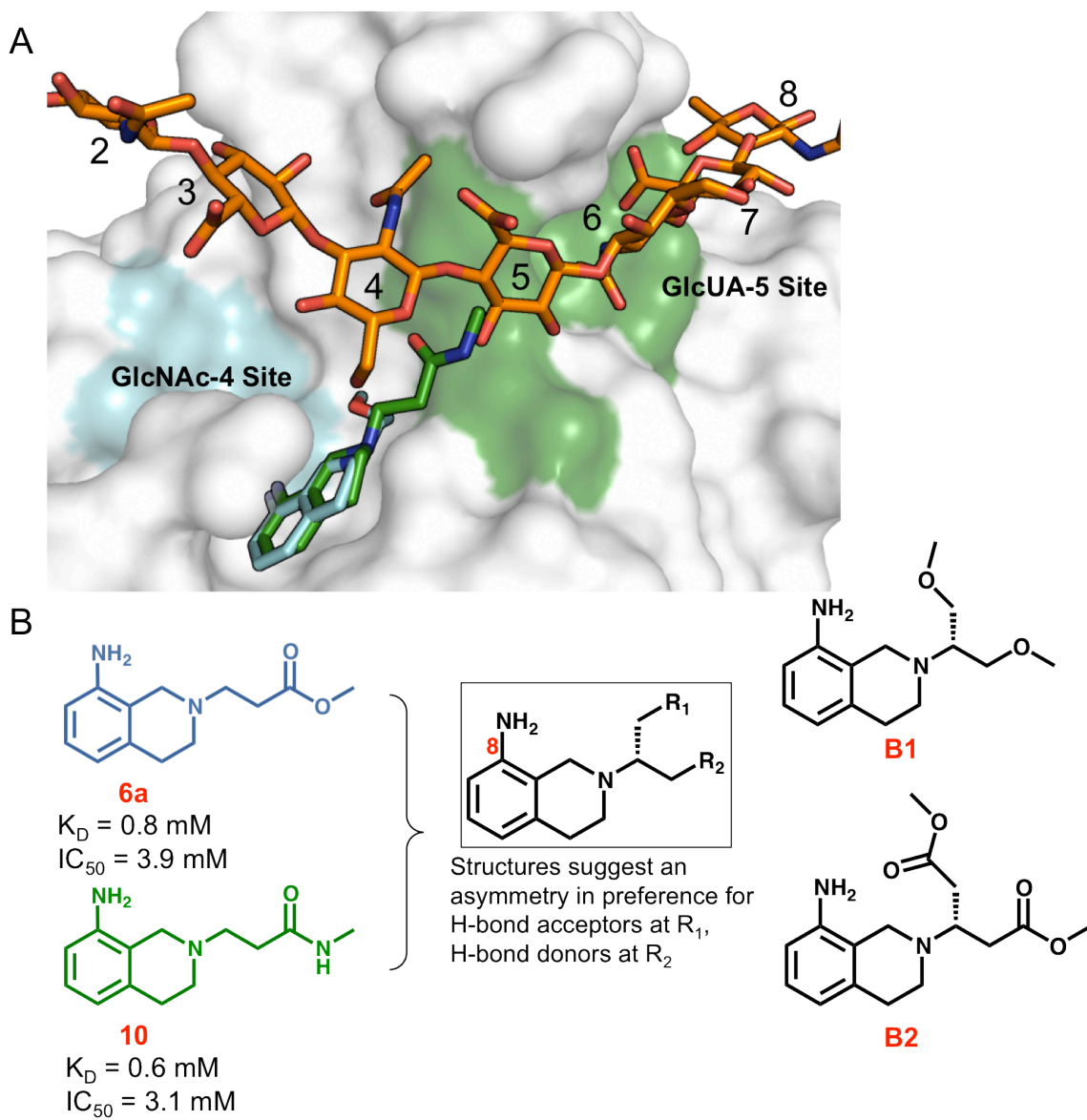
Figure 4-10. (A) Superposition of compound **6a** and 1**0**-bound CD44 complexes inspires (B) ideas for "branched" THIQs such as B1 and B2.
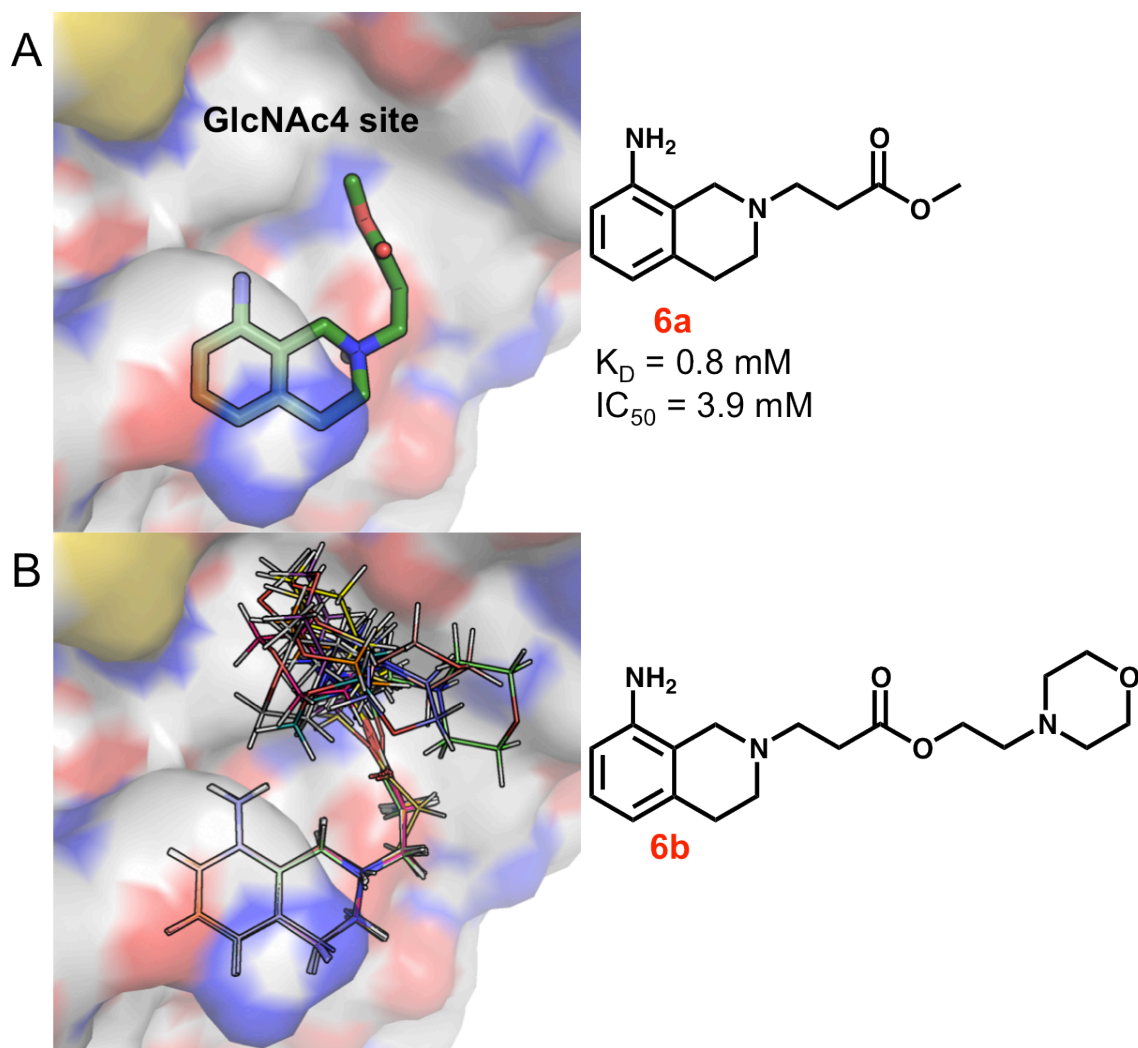
Figure 4-11. (A) The binding of **6a**-bound CD44 complexes into the cavity beneath GlcNAc4 inspires analogs with further extension to gain specificity. (B) Top docking poses of an envisioned molecule **6b**.

Perhaps the best way to improve these molecules is to fuse the *8-amino*-THIQ pharmacophore to HA-derived saccharides that can occupy subsites in the center of the HA binding groove. From our complex of *m*HABD and HA_4 (PDB-ID 4MRD), it is evident that functional groups on GlcUA5 and GlcNAc6 make very specific hydrogen bonds to CD44, and that these groups contribute much of the binding affinity of longer HA fragments. The ideal geometry for a linker to the THIQ is revealed by one of the complexes that were characterized during lead optimization. The orientation of the methylamide of **10** is clearly resolved in the structure, and extends with good geometry to precisely overlay the C3 hydroxyl of GlcUA5 of bound HA (Figure 4-12 A). A hydrocarbon linker of three methylene groups should connect the endocyclic nitrogen of the THIQ to the 3-β-hydroxyl of a saccharide occupying the GlcUA-5 subsite (Figure 4-12 B).

A few simple molecules (**14**, **15a** and **15b**) were made specifically to evaluate this design of a linker for use in THIQ-HA conjugates. In particular, the orientation of the tetrahydropyran of **15a** revealed in a complex co-crystal structure validated this linker design (Figure 4-6 C). The morpholino of **12** is connected to the THIQ via a linker that is one carbon too short; the morpholino ring lies just short of the GlcUA5 subsite (Figure 4-6 B). It would be interesting to see how a counterpart with an extra methylene before the morpholino head group would bind.

We expect that a "bidentate inhibitor" conjugating an *8-amino*-THIQ to saccharide moeities binding in both the THIQ subsite and the HA groove would have significantly improved binding affinity and efficacy in HA-binding inhibition. The surface area of

interaction of a THIQ-HA$_4$ conjugate conjoined with this linker is estimated to be twice that of either the HA$_4$ or THIQ moieties alone ($\sim 400^2$ Å), and conjugated tetrasaccharide should bind with a low micro-molar affinity comparable to that possessed by HA$_8$, which is known to effect CD44 receptor clustering and cell motility [6, 18]. Computational methods used to predict stabilization energies from docking experiments lead to a compelling hypothesis that THIQ-HA conjugates we propose will be potent inhibitors of HA binding with biological activity (Figure 4-13).

A 2,3,4,5-tetrahydro-1H-benzodiazepine (BZD) is a promising alternative scaffold of THIQ for lead optimization. Our collaboration with Dr. Courtney Aldrich has resulted in synthetically accessible analogs that paralleled substitutions made to the THIQs. Preliminary SAR studies have concluded specific stereochemistry of C3 substituents may transfer activity. The additivity expected from the addition of an exocyclic amine is not as good in this series, but other substitutions as mentioned should also be explored in the future.

Further exploration the SAR of the *8-amino*-THIQ scaffold is justified, and several routes proposals for how potency might be further increased are presented as part of this discussion. More potent tool compounds will be needed to validate CD44 as a therapeutic target. Such exploration of structural activity relationships will help to define the design criteria needed for effective binding and future drug optimization.
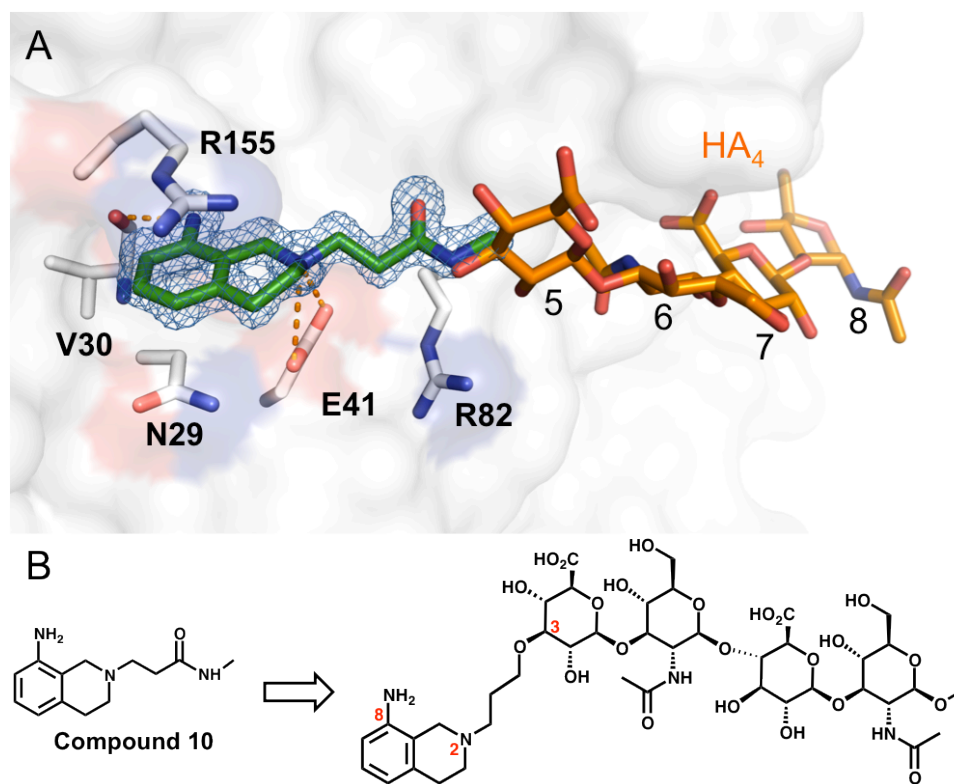
Figure 4-12. (A) Superposition of compound **10** and HA$_4$-bound CD44 complexes. The 2F$_O$–F$_C$ density for the ligand is shown in blue. (B) The superposition inspires ideas for making a "bidentate" conjugate as illustrated.
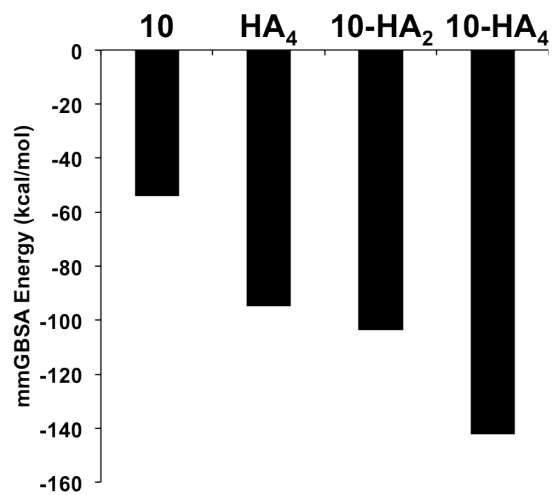
Figure 4-13. mmGBSA calculation predicts relative stabilization energy of compound **10**, o-HA4 and the "bidentate" conjugates.

CHAPTER 5

EVALUATE EFFECTS OF ANTAGONISTS OF CD44-HA

BINDING ON CELLS

5.1 Background

The accumulated understanding of ECM biology lends strong support to our hypotheses that tumor progression can be inhibited by interfering with CD44-HA interactions, and that small molecules can selectively compete for HA binding on CD44. In Chapter 4, we demonstrate that improved THIQs can compete for binding to CD44 HABD *in vitro*, and that some of them are as potent as HA tetrasaccharides (1~3 mM by SPR). It is important to measure the activity of these compounds on cells to assess their value as research tools for the study of cellular migration and growth events in biological systems. In this chapter, we will evaluate whether *in vitro* activity translates to inhibition of HA binding by CD44 on cells.

We set out to first evaluate the effect of these compounds on HA binding to tumor cells using flow cytometry. This assay has been introduced previously by others to measure the HA binding in heterogeneous populations of cancer cells using a fluorophore-labeled HA (F-HA) [89]. Different populations of cancer cells express different patterns of surface receptors that can be probed by binding to F-HA and correlated with tumor progression. We hoped to adapt this assay for a competition study to assess our small molecule inhibitors. Small molecule interference with the binding of F-HA by cells should parallel binding to the CD44 HABD quantified by the SPR HA competition assay (Chapter 2). A demonstration that competition for HA on *in vitro* binding domains translates into binding by cells will justify testing these inhibitors in more complex cell-based *in vitro* assays. Once the ability of CD44-HA antagonists to block HA binding is confirmed,

inhibitors can be validated in a scratch-wound migration assays of CD44-mediated motility [42, 90, 91], or anchorage-independent growth assays for CD44-mediated metastasis and growth [92, 93]. Relevant signaling pathways may also be explored *in vitro* [18, 72] in order to justify future preclinical animal studies.

To do these studies, CD44-null and wild-type glioblastoma cells (GBM) will be used primarily. These cells were cultured from the brain tumors of the transgenic CD44 wild-type and null mice created in the Ohlfest and McCarthy laboratories (UMN) [Decker *et al*. submitted]. The relevance of CD44 to GBM tumor progression has previously been shown [18, 91]. Also, a comparison of CD44 wild-type and null cells will help to confirm molecular specificity in the action of the small molecules. To further verify CD44 specificity, we also have access to metastatic human prostate cancer cell lines (e.g. PC3M-LN4) [89] and breast cancer cell lines (e.g. MDA-MB-231) [92] that overproduce cell surface CD44 and HA, and rely on endogenous HA for anchorage independent growth. We will also use non-cancerous RHAMM$^{-/-}$/CD44$^{+/+}$ embryonic fibroblasts provided by Dr. Eva Turley (Western Ontario, Canada) [94].

During the course of our investigation, several unexpected technical challenges were encountered that complicated attempts to quantify cell surface binding of HA in the flow experiments, primarily because the cells have mechanisms for HA turnover or internalization that are independent of CD44. Because cells have these mechanism for HA uptake that are independent of CD44, the assay (i.e. flow cytometry) cannot adequately measure cell-surface signal involving CD44 either. Similar flow cytometry assays has been applied to quantify the binding of F-HA, but they didn't distinguish

surface binding from more general cell-association [89]. Moreover, apparent cytotoxicity observed with representative compounds in CD44 null cells complicate our interpretation of the readouts in binding studies. Here, we will address these issues in the context of the binding experiments and explain why the original plan had to be altered, what turned out to be valuable, and what else can be done.

All cell biology resources were generously shared by the long time collaborator Dr. James B. McCarthy at the Department of Laboratory Pathology (UMN), whom also served as co-investigator for the AHC SEED GRANT and the MCC Brainstorm Award that supported this part of the studies.

5.2 Materials and Methods

5.2.1 Materials

Neural stem cell media for culturing GBM cells consists of DMEM/F12 (1:1) with L-glutamine and 2.438 g/l sodium bicarbonate (500 ml; Gibco), supplement with B-27 without vitamin A (Gibco), N-2 (Gibco), penicillin-streptomycin solution (Cellgro) and normocin (Invitrogen). Human EGF and FGF (10 μg/ml) were only added for cell passage. Hyaluronan sodium salt from rooster comb, chemical reagents, and other general buffer reagents were from Sigma (St. Louis, MO).

5.2.2 Tissue culture and cell lines

Maintenance of primary culture and subculturing GBM cell lines was followed as described for GBM stem cells (J. McCarthy, Personal Communication). These tumor cells isolated from CD44 +/+ WT mice are 094m2, and the cell lines 096m2 are cultured from CD44 −/− null mice. 0189m22, is a CD44 overexpressing cell line. They are denoted as WT for 094m2, KO for 096m2 and OE for 0189m22.

Cells grow as a mixed population of adherent and non-adherent cells in the neural stem cell media at 37°C under $5\%CO_2/5\%O_2$. Following plating, the media was changed twice per week or as needed depending on the number of cells and how fast they proliferate (which can be judged by the yellowing of the media). When the cells need splitting, they can be dissociated from the dishes by adding either non-enzymatic dissociation reagents or TrypLE (Gibco). Cells were subcultured by pooling the tissue culture media supernatant that contains non-adherent cells with the dissociated adherent cells prior to centrifugation. The cell pellet is resuspended, counted, and 1 X 10^6 cells were seeded in a new T25 dish in 10 ml of complete growth media ready for use or for cell passage. Fresh growth factors were added to the cells every 3-4 days for passage.

5.2.3 Migration assays

GBM cells were seeded into pre-attached wound-forming inserts at a high density (10^6) in a laminin-coated culture plate with growth medium and grown to confluence (24-36 h).

Confluent cell monolayers were formed around the wound area after removing the inserts and loose cells by washing the wells twice with medium. A count of cells migrating into the wound area on the cell culture dish was used as a measure of CD44-mediated cell motility. The migration area (%) was analyzed at 24 h from triplicate wells.

5.2.4 Anchorage-independent growth in soft agar

A layer of 1% agarose in growth medium was pipetted into six-well plates (2 ml/well) and allowed to solidify for 15 min at 4°C. Cells were suspended in 4 ml growth medium at 10^4 cells/ml and incubated at 37°C for 15 min before mixed with a softer agarose layer (0.6%). For conditions involving inhibitors or antibodies, reagents were added at the indicated concentration prior to incubation at 37°C. Four milliliters of 1.2% agarose was then mixed thoroughly with the cells, and 2 ml of the cell suspension was pipetted into triplicate wells. The wells were overlaid with 2 ml of growth medium after the softer agar layer solidified and incubated at 37°C/5% CO_2 for 21 days. The medium was replaced every 3 days. Colonies were counted in 10 random fields per well and data are shown as the average number of colonies from triplicate wells.

5.2.5 Synthesis of AlexaFluor hyaluronan conjugates (F-HA)

Fluorophore-labeling hyaluronan was accomplished with hydrazide functionalized AlexaFluor-647 (Life Technologies, CA), which covalently couples to the glucuronic acid carboxylate. To prepare the AlexaFluor hyaluronan conjugates, 2.8 g/ml 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) was first dissolved into a 1ml-buffer of 20 mM MES (pH 6.5) and 30% ethanol. 200 µl 1% HA (220 kDa, Hyalose) and 300 µl of 1 mg/ml AlexaFluor 647 hydrazide dye were added to the coupling solution and shaken at room temperature for 24 hours. Primarily, we used 220 kDa of HA to make the conjugate for most of the experiments, although other sizes were also prepared (17 or 50 kDa). The conjugated HA solution was dialyzed for four days against 2 liters of phosphate buffer saline (PBS) using a molecular weight cut-off 50 kDa membrane and the buffer was refreshed each day to remove extra free dye.

5.2.6 Flow cytometry

The GBM cells were released from the cell culture by PBS containing 0.53 mM of EDTA and washed with PBS once to remove culture medium. The cells were re-suspended in PBS and 500 µl were pipetted into individual culture tubes as 10^6 cells/ml. The appropriate amount of HA-A647 was added to those samples for HA binding experiments (typically 50 µl dye/5 X 10^5 cells) and incubated at 4°C for 60 min. When

assessing effects of inhibitors and primary antibodies, the cells were pre-treated with the reagents before added the F-HA for the same amount of time. The compounds were dissolved into DMSO added to the cell culture to give no more than 0.5% DMSO at the indicated concentration prior to incubation at room temperature. For the competition experiments CD44-specific antibodies or their isotype controls (5 or 10 μg/ml) were incubated with the cells at 4°C for 30 min, and for samples needed fluorophore-tagged secondary antibodies, the cells were washed with PBS twice, re-suspended and incubated with the antibodies at 4°C for another 15 min.

These cells were not removed from inhibitors or primary antibodies. After incubation with HA or secondary antibodies, cells were washed three times and re-suspended in 400 μl PBS on ice for flow experiments. HA binding was monitored as fluorescent events gated at 10,000 at flow cytometry Accuri C6 (BD Biosciences). Positive binding was determined by FACS median fluorescence values of populations above the background fluorescent level of intact cells before addition of F-HA.

5.2.7 MTS viability assay

Cell viability was measured with the MTS cell proliferation assay (Promega). Absorbance of UV-detectable formazan converted from the MTS reagent was used to estimate the abundance of viable cells. Growing cells were released, washed and seeded at 2×10^4 cells per well for 100 μl into 96-well microplates to adhere and grow overnight.

Then culture medium was replaced by medium containing inhibitors or vehicles at the concentrations and conditions as indicated in figure legends. Ten microliters of the MTS reagent at the suggested concentration (~300 μg/ml) was added to each well and incubated at 37°C with 5% CO_2 atmosphere. Absorbance at 490 nm was recorded in 120 min. Absorbance shown in the figures was obtained by averaging six replicates, subtracting the absorbance of cell-free equivalents and normalized to vehicle controls.

5.2.8 F-HA binding monitored by confocal microscopy

Localization of F-HA was assessed by confocal microscopy. In general, 10^5 cells were split from the sample readied for flow cytometry experiments, and then added to the poly-L-lysine-coated coverslips and fixed with 4% paraformaldehyde, then mounted to slide-flasks.

To prepare samples of adherent cells, cells were cultured on the coverslips. Growing cells were seeded to the slips (2 x 10^5 cells per well for 250 μl) and allowed to adhere and grow overnight. Next day, the culture medium was replaced with medium containing inhibitors, vehicles or antibodies for 30 minutes at 4°C before adding ~500 μg/ml F-HA and cells were then incubated further for 60 minutes.

5.3 Results

5.3.1 HA binding to CD44 expressing cells monitored by flow cytometry

In the previous chapter, we characterize the binding of a series of THIQ analogs that bind CD44 HA binding domains and assessed their ability to compete with polymeric HA using the immobilized HA SPR assay (See Chapter 2). To quantify interference with the binding of HA by cells, we have used fluorescent-labeled HA (F-HA) and GBM cells in flow cytometry to evaluate how much F-HA binds to the cells and how well small molecules or blocking antibodies (as positive controls) decrease binding (Figure 5-1).
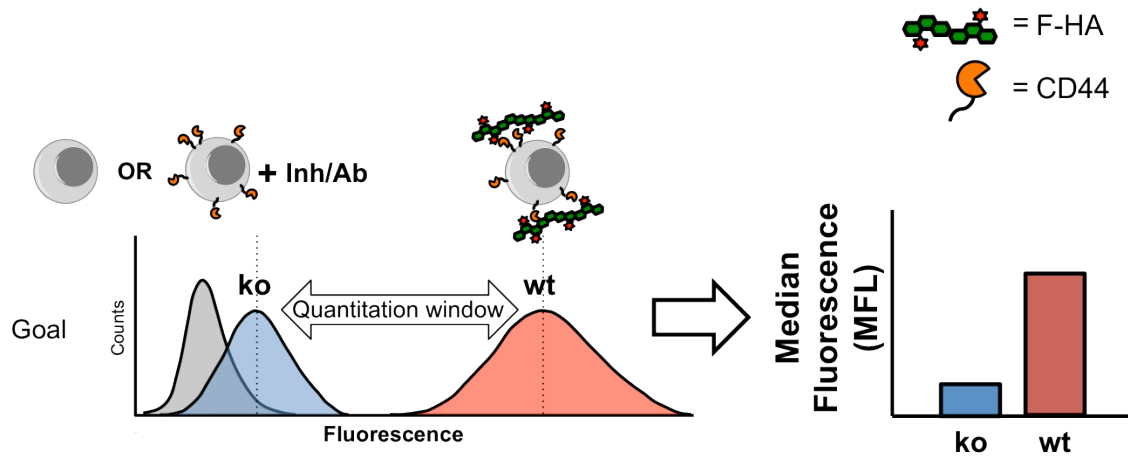
Figure 5-1. Fluorophore-labeled HA-binding measured by flow cytometry. Binding will be confirmed when the fluorescence of a population rises significantly above the background fluorescence of intact cells before addition of F-HA (grey, left panel). We expect to see a quantitative signal separation in F-HA binding between the CD44 expressing cells and cells pretreated with CD44 blocking controls (goal). The data can be re-plotted as a bar graph that compares the median fluorescence (MFL) of each individual samples (right panel).

We first characterized the prevalence of CD44 on the surface of KO, WT and OE glioma cells. Cells pre-treated with an anti-CD44 antibody IM7 were stained with a fluorophore-labeled secondary antibody (anti-mouse IgG-AlexaFluor647), and analyzed in flow cytometry. OE glioma cells acquire a fluorescence level ten times higher than the WT cells (10^6 vs. 10^5), whereas the gain of fluorescence by KO glioma cells was negligible (Figure 5-2 A). This result correlates well with our Western blot analysis of these cells; CD44 OE cells produce much more CD44 (Figure 5-2 B).

Next, we measured the binding of F-HA of these cells by flow cytometry 1 hour after addition of labeled HA to cells at 4°C. However, the apparent binding by OE or even WT cells was not elevated as was anticipated. Rather, CD44 null cells consistently showed higher fluorescence consistent with increased HA binding than CD44 positive cells (Figure 5-3).

We considered the possibility that F-HA might be internalized by CD44-independent endocytosis. A series of endocytic inhibitors known for controlling receptor- or non-receptor-mediated endocytosis were investigated [95], including chlorpromazine for clathrin-mediated endocytosis, nystatin for caviolin dependent endocytosis, methyl-β-cyclodextrin for cholesterol depletion, cytochalasin D as an actin polymerization inhibitor, genistein as a protein tyrosine kinase inhibitor, and amiloride for macropinocytosis [96]. When cells were pretreated with these endocytic inhibitors, only chlorpromazine reduced some internalization. In subsequent flow experiments, chlorpromazine was used it to try to minimize non-specific HA-uptake, but with minimal effect.

172

When unlabeled HA of the same approximate molecular weight (220 kDa) is co-administered to compete with the fluorophore labeled HA, a partial reduction to 60% of the baseline is observed (Figure 5-4 A). Anti-CD44 antibodies that reportedly block HA binding (IM7 and KM201 [72, 89]) produced only modest inhibition (~ 30%) in CD44 OE GBM cells (Figure 5-4 B), suggesting only 30% of the F-HA signal appears to be CD44-related.
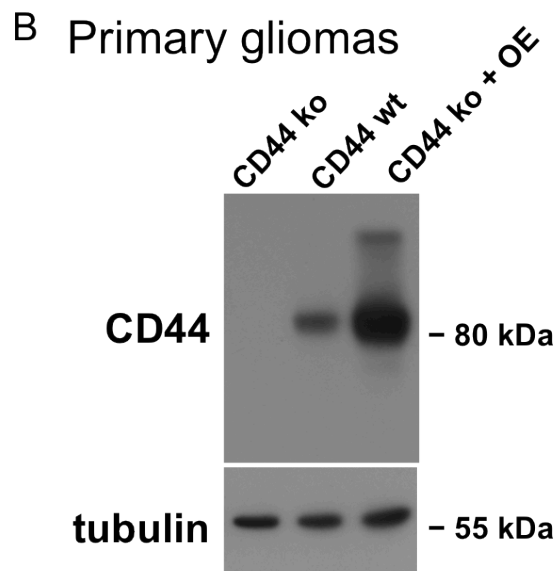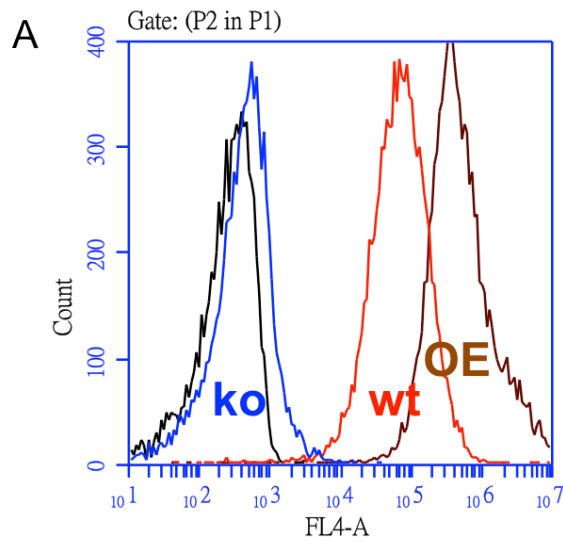
Figure 5-2. GBM cell surface expression of CD44 has been confirmed by flow cytometry with a fluorophore-labeled anti-CD44 antibody, anti-mouse IgG-AlexaFluor647 (A). The cells over-produce CD44 (OE) has the median fluorescence at 10^6 compared to the WT at 10^5, whereas the KO cells has similar signals as the untreated sample (black). CD44 production of these gliomas was also analyzed by Western blotting (B). Shown a band matched the target mass of CD44 at 85 kDa. Tubulin was used as a loading control.
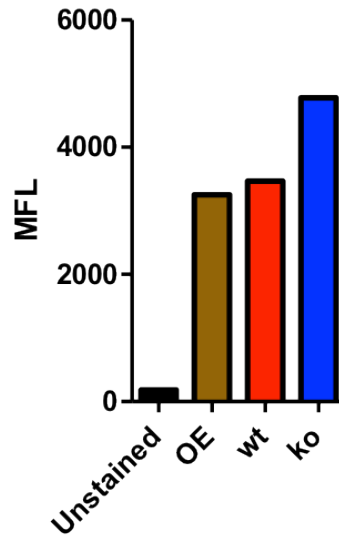
Figure 5-3. Glioma cells that express different amounts of CD44 were used to investigate whether the HA binding is CD44-specific by pre-treating the cells with F-HA. The binding of F-HA of these cells by flow cytometry 1 hour after addition of labeled HA to cells at 4°C. The basal level of F-HA binding (shown as median fluorescence) is independent of cell surface CD44 expression: the KO cells associate with much more F-HA than the OE cells.
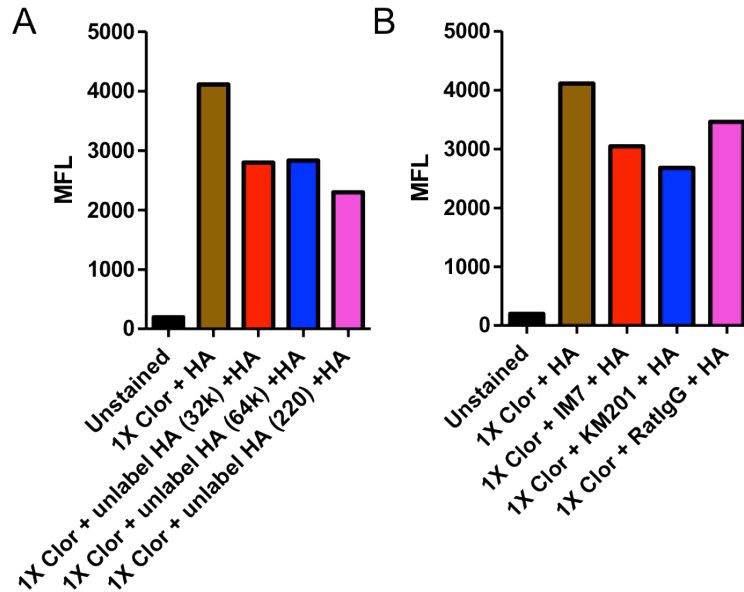
Figure 5-4. CD44-blocking reagents were used in the competition experiments to assess CD44-related F-HA binding to cells overproducing CD44. CD44-specific antibodies or their isotype controls (5 or 10 µg/ml) were pre-incubated with the cells at 4°C for 30 min before addition of F-HA for an hour. Unlabeled HA of various sizes only partially block the F-HA binding (A). Larger sizes of unlabeled HA (220 kDa) block F-HA (220 kDa) more efficiently, to 60% of the baseline (A). Anti-CD44 antibodies also inhibit some of the F-HA binding (B). These experiments were done with reagents indicated followed by pre-incubated the cells with 1x chlorpromazine (10 µg/ml) as an endocytic inhibitor that reduces part of the HA internalization.

Compound **5a** was evaluated to see if it could alter F-HA binding by cells. These cells were pre-treated with **5a** at 1 mM before F-HA was added for 30 minutes. However, there is no clear indication of CD44 dependent response of glioma cells to the compound. Compound **5a** can partially decrease the F-HA binding observed to CD44-expressing glioma cells (both WT and OE) in a dose-dependent manner, but also decreases F-HA binding by CD44-null cells. For example, compound **5a** resulted in a collective decrease (40%) in F-HA binding at 1 mM when compared with the vehicle control (Figure 5-5). The reduction of F-HA binding is dose-dependent to these cells when treated **5a** at lower concentrations (not shown).

Similar experiments were repeated using fibroblasts and prostate cancer cells that overexpress CD44 and bind lots of F-HA on the cell surface (See 5.3.2). We expected to see more CD44-mediated F-HA binding by these cells and a greater F-HA binding should be blocked by CD44 inhibitors. But when pre-treated with anti-CD44 antibodies, both cell lines were able to associate with much of F-HA, suggesting that CD44 may not solely account for F-HA binding (Figure 5-6 or 5-7 A). Unexpectedly, when pretreated with compound **5a**, these cells show a robust increase in HA staining over the DMSO control (Figure 5-6 & 5-7 B). We suspected if this was a single-molecule effect, so **3**, **8**, **8a** and **9** were also evaluated. Compound **8** and **8a** seem to profoundly exaggerate this effect. Inhibitor titration reveals a dose-dependent increase in binding of F-HA to cells, regardless the molecular weight of F-HA used (17, 50 or 220 kDa) (not shown).
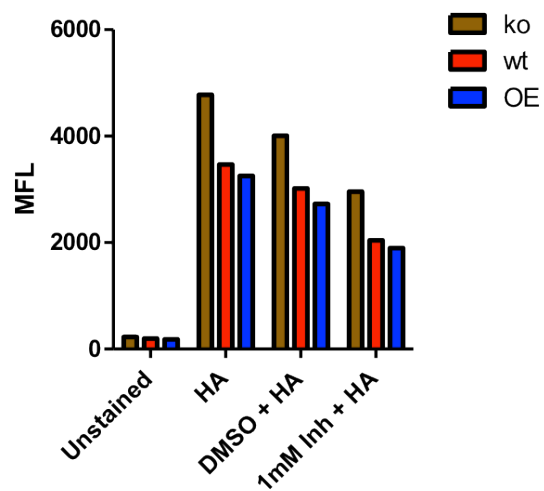
Figure 5-5. Compound **5a** alters F-HA binding by GBM cells. Cells expressing different amount of CD44 were pre-treated with **5a** before F-HA was added for 30 minutes at 4°C and measured by flow cytometry. Compound **5a** can partially decrease the HA binding to all the GBM cells. A collective decrease (40%) in F-HA binding was observed with **5a** of 1 mM when compared with the vehicle control.
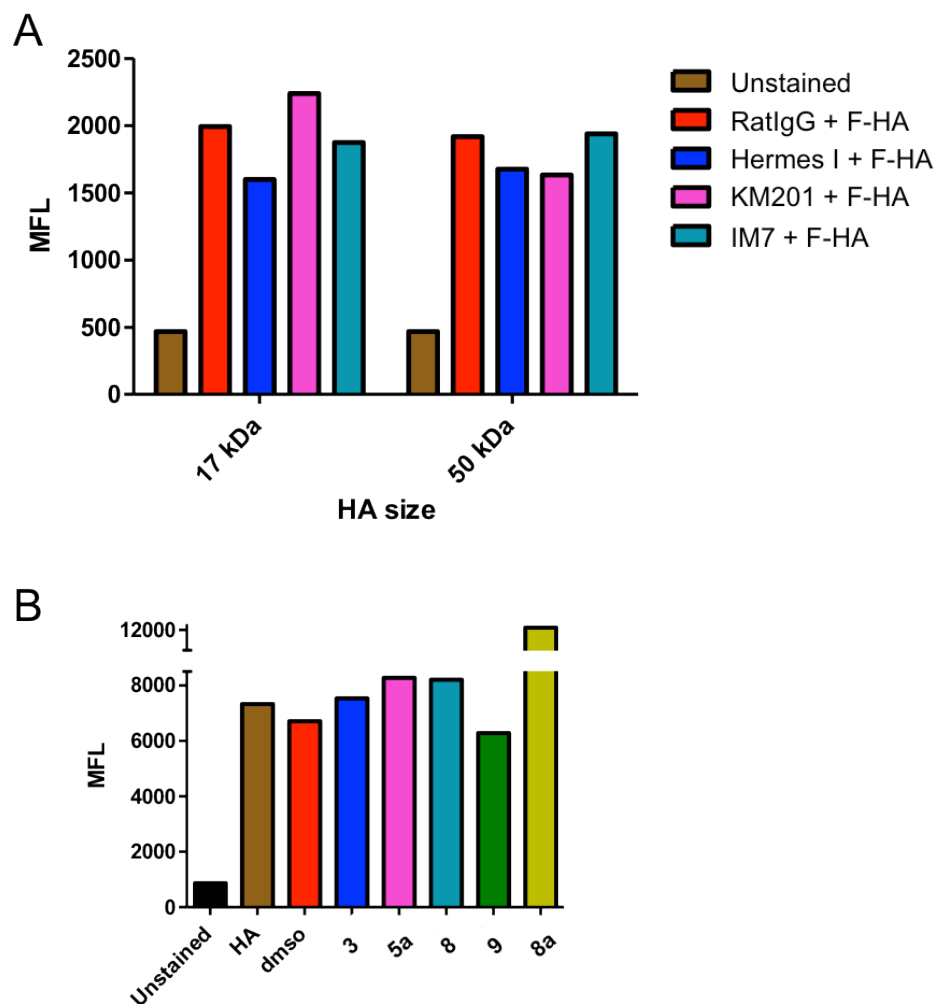
Figure 5-6. F-HA binding on prostate cancer cells (PC3M-LN4) that overexpress CD44 measured by flow cytometry. Antibodies (5 or 10 μg/ml) or THIQ compounds (400 μM) were pre-incubated with the cells at 4°C for 30 min before addition of F-HA for an hour. (A) Anti-CD44 antibodies (Hermes I, KM201 and IM7) only partially inhibit the F-HA binding to the cells regardless of the size of F-HA (17 or 50 kDa) when compared to the rat isotype control IgG (red). (B) Some selected compounds can enhance the F-HA binding to cells as compared to the F-HA treated control (brown).
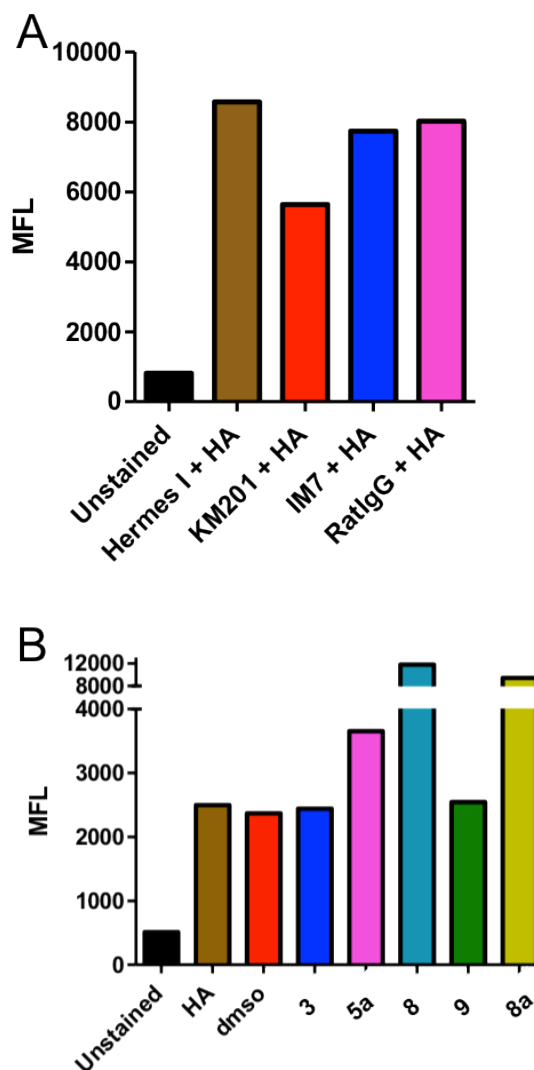
Figure 5-7. F-HA binding on embryonic fibroblasts ($RHAMM^{-/-}/CD44^{+/+}$) that overexpress CD44 was observed by flow cytometry. Antibodies (5 or 10 μg/ml) or THIQ compounds (400 μM) were pre-incubated with the cells at 4°C for 30 min before addition of F-HA for an hour. (A) Anti-CD44 antibodies only partially block the HA binding to the cells when compared to the rat isotype control IgG (pink). (B) Some selected compounds (e.g., **5a**, **8** and **8a**) can greatly enhance the F-HA binding to cells when compared to the F-HA treated control (brown).

5.3.2 HA binding to CD44 expressing cells monitored by confocal microscopy

As we considered the possibility that F-HA might be internalized by the cells we selected, we performed confocal microscopy to monitor the cellular localization of F-HA binding. We used prostate LN4 cells instead of GBM cells because they stained much brighter by F-HA. The sample readied for flow cytometry experiments was used for the confocal imaging and Z-sections of the images were taken from the apical to basal side. When **5a** is co-administered with F-HA, a great increase of the F-HA binding is observed in both cell surface and cytosol as compared to the vehicle-treated sample with only F-HA, indicating that both surface binding and internalization of F-HA were enhanced by **5a** (Figure 5-8).

To test whether adherent cells may exhibit differential binding, we prepared LN4 cells on cover slips, and quantified F-HA uptake by confocal microscopy with these cells in the presence of antibodies and compound **5a** (Figure 5-9). Strong nuclear staining of cells was observed when an isotype control antibody IgG was co-administered with F-HA, and a blocking antibody KM201 did not have any effect. Nuclear staining was not prominent in the presence of DMSO. Interestingly, when compound **5a** is pre-treated at 0.4 mM, both surface and intracellular binding was greatly reduced. Additional investigation revealed that the reduction was not in a dose-dependent manner (not shown).
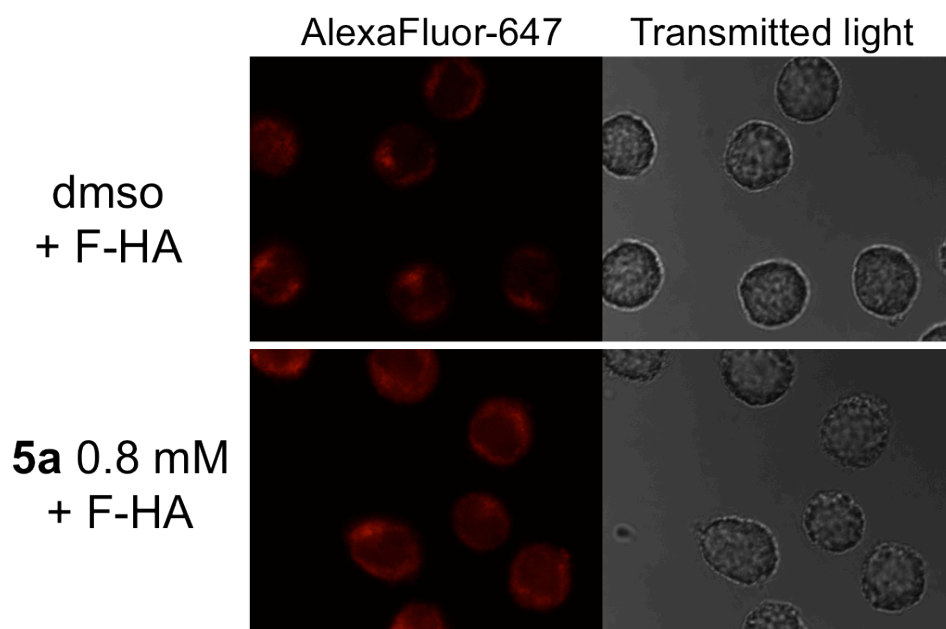
Figure 5-8. A middle Z-section of confocal microscopy images reveals that non-adherent prostate cancer PC3M-LN4 cells pretreated with compound **5a** show a robust increase in F-HA staining over the DMSO control.
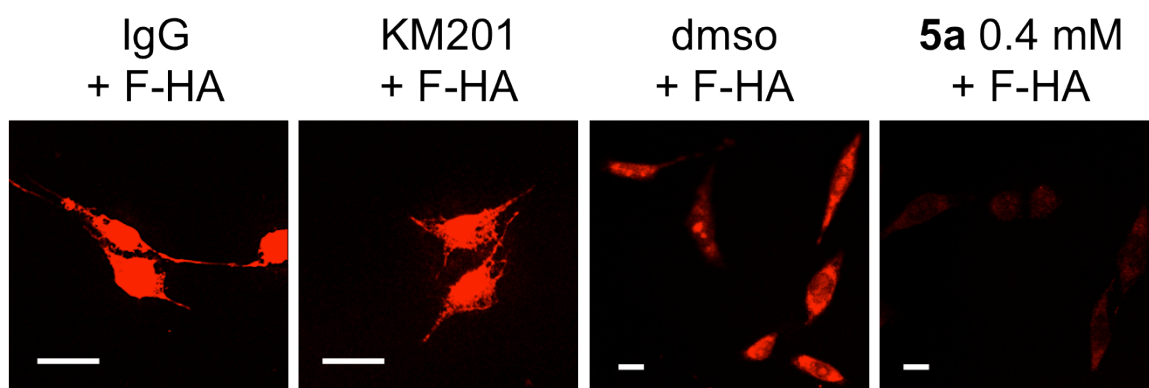
Figure 5-9. A middle Z-section of confocal microscopy images reveal the cellular localization of F-HA in adherent prostate cancer PC3M-LN4 cells, when pretreated with antibodies, DMSO and compound **5a**. The scale bar in the panels is 10 μm.

5.3.3 Compound 5a inhibits the growth of LN4 cells

We evaluated the effect of compound **5a** on the growth of prostate cancer LN4 cells in 3D soft agar culture. The compound inhibits anchorage-independent growth in a dose-dependent manner with an EC_{50} of 285 μM, showing a moderate level of tumor cell survival in soft agar (Figure 5-10 A). Our cell proliferation data support a conclusion that **5a** slows down the cell growth rate in the dose range over five days and limits exponential growth of the cells at 400 μM (Figure 5-10 B). There was some initial killing of seeded cells in 24 hour with an increase of inhibitor concentration, but some populations continue to grow out and proliferate over time. No growth of these cells was observed at higher **5a** concentrations (800 μM). These results suggest that the inhibitor may negatively regulate the survival of LN4 cells.

Indeed, **5a** inhibits cell survival with EC_{50} of 600 μM (Figure 5-10 C), confirming the cytotoxic activity. We next analyzed the change of cell numbers over two days for individual concentrations. When exposed to a single dose of **5a** at 400 μM, cultured cells stay cytostatic and maintain constant cell number for two days, but then growth resumes at lower concentrations of **5a** (Figure 5-10 D).
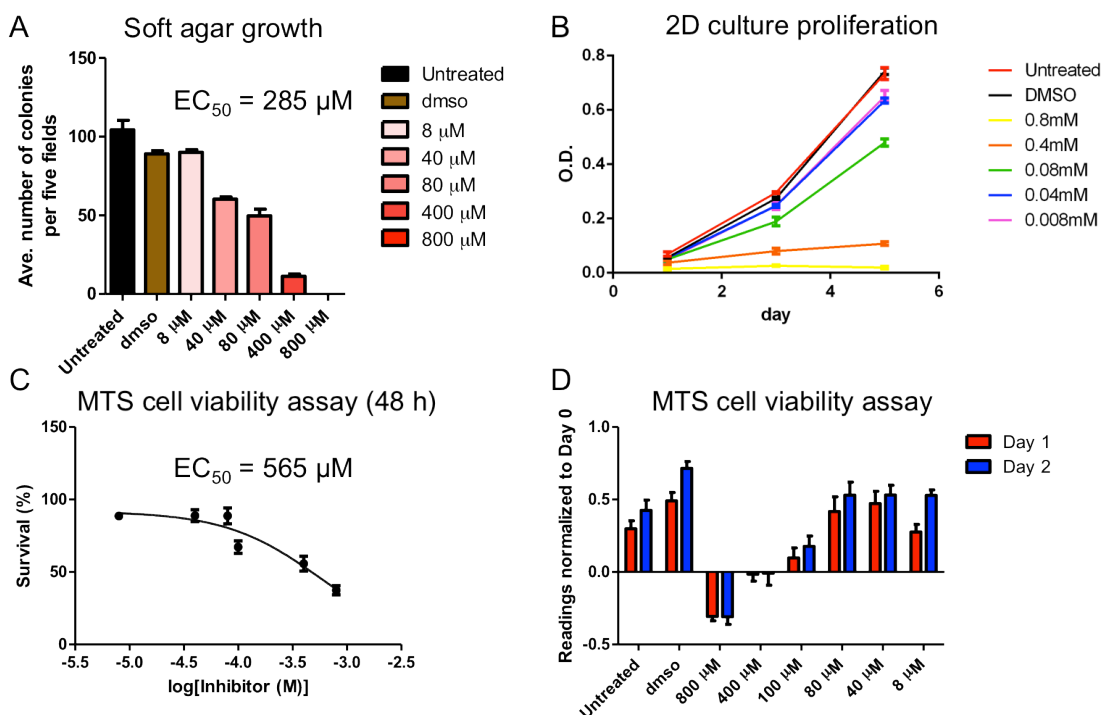
Figure 5-10. Compound **5a** inhibits the growth of LN4 cells. (A) Soft agar for anchorage-independent growth. Single cells (10^4) were suspended in soft agar supplement with growth media and **5a** for colony formation in two weeks. Colony counts are plotted for observations made with varying concentration of **5a** and DMSO (n = 3). (B) Cell proliferation assay for cell growth over five days. Cells (10^3) were plated in a culture dish and counted on day 1, 3 and 5. Optical density of cells (490 nm) reflects the amount of viable cells. No exponential growth of these cells was observed at concentrations higher than 400 μM (n = 3). (C) Cell viability was assayed in 2 days (n = 6). The efficacy of **5a** was expressed as an effective concentration where 50% of cells are viable (EC_{50}). (D) The cell viability assay plotted for the change of viable cell number in either 1 or 2 days (n = 6).

5.3.4 Verify CD44 specificity of small molecule inhibitors in growth inhibition using gliomas

We re-examined the toxicity of **5a** toward glioma cell lines that express different amounts of CD44 to assess the relationship between CD44 expression and the ability of **5a** to effect anchorage-independent growth and survival. We observed marginal toxicity of **5a** to gliomas in the MTS cell viability assay. When cells were pre-treated with **5a** at various concentrations for 30 minutes, CD44 WT and OE glioma cells show a reduction of viable cells to 75% of the baseline, whereas the KO cells are less affected (90%), suggesting that CD44 positive cells are less tolerant to **5a** than CD44 null cells. There is some indication that the survival of glioma cells treated with compound **5a** may depend on the amount of CD44 expressed by those cells. Cells are not affected with compound **3** at comparable doses, suggesting that the toxicity does not extend to all compounds with the THIQ pharmacophore (Figure 5-11). When treated with **5a** for 48 hours, the CD44 positive cells show a lower EC_{50} than the CD44 null cells (220 μM vs. 650 μM) (Figure 5-12 A).

The wild-type cells grow more colonies than the null cells in the anchorage-independent environment. An anti-CD44 antibody (KM201) apparently inhibits growth of the wild type cells, but is ineffective to the null cells, suggesting that inhibition of growth is a CD44-dependent process (not shown). Compound **5a**, however, appears equally capable of inhibiting the growth of either cell line, with an EC_{50} of 171 μM and 73 μM, respectively (Figure 5-12 B).

We also looked at the effect of other molecules in the series (Table 5-1). By SPR, compound **8** inhibits HA binding *in vitro* with an IC_{50} (0.95 mM) 3.5-fold more potent than **5a**. It inhibits growth of both CD44 positive and GBM null cells with EC_{50} of 42 µM and 38 µM, respectively. It is a more effective inhibitor of WT glioma growth, but the compound is even more indiscriminately cytotoxic than **5a**.

Compounds **6a** and **9** utilize a different binding mode that curls into the GlcNAc4 subsite (See Chapter 4.2.3). The *in vitro* binding affinity and competition activity of compounds **6a** and **9** are slightly improved over that seen for *8-amino*-THIQ compound **7**. Interestingly, neither these compounds have any significant effect on the survival of either cell line at any dose. At the highest dose, compound **9** killed the cells, but at nearly every other dose it had no effect, with an EC_{50} of 0.88 mM for the CD44 wild type cells and 1.05 mM for the CD44 null cells. These results suggest that these compounds do not share the cytotoxic properties of **5a** or **8**.
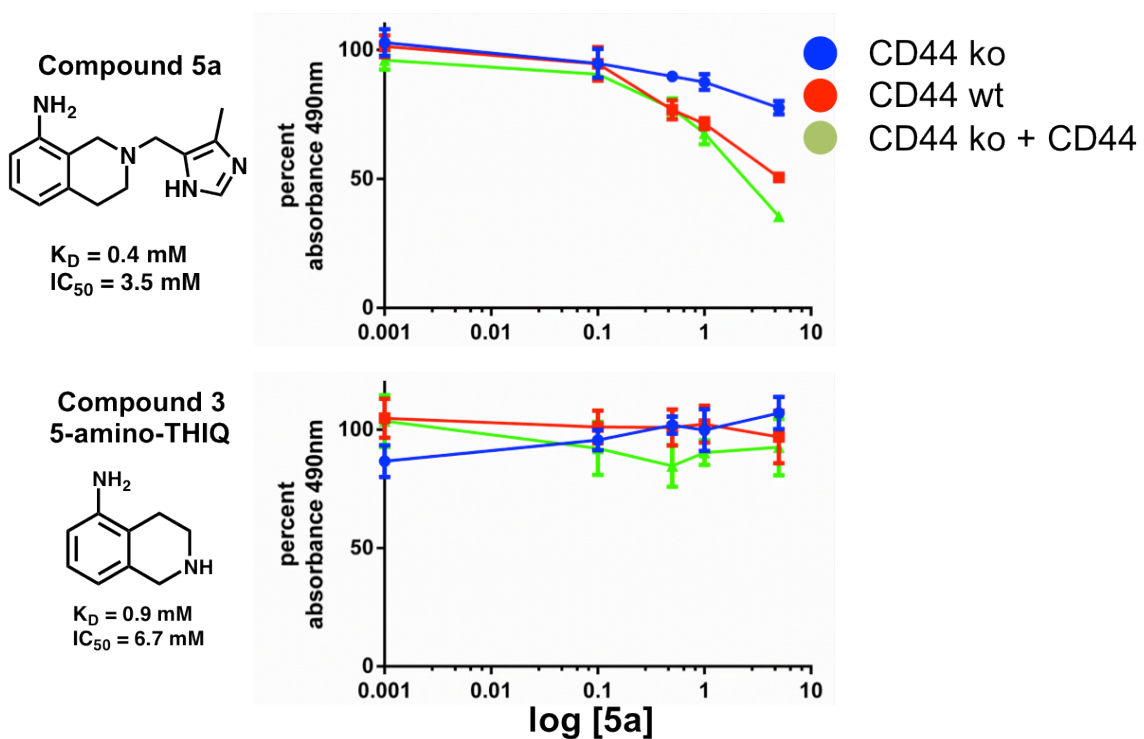
Figure 5-11. Compound **5a** but not compound **3** confers cytotoxicity against glioma cells expressing CD44 in the MTS viability assay. Cell viability was assessed after compounds were added for 30-min pre-incubation. Absorbance at 490 nm for individual conditions was normalized to the DMSO-treated sample.
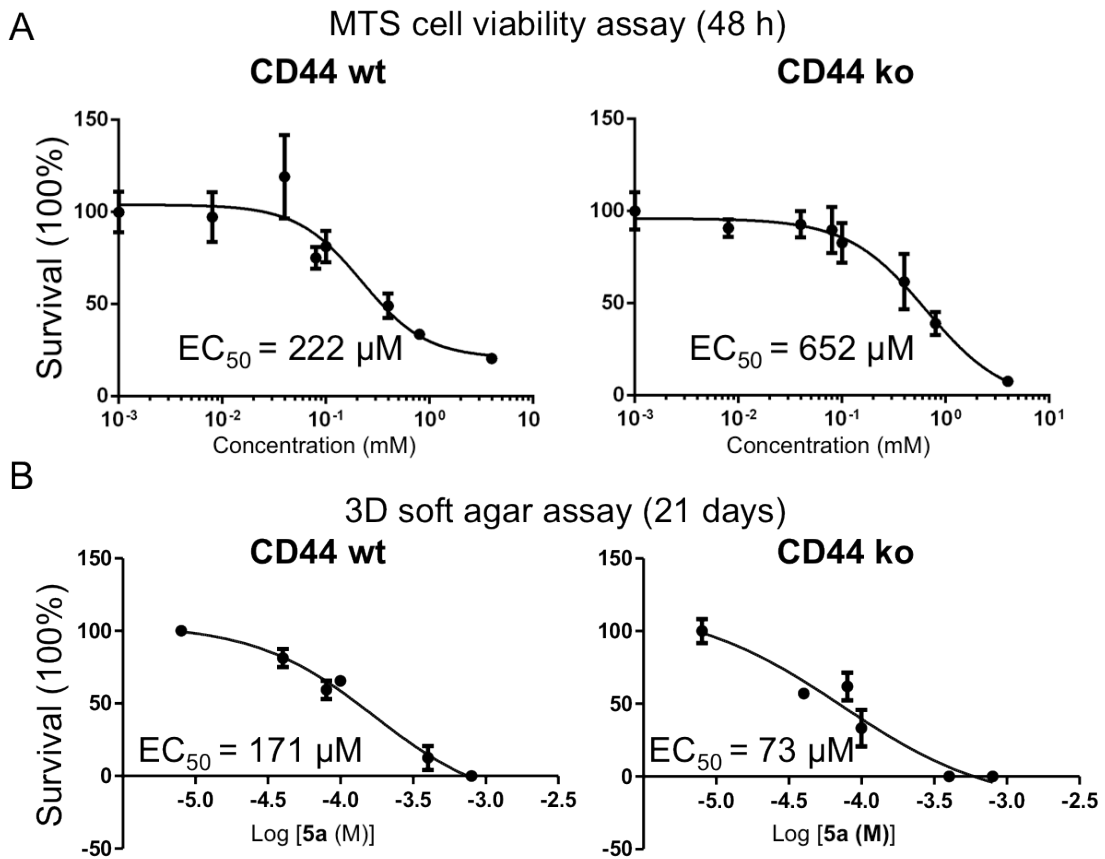
Figure 5-12. Compound **5a** inhibits the growth of glioma cells. (A) MTS cell viability assay in 2 days. (B) 3D soft agar growth in 21 days.

5.3.5 Establish cell migration assay using gliomas

We attempted to evaluate the motility of glioma cell lines that express different amounts of CD44. Smooth muscle cell migration and motility have previously been evaluated using a scratch-wound migration assay [42]. Therefore a scratch wound assay was used, in which wound closure on laminin-coated wells can be used as a measure of CD44-mediated cell motility. We expected to see a clear difference in the migratory ability of the CD44 bearing cells (WT) compared to those tumor cells that lack CD44 (KO). However, we were unable to plate the cells onto the dish as a confluent monolayer so that a count of cells migrating into the scratched area on the cell culture dish cannot be measured. Future directions should include different substrates of CD44 (e.g., HA, collagen or fibronectin) in place of laminin to be coated on the dish [72].

5.4 Discussion

The objective of our research is to create potent and highly selective inhibitors of HA binding to CD44 so that the pharmacology of such novel molecules may be explored. We had assessed the cell-based activity of THIQ inhibitors toward HA binding, cell growth and cytotoxicity, and demonstrated that these compounds do not act selectively for cells expressing CD44, although there were some hints of CD44 dependency. We have been working to establish *in vitro* bioassays that make use of CD44 positive and CD44 null

glioma cell lines isolated from mice in order to confirm that the small molecules interfere with the binding of HA facilitated by CD44 on these cells.

We have used fluorescent-labeled HA (F-HA) in flow cytometry to evaluate how much F-HA binds to the cell surface. However, the flow cytometry experiments have revealed that an unexpected amount of HA is taken into cells independent of CD44. This constitutes a problem for the initial experimental plan: most HA binding was expected to be CD44-mediated. The fact that some F-HA was present inside the cells increases the background and therefore limits the ability to assess the effect of CD44-specific binding for blocking reagents or inhibitors using this methodology. Attempts to prevent cellular uptake of HA by conducting experiments at 4°C or by using endocytic inhibitors were not entirely successful. We also evaluated F-HA binding by LN4 cells and fibroblasts that over-produce CD44 on the cell surface with similar studies and they behaved similarly. We think this is likely true for any other cells.

Confocal microscopy revealed an increase in both surface binding and internalization of the labeled HA in human carcinomas and embryonic fibroblasts caused upon the addition of the THIQ compounds (Figure 5-8). This is a very unusual activity. Interestingly, not all the THIQ-based inhibitors increase F-HA binding (Figure 5-6 & 5-7). These compounds must alter the way cells recognize and/or process HA, potentially through binding to CD44. The *5-amino*-THIQ (compound **3**) itself does not affect F-HA binding, nor does compound **9** bearing an ethyl ether linker, whereas equally potent inhibitors that have aromatic rings or charged groups (e.g., **5a**, **8** and **8a**) generate robust F-HA signals. The molecular mechanism of increased binding is unknown. It has been

191

shown that oligo-HA binding induces CD44 internalization [18, 97]. It is likely that HA catabolism or receptor recycling by these two cell types has changed upon binding to specific compounds. Some of these compounds that change the local HA concentration of fibroblast may have potential value for research in wound healing or tumorigenesis.

F-HA binding was measured on suspended cells using flow cytometry, but cells may bind HA differently when cultured at a surface. For instance, fibroblasts bind HA via CD44 in the apical membranes, but via RHAMM when suspended. Previous studies have shown that a CD44-specific antibody (KM201) blocked the binding and uptake of F-HA to 2D-cultured human metastatic breast cancer cells (MDA-MB-231) that overexpress CD44 and exhibit high F-HA binding [89]. We have shown that KM201 is not sufficient to inhibit the binding or uptake of F-HA to the prostate cancer cells (LN4) in 2D culture. Strong nuclear staining of cells when pre-treated with antibodies was observed at 4 °C, suggesting that these cells were capable of receptor trafficking upon binding to F-HA at low temperature. Contrary to the non-adherent cells, incubation with compound **5a** did not generate robust increase in F-HA uptake above the control levels. Instead, compound **5a** reduced the cytoplasmic and nuclear accumulation of F-HA in LN4 cells, albeit in a dose-independent manner. In both cases, F-HA binding analyzed by confocal microscopy allows us to evaluate the consequences of inhibitor treatment by comparing the amount of cytosolic or surface-bound F-HA among samples.

The pronounced inhibition of the anchorage-independent growth of human prostate carcinoma [98] by **5a** sustains our interest in the possibility that this effect is CD44-mediated (Figure 5-10 A). These cells require CD44 to interact with endogenous HA for

malignant growth in an anchorage-independent condition like soft agar. Compound **5a** inhibited colony growth of LN4 cells in a dose-dependent manner; the colony count decreased by 50% between the dose range of 80 μM and 400 μM (Figure 5-10 A). Cells can continue to grow out over a long period of time when less than 400 μM of **5a** is present, and can remain cytostatic at higher concentrations (up to 5 mM). This is a very different cellular activity than that observed from F-HA binding measurements.
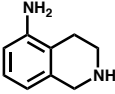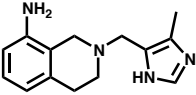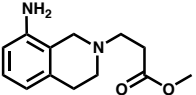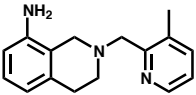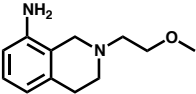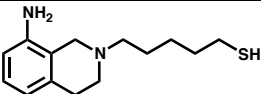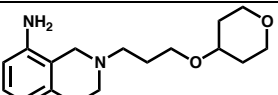
In order to further characterize the THIQ compounds as inhibitors of growth, we examined cell viability (or survival) using MTS reagents. Cell proliferation assays confirm that **5a** negatively regulates the survival of LN4 cells (Figure 5-10 B–D). To determine whether or not this effect on survival has anything to do with CD44, we re-examined the effect on the anchor-independent growth and survival of gliomas that express different amounts of CD44. Our data suggest that there is some CD44-dependent killing by **5a**. CD44 positive cells are less tolerant of **5a** than CD44 null cells (Figure 5-12 A). As for anchorage-independent growth, although CD44 null cells grow fewer colonies due to the lack of CD44, **5a** is equally capable of inhibiting the growth of either cell lines (Figure 5-12 B). These results affirm a need to identify more potent antagonists that can be dosed at more reasonable concentrations.

We have examined the activity of other THIQ inhibitors for the cell survival of gliomas, and explored the possibility of a correlation between binding mode, effective HA competition and CD44-mediated cytotoxicity, but none of the current compounds stand out. Only specific compounds negatively regulate the survival of gliomas, but the activity is not selective to CD44 positive cells (Table 5-1). The pharmacophore *5-amino-*

THIQ does not affect cell viability (Figure 5-11). Only compound **5a** has moderate activity and slightly better selectivity to CD44 positive cells. Both compound **8** and **13** are more effective against CD44 positive cells, but they are even less selective. Their activity cannot be attributed to binding mode because compound **8** occupied the same footprint as **5a**, but they are both more effective in HA competition. Nevertheless, the potential off-target effect could mask any apparent CD44-related effect through receptor-independent toxicity. It would be reasonable trying to do either the flow experiments or the migration with one of these compounds that is less cytotoxic (e.g. **6a**, **9** and **15a**).

A variety of different studies have shown that o-HAs can influence cell-adhesion processes through CD44. It is likely that many of the biological effects on cells attributed to o-HAs can occur upon inhibitor binding simply due to the dissociation of receptors bound to the p-HA of the ECM. Our current results motivate a future search for more appropriate cell systems for binding studies by flow cytometry, as well as for more potent and selective antagonists. An effective HA inhibitor will likely need to possess a binding affinity that is stronger than that measured for polymeric HA (~20 μM) [32, 51], and more importantly, likely act selectively at CD44 without undesirable cytotoxicity.

Table 5-1. Inhibitory Activity of THIQs for GBM Cell Survival

Structure	No	EC$_{50}$ (µM)		IC$_{50}$ (mM)
		CD44$^{+/+}$	*CD44$^{-/-}$*	*h*HABD
	3	> 2000	> 2000	6.7
	5a	222	652	3.5
	6a	> 2000	> 2000	3.9
	8	42	38	0.95
	9	880	1050	4.9
	13	18	38	1.1
	15a	566	667	3.4

EC$_{50}$s shown were calculated by averaging absorbance at 490 nm in six replicates, subtracting the absorbance of cell-free equivalents and normalized to vehicle controls.

Bibiography

1. Lennon, F. E.; Singleton, P. A. Hyaluronan regulation of vascular integrity. *American journal of cardiovascular disease* **2011,** 1, 200-213.

2. Toole, B. P. Hyaluronan-CD44 Interactions in Cancer: Paradoxes and Possibilities. *Clinical Cancer Research* **2009,** 15, 7462-7468.

3. Jiang, D.; Liang, J.; Noble, P. W. Hyaluronan as an immune regulator in human diseases. *Physiological reviews* **2011,** 91, 221-264.

4. Toole, B. P. Hyaluronan: from extracellular glue to pericellular cue. *Nature reviews. Cancer* **2004,** 4, 528-539.

5. Watanabe, K.; Yamaguchi, Y. Molecular identification of a putative human hyaluronan synthase. *The Journal of biological chemistry* **1996,** 271, 22945-22948.

6. Slevin, M.; Kumar, S.; Gaffney, J. Angiogenic oligosaccharides of hyaluronan induce multiple signaling pathways affecting vascular endothelial cell mitogenic and wound healing responses. *The Journal of biological chemistry* **2002,** 277, 41046-41059.

7. McKee, C. M.; Penno, M. B.; Cowman, M.; Burdick, M. D.; Strieter, R. M.; Bao, C.; Noble, P. W. Hyaluronan (HA) fragments induce chemokine gene expression in alveolar macrophages - The role of HA size and CD44. *Journal of Clinical Investigation* **1996,** 98, 2403-2413.

8. Ponta, H.; Sherman, L.; Herrlich, P. A. CD44: from adhesion molecules to signalling regulators. *Nature reviews. Molecular cell biology* **2003,** 4, 33-45.

9. Geiger, B.; Bershadsky, A.; Pankov, R.; Yamada, K. M. Transmembrane crosstalk between the extracellular matrix and the cytoskeleton. *Nature reviews. Molecular cell biology* **2001,** 2, 793-805.

10. Hansson, G. K.; Libby, P. The immune response in atherosclerosis: a double-edged sword. *Nature reviews. Immunology* **2006,** 6, 508-519.

11. Pandey, M. S.; Weigel, P. H. Hyaluronic acid receptor for endocytosis (HARE)-mediated endocytosis of hyaluronan, heparin, dermatan sulfate, and acetylated low density lipoprotein (AcLDL), but not chondroitin sulfate types A, C, D, or E, activates NF-κB-regulated gene expression. *The Journal of biological chemistry* **2014,** 289, 1756-1767.

12. Slomiany, M. G.; Dai, L.; Tolliver, L. B.; Grass, G. D.; Zeng, Y.; Toole, B. P. Inhibition of Functional Hyaluronan-CD44 Interactions in CD133-positive Primary Human Ovarian Carcinoma Cells by Small Hyaluronan Oligosaccharides. *Clinical Cancer Research* **2009,** 15, 7593-7601.

13. Lebel, L.; Smith, L.; Risberg, B.; Laurent, T. C.; Gerdin, B. Increased lymphatic elimination of interstitial hyaluronan during E. coli sepsis in sheep. *The American journal of physiology* **1989,** 256, H1524-31.

14. Takahashi, Y.; Li, L.; Kamiryo, M.; Asteriou, T.; Moustakas, A.; Yamashita, H.; Heldin, P. Hyaluronan fragments induce endothelial cell differentiation in a CD44- and CXCL1/GRO1-dependent manner. *The Journal of biological chemistry* **2005,** 280, 24195-24204.

15. Matou-Nasri, S.; Gaffney, J.; Kumar, S.; Slevin, M. Oligosaccharides of hyaluronan induce angiogenesis through distinct CD44 and RHAMM-mediated signalling pathways involving Cdc2 and gamma-adducin. *International journal of oncology* **2009,** 35, 761-773.

16. Sugahara, K. N.; Murai, T.; Nishinakamura, H.; Kawashima, H.; Saya, H.; Miyasaka, M. Hyaluronan oligosaccharides induce CD44 cleavage and promote cell migration in CD44-expressing tumor cells. *The Journal of biological chemistry* **2003,** 278, 32259-32265.

17. Sugahara, K. N.; Hirata, T.; Hayasaka, H.; Stern, R.; Murai, T.; Miyasaka, M. Tumor cells enhance their own CD44 cleavage and motility by generating hyaluronan fragments. *The Journal of biological chemistry* **2006,** 281, 5861-5868.

18. Gilg, A. G.; Tye, S. L.; Tolliver, L. B.; Wheeler, W. G.; Visconti, R. P.; Duncan, J. D.; Kostova, F. V.; Bolds, L. N.; Toole, B. P.; Maria, B. L. Targeting hyaluronan interactions in malignant gliomas and their drug-resistant multipotent progenitors. *Clinical Cancer Research* **2008,** 14, 1804-1813.

19. Day, A. J. Hyaluronan-binding Proteins: Tying Up the Giant. *Journal of Biological Chemistry* **2001,** 277, 4585-4588.

20. Gao, F.; Yang, C. X.; Mo, W.; Liu, Y. W.; He, Y. Q. Hyaluronan oligosaccharides are potential stimulators to angiogenesis via RHAMM mediated signal pathway in wound healing. *Clinical & Investigative Medicine* **2008,** 31, E106-16.

21. Campo, G. M.; Avenoso, A.; Campo, S.; D'Ascola, A.; Nastasi, G.; Calatroni, A. Small hyaluronan oligosaccharides induce inflammation by engaging

both toll-like-4 and CD44 receptors in human chondrocytes. *Biochemical Pharmacology* **2010,** 80, 480-490.

22. Stern, R.; Asari, A. A.; Sugahara, K. N. Hyaluronan fragments: an information-rich system. *European journal of cell biology* **2006,** 85, 699-715.

23. Kohda, D.; Morton, C. J.; Parkar, A. A.; Hatanaka, H.; Inagaki, F. M.; Campbell, I. D.; Day, A. J. Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell* **1996,** 86, 767-775.

24. Jackson, D. G. Immunological functions of hyaluronan and its receptors in the lymphatics. *Immunological reviews* **2009,** 230, 216-231.

25. Mahoney, D. J.; Blundell, C. D.; Day, A. J. Mapping the hyaluronan-binding site on the link module from human tumor necrosis factor-stimulated gene-6 by site-directed mutagenesis. *The Journal of biological chemistry* **2001,** 276, 22764-22771.

26. Baranova, N. S.; Nilebäck, E.; Haller, F. M.; Briggs, D. C.; Svedhem, S.; Day, A. J.; Richter, R. P. The inflammation-associated protein TSG-6 cross-links hyaluronan via hyaluronan-induced TSG-6 oligomers. *The Journal of biological chemistry* **2011,** 286, 25675-25686.

27. YANG, B.; YANG, B.; SAVANI, R.; Turley, E. Identification of a Common Hyaluronan-Binding Motif in the Hyaluronan-Binding Proteins Rhamm, Cd44 and Link Protein. *Embo Journal* **1994,** 13, 286-296.

28. Zöller, M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nature reviews. Cancer* **2011,** 11, 254-267.

29. Banerji, S.; Day, A. J.; Kahmann, J. D.; Jackson, D. G. Characterization of a functional hyaluronan-binding domain from the human CD44 molecule expressed in Escherichia coli. *Protein expression and purification* **1998,** 14, 371-381.

30. Adedeji, A. O.; Severson, W.; Jonsson, C.; Singh, K.; Weiss, S. R.; Sarafianos, S. G. Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms. *Journal of virology* **2013,** 87, 8017-8028.

31. Teriete, P.; Banerji, S.; Noble, M.; Blundell, C. D.; Wright, A. J.; Pickford, A. R.; Lowe, E.; Mahoney, D. J.; Tammi, M. I.; Kahmann, J. D.; Campbell, I. D.; Day, A. J.; Jackson, D. G. Structure of the regulatory hyaluronan binding domain in the inflammatory leukocyte homing receptor CD44. *Molecular Cell* **2004,** 13, 483-496.

32. Banerji, S.; Wright, A. J.; Noble, M.; Mahoney, D. J.; Campbell, I. D.; Day, A. J.; Jackson, D. G. Structures of the Cd44-hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nature structural & molecular biology* **2007,** 14, 234-239.

33. Underhill, C. B.; Chi-Rosso, G.; Toole, B. P. Effects of detergent solubilization on the hyaluronate-binding protein from membranes of simian virus 40-transformed 3T3 cells. *Journal of Biological Chemistry* **1983**, 258, 8086-8091.

34. Lesley, J; Hascall; V. C.; Tammi, M; Hyman, R. Hyaluronan Binding by Cell Surface CD44. *Journal of Biological Chemistry* **2000**, 275, 26967-26975.

35. Ghatak, S.; Misra, S.; Toole, B. P. Hyaluronan oligosaccharides inhibit anchorage-independent growth of tumor cells by suppressing the phosphoinositide 3-

kinase/Akt cell survival pathway. *The Journal of biological chemistry* **2002,** 277, 38013-38020.

36. Yang, C.; Cao, M.; Liu, H.; He, Y.; Xu, J.; Du, Y.; Liu, Y.; Wang, W.; Cui, L.; Hu, J.; Gao, F. The High and Low Molecular Weight Forms of Hyaluronan Have Distinct Effects on CD44 Clustering. *The Journal of biological chemistry* **2012,** 287, 43094-43107.

37. Qin, Z.; Dai, L.; Bratoeva, M.; Slomiany, M. G.; Toole, B. P.; Parsons, C. Cooperative roles for emmprin and LYVE-1 in the regulation of chemoresistance for primary effusion lymphoma. *Leukemia* **2011,** 25, 1598-1609.

38. Lüke, H. J.; Prehm, P. Synthesis and shedding of hyaluronan from plasma membranes of human fibroblasts and metastatic and non-metastatic melanoma cells. *The Biochemical journal* **1999,** 343 Pt 1, 71-75.

39. Ye, L.; Mora, R.; Akhayani, N.; Haudenschild, C. C.; Liau, G. Growth factor and cytokine-regulated hyaluronan-binding protein TSG-6 is localized to the injury-induced rat neointima and confers enhanced growth in vascular smooth muscle cells. *Circulation research* **1997,** 81, 289-296.

40. Toole, B. P.; Wight, T. N.; Tammi, M. I. Hyaluronan-cell interactions in cancer and vascular disease. *The Journal of biological chemistry* **2002,** 277, 4593-4596.

41. Krettek, A.; Sjöberg, S. CD44 - a New Cardiovascular Drug Target or Merely an Innocent Bystander? *Cardiovascular & Haematological Disorders-Drug Targets* **2011,** 9, 293-302.

42. Kashima, Y.; Takahashi, M.; Shiba, Y.; Itano, N.; Izawa, A.; Koyama, J.; Nakayama, J.; Taniguchi, S. a. i.; Kimata, K.; Ikeda, U. Crucial role of hyaluronan in neointimal formation after vascular injury. *PloS one* **2013,** 8, e58760.

43. Zeidler, A.; Bräuer, R.; Thoss, K.; Bahnsen, J.; Heinrichs, V.; Jablonski-Westrich, D.; Wroblewski, M.; Rebstock, S.; Hamann, A. Therapeutic effects of antibodies against adhesion molecules in murine collagen type II-induced arthritis. *Autoimmunity* **1995,** 21, 245-252.

44. Mikecz, K.; Brennan, F. R.; Kim, J. H.; Glant, T. T. Anti-CD44 treatment abrogates tissue oedema and leukocyte infiltration in murine arthritis. *Nature medicine* **1995,** 1, 558-563.

45. Rupp, U.; Schoendorf-Holland, E.; Eichbaum, M.; Schuetz, F.; Lauschner, I.; Schmidt, P.; Staab, A.; Hanft, G.; Huober, J.; Sinn, H.-P.; Sohn, C.; Schneeweiss, A. Safety and pharmacokinetics of bivatuzumab mertansine in patients with CD44v6-positive metastatic breast cancer: final results of a phase I study. *Anti-cancer drugs* **2007,** 18, 477-485.

46. Riechelmann, H.; Sauter, A.; Golze, W.; Hanft, G.; Schroen, C.; Hoermann, K.; Erhardt, T.; Gronau, S. Phase I trial with the CD44v6-targeting immunoconjugate bivatuzumab mertansine in head and neck squamous cell carcinoma. *Oral oncology* **2008,** 44, 823-829.

47. Hirota-Takahata, Y.; Harada, H.; Tanaka, I.; Nakata, T.; Nakajima, M.; Takahashi, M. F-19848 A, a Novel Inhibitor of Hyaluronic Acid Binding to Cellular Receptor CD44. *The Journal of Antibiotics* **2007,** 60, 633-639.

48. Bajorath, J.; Greenfield, B.; Munro, S. B.; Day, A. J.; Aruffo, A. Identification of CD44 residues important for hyaluronan binding and delineation of the binding site. *The Journal of biological chemistry* **1998,** 273, 338-343.

49. Jana, M.; Bandyopadhyay, S. Conformational flexibility of a protein-carbohydrate complex and the structure and ordering of surrounding water. *Physical chemistry chemical physics : PCCP* **2012,** 14, 6621-6631.

50. Peach, R. J.; Hollenbaugh, D.; Stamenkovic, I.; Aruffo, A. Identification of hyaluronic acid binding sites in the extracellular domain of CD44. *The Journal of cell biology* **1993,** 122, 257-264.

51. Ogino, S.; Nishida, N.; Umemoto, R.; Suzuki, M.; Takeda, M.; Terasawa, H.; Kitayama, J.; Matsumoto, M.; Hayasaka, H.; Miyasaka, M.; Shimada, I. Two-State Conformations in the Hyaluronan-Binding Domain Regulate CD44 Adhesiveness under Flow Condition. *Structure* **2010,** 18, 649-656.

52. Takeda, M.; Terasawa, H.; Sakakura, M.; Yamaguchi, Y.; Kajiwara, M.; Kawashima, H.; Miyasaka, M.; Shimada, I. Hyaluronan recognition mode of CD44 revealed by cross-saturation and chemical shift perturbation experiments. *The Journal of biological chemistry* **2003,** 278, 43550-43555.

53. Takeda, M.; Ogino, S.; Umemoto, R.; Sakakura, M.; Kajiwara, M.; Sugahara, K. N.; Hayasaka, H.; Miyasaka, M.; Terasawa, H.; Shimada, I. Ligand-induced structural changes of the CD44 hyaluronan-binding domain revealed by NMR. *The Journal of biological chemistry* **2006,** 281, 40089-40095.

54. Alex, A. A.; Flocco, M. M. Fragment-based drug discovery: what has it achieved so far? *Current topics in medicinal chemistry* **2007,** 7, 1544-1567.

55. Coyne, A. G.; Scott, D. E.; Abell, C. Drugging challenging targets using fragment-based approaches. *Current opinion in chemical biology* **2010,** 14, 299-307.

56. Scott, D. E.; Ehebauer, M. T.; Pukala, T.; Marsh, M.; Blundell, T. L.; Venkitaraman, A. R.; Abell, C.; Hyvönen, M. Using a fragment-based approach to target protein-protein interactions. *Chembiochem : a European journal of chemical biology* **2013,** 14, 332-342.

57. Elinder, M.; Geitmann, M.; Gossas, T.; Kallblad, P.; Winquist, J.; Nordstrom, H.; Hamalainen, M.; Danielson, U. H. Experimental Validation of a Fragment Library for Lead Discovery Using SPR Biosensor Technology. *Journal of Biomolecular Screening* **2011,** 16, 15-25.

58. Navratilova, I.; Hopkins, A. L. Fragment Screening by Surface Plasmon Resonance. *ACS Medicinal Chemistry Letters* **2010**, 1, 44-48.

59. Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2007,** 2, 2212-2221.

60. Giannetti, A. M. From experimental design to validated hits a comprehensive walk-through of fragment lead identification using surface plasmon resonance. *Methods in enzymology* **2011,** 493, 169-218.

61. Stols, L.; Zhou, M.; Eschenfeldt, W. H.; Millard, C. S.; Abdullah, J.; Collart, F. R.; Kim, Y.; Donnelly, M. I. New vectors for co-expression of proteins: structure of

Bacillus subtilis ScoAB obtained by high-throughput protocols. *Protein expression and purification* **2007,** 53, 396-403.

62.   Kabsch, W. XDS. *Acta crystallographica. Section D, Biological crystallography* **2010,** 66, 125-132.

63.   Evans, P. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography* **2005,** 62, 72-82.

64.   McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phasercrystallographic software. *Journal of Applied Crystallography* **2007,** 40, 658-674.

65.   Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **2010,** 66, 486-501.

66.   Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **2011,** 67, 355-367.

67.   Lebedev, A. A.; Young, P.; Isupov, M. N.; Moroz, O. V.; Vagin, A. A.; Murshudov, G. N. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta crystallographica. Section D, Biological crystallography* **2012,** 68, 431-440.

68.   Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-

atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **2010,** 66, 12-21.

69. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic acids research* **2000,** 28, 235-242.

70. Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *Journal of Biomolecular Screening* **2001,** 6, 429-440.

71. Ziebell, M. R.; Zhao, Z. G.; Luo, B.; Luo, Y.; Turley, E. A.; Prestwich, G. D. Peptides that mimic glycosaminoglycans: high-affinity ligands for a hyaluronan binding domain. *Chemistry & Biology* **2001,** 8, 1081-1094.

72. Piotrowicz, R. S.; Damaj, B. B.; Hachicha, M.; Incardona, F.; Howell, S. B.; Finlayson, M. A6 Peptide Activates CD44 Adhesive Activity, Induces FAK and MEK Phosphorylation, and Inhibits the Migration and Metastasis of CD44-Expressing Cells. *Molecular cancer therapeutics* **2011,** 10, 2072-2082.

73. Hibino, S.; Shibuya, M.; Engbring, J. A.; Mochizuki, M.; Nomizu, M.; Kleinman, H. K. Identification of an active site on the laminin alpha5 chain globular domain that binds to CD44 and inhibits malignancy. *Cancer research* **2004,** 64, 4810-4816.

74. Ugarte-Berzal, E.; Bailón, E.; Amigo-Jiménez, I.; Albar, J. P.; García-Marco, J. A.; García-Pardo, A. A Novel CD44-binding Peptide from the pro-Matrix Metalloproteinase-9 Hemopexin Domain Impairs Adhesion and Migration of
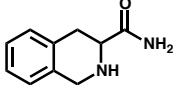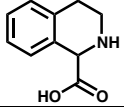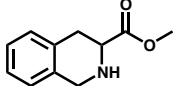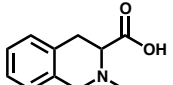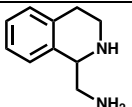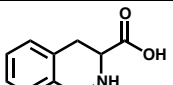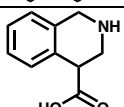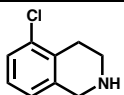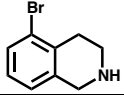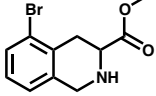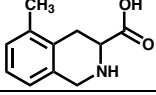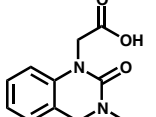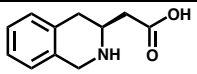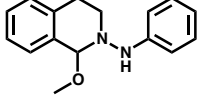
Chronic Lymphocytic Leukemia (CLL) Cells. *The Journal of biological chemistry* **2014**, 289, 15340-15349.

75.   Mishima, K.; Mazar, A. P.; Gown, A.; Skelly, M.; Ji, X. D.; Wang, X. D.; Jones, T. R.; Cavenee, W. K.; Huang, H. J. A peptide derived from the non-receptor-binding region of urokinase plasminogen activator inhibits glioblastoma growth and angiogenesis in vivo in combination with cisplatin. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** 97, 8484-8489.

76.   Ghamande, S. A.; Silverman, M. H.; Huh, W.; Behbakht, K.; Ball, G.; Cuasay, L.; Würtz, S. O.; Brunner, N.; Gold, M. A. A phase 2, randomized, double-blind, placebo-controlled trial of clinical activity and safety of subcutaneous A6 in women with asymptomatic CA125 progression after first-line chemotherapy of epithelial ovarian cancer. *Gynecologic oncology* **2008,** 111, 89-94.

77.   Park, H.-Y.; Lee, K.-J.; Lee, S.-J.; Yoon, M.-Y. Screening of peptides bound to breast cancer stem cell specific surface marker CD44 by phage display. *Molecular biotechnology* **2012,** 51, 212-220.

78.   Siegal, G.; Ab, E.; Schultz, J. Integration of fragment screening and library design. *Drug discovery today* **2007,** 12, 1032-1039.

79.   Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical biochemistry* **2006,** 357, 289-298.

80. Lavinder, J. J.; Hari, S. B.; Sullivan, B. J.; Magliery, T. J. High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *Journal of the American Chemical Society* **2009,** 131, 3794-3795.

81. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug discovery today* **2003,** 8, 876-877.

82. Hämäläinen, M. D.; Zhukov, A.; Ivarsson, M.; Fex, T.; Gottfries, J.; Karlsson, R.; Björsne, M. Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. *Journal of Biomolecular Screening* **2008,** 13, 202-209.

83. Larsson, A.; Jansson, A.; Åberg, A.; Nordlund, P. Efficiency of hit generation and structural characterization in fragment-based ligand discovery. *Current opinion in chemical biology* **2011,** 15, 482-488.

84. Hubbard, R. E.; Murray, J. B. Experiences in fragment-based lead discovery. *Methods in enzymology* **2011,** 493, 509-531.

85. Rich, R. L.; Myszka, D. G. Kinetic analysis and fragment screening with Fujifilm AP-3000. *Analytical biochemistry* **2010,** 402, 170-178.

86. Davis, B. J.; Erlanson, D. A. Learning from our mistakes: the 'unknown knowns' in fragment screening. *Bioorganic & Medicinal Chemistry Letters* **2013,** 23, 2844-2852.

87. Tan, K. P.; Nguyen, T. B.; Patel, S.; Varadarajan, R.; Madhusudhan, M. S. Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-
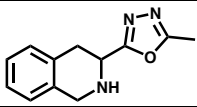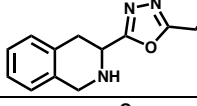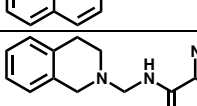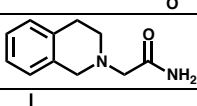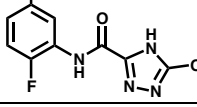
binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic acids research* **2013,** 41, W314-21.
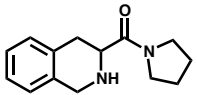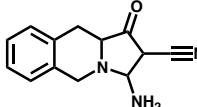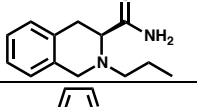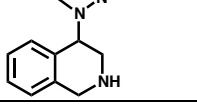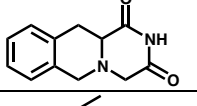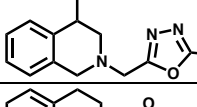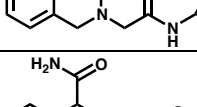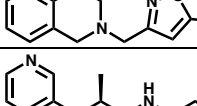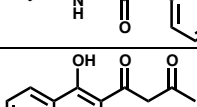
88.   Higman, V. A.; Blundell, C. D.; Mahoney, D. J.; Redfield, C.; Noble, M. E. M.; Day, A. J. Plasticity of the TSG-6 HA-binding Loop and Mobility in the TSG-6-HA Complex Revealed by NMR and X-ray Crystallography. *Journal of molecular biology* **2007,** 371, 669-684.

89.   Veiseh, M.; Kwon, D. H.; Borowsky, A. D.; Tolg, C.; Leong, H. S.; Lewis, J. D.; Turley, E. A.; Bissell, M. J. Cellular heterogeneity profiling by hyaluronan probes reveals an invasive but slow-growing breast tumor subset. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** 111, E1731-9.

90.   Tolg, C.; Hamilton, S. R.; Zalinska, E.; McCulloch, L.; Amin, R.; Akentieva, N.; Winnik, F.; Savani, R.; Bagli, D. J.; Luyt, L. G.; Cowman, M. K.; McCarthy, J. B.; Turley, E. A. A RHAMM mimetic peptide blocks hyaluronan signaling and reduces inflammation and fibrogenesis in excisional skin wounds. *The American journal of pathology* **2012,** 181, 1250-1270.

91.   Chetty, C.; Vanamala, S. K.; Gondi, C. S.; Dinh, D. H.; Gujrati, M.; Rao, J. S. MMP-9 induces CD44 cleavage and CD44 mediated cell migration in glioblastoma xenograft cells. *Cellular signalling* **2012,** 24, 549-559.

92.   Bullard, K. M.; Kim, H.-R.; Wheeler, M. A.; Wilson, C. M.; Neudauer, C. L.; Simpson, M. A.; McCarthy, J. B. Hyaluronan synthase-3 is upregulated in metastatic colon carcinoma cells and manipulation of expression alters matrix retention and cellular growth. *International Journal of Cancer* **2003,** 107, 739-746.
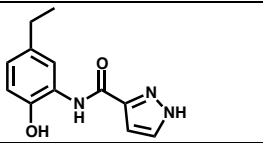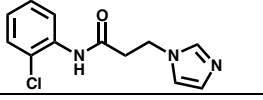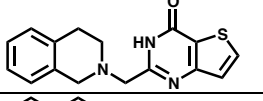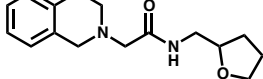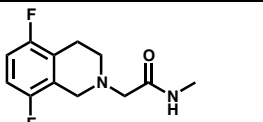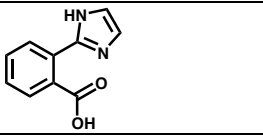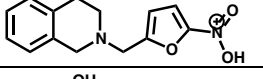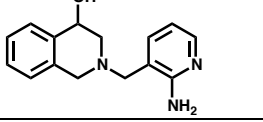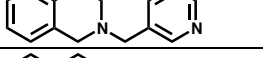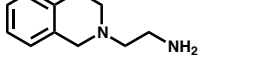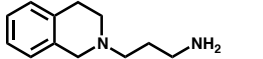
93. Yang, J.; Price, M. A.; Li, G. Y.; Bar-Eli, M.; Salgia, R.; Jagedeeswaran, R.; Carlson, J. H.; Ferrone, S.; Turley, E. A.; McCarthy, J. B. Melanoma proteoglycan modifies gene expression to stimulate tumor cell motility, growth, and epithelial-to-mesenchymal transition. *Cancer research* **2009,** 69, 7538-7547.

94. Tolg, C.; Hamilton, S. R.; Morningstar, L.; Zhang, J.; Zhang, S.; Esguerra, K. V.; Telmer, P. G.; Luyt, L. G.; Harrison, R.; McCarthy, J. B.; Turley, E. A. RHAMM promotes interphase microtubule instability and mitotic spindle integrity through MEK1/ERK1/2 activity. *Journal of Biological Chemistry* **2010,** 285, 26461-26474.

95. Ivanov, A. I. Pharmacological inhibition of endocytic pathways: is it specific enough to be useful? *Methods in molecular biology (Clifton, N.J.)* **2008,** 440, 15-33.

96. Racine, R.; Mummert, M. E. Hyaluronan Endocytosis: Mechanisms of Uptake and Biological Functions. *Molecular Regulation of Endocytosis* **2012**, Chapter 14.

97. Ouasti, S.; Kingham, P. J.; Terenghi, G.; Tirelli, N. The CD44/integrins interplay and the significance of receptor binding and re-presentation in the uptake of RGD-functionalized hyaluronic acid. *Biomaterials* **2012**, 33, 1120-1134.

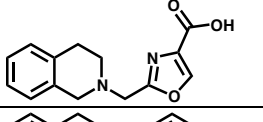98. Simpson, M. A.; Wilson, C. M.; Furcht, L. T.; Spicer, A. P.; Oegema, T. R.; McCarthy, J. B. Manipulation of hyaluronan synthase expression in prostate adenocarcinoma cells alters pericellular matrix retention and adhesion to bone marrow endothelial cells. *The Journal of biological chemistry* **2002,** 277, 10050-10057.

Appendix A. Index of the THIQ analogs tested

Catalog Number	Structure	Thermal shift (°C)	Affinity (mM)	Source
4027538		0	6.6	Cambridge
5926716		0.5	n.b.	Cambridge
8924006		0.5	n.b.	Cambridge
4400181		0.5	n.b.	Cambridge
5547653		0	n.b.	Cambridge
CDS005312		−0.75	2.4	Sigma
CDS012933		−0.5	5.6	Sigma
4002741		−1	7.8	Cambridge
CDS015572		−1	n.b.	Sigma
CDS009155		−0.5	n.b.	Sigma
CDS008658		0	n.b.	Sigma
9039842		0	n.b.	Cambridge
CDS008707			11.2	Sigma
S807389			2.8	Sigma

Z10369630069		−20	4.0	Enamine
Z10373739821 20		0	2.0	Enamine
Z31630307237		0	2.3	Enamine
Z33606633419		−0.5	13.5	Enamine
Z117165043149		−0.25	2.7	Enamine
Z10369630296		−0.75	n.b.	Enamine
Z10369654239		0	n.b.	Enamine
Z295465724		−0.5	9.9	Enamine
Z1147449695190		0	3.0	Enamine
Z5690688768		−0.75	3.0	Enamine
Z1269114463118		−0.5	4.9	Enamine
Z103695912244		−0.25	n.b.	Enamine
EN300-00696112		−0.75	2.9	Enamine
Z104377982			13.8	Enamine
Z228586634			n.b.	Enamine

EN300-26958			14	Enamine
Z255150250			10.4	Enamine
Z607864950			5.9	Enamine
Z607864952			6.7	Enamine
Z644181556			15.6	Enamine
EN300-80596			10.6	Enamine
Z281880906			13.3	Enamine
Z1205404571			9.6	Enamine
Z46161999			9.9	Enamine
Z1152150465			14	Enamine
Z1182683606			5.7	Enamine
Z17871331			n.b.	Enamine
Z56793490			n.b.	Enamine
Z54752169			n.b.	Enamine
Z1385772080			n.b.	Enamine

Z1317184614			n.b.	Enamine
Z234887794			n.b.	Enamine
Z44491159			n.b.	Enamine
Z55130392			n.b.	Enamine
Z1373561797			n.b.	Enamine
4002463			n.b.	Cambridge
5268075			n.b.	Cambridge
34220941			n.b.	Cambridge
26391192			n.b.	Cambridge
5260352			n.b.	Cambridge
4003442			n.b.	Cambridge
Z271005928			n.b.	Enamine

n.b. = No binding

Appendix B. Crystallographic refinement status of THIQ analogs

Ligand	Reso	R/R$_{free}$	Date-Site-Puck-Sample	FXID	PDB
D1	1.60	0.22/0.24	20111215-GMCA-070B-10	0712	
D2	1.80	0.22/0.24	20111105-GMCA-031R-02	0713	
D5	1.84	0.25/0.28	20120226-GMCA-031Y-06	0714	
D17	1.84	0.26/0.30	20120226-GMCA-031Y-09	0715	
D18	2.46	0.26/0.27	20120226-GMCA-031B-10	0716	
S4	1.86	0.22/0.26	20121216-ALS-070Y-08	0717	
S5	1.72	0.23/0.26	20121216-ALS-070R-06	0718	
3	1.69	0.16/0.20	20121216-ALS-070Y-05	0727	4MRG
5	1.75	0.17/0.21	20130420-APS-031Y-11	0728	4NP2
5a	1.61	0.17/0.22	20130710-ALS-031Y-03	0729	4NP3
5b	1.40	0.18/0.20	20131215-IMCA-070R-01	0070	5BZJ
6	1.40	0.18/0.21	20130316-IMCA-070P-01	0719	5BZK
6a	1.23	0.17/0.18	20140315-IMCA-070Y-04	0285	5BZL
7	1.32	0.17/0.20	20140315-IMCA-070G-11	0304	5BZI
7a	2.23	0.21/0.25	20130129-ALS-031R-02	0720	
7b	2.14	0.22/0.26	20130129-ALS-070P-11	0721	
7e	1.85	0.21/0.24	20130420-IMCA-031Y-05	0722	
7f	1.30	0.17/0.21	20130420-APS-031Y-11	0723	
7g	1.95	0.18/0.22	20130129-ALS-070R-11	0724	5BZC
7h	1.31	0.16/0.18	20130316-IMCA-070B-11	0725	5BZE
7i	2.70	0.20/0.28	20130129-ALS-031Y-06	0726	5BZF
7j	1.39	0.23/0.24	20130316-IMCA-070R-06	0207	
7l	1.30	0.24/0.26	20130420-IMCA-031Y-07	0208	
7m	1.35	0.25/0.27	20130420-IMCA-070R-07	0209	
7n	2.81	0.19/0.27	20130129-ALS-070R-07	0731	
7o	2.19	0.19/0.24	20130129-ALS-070P-03	0732	5BZG
7q	1.95	0.21/0.25	20131112-in-house	0733	
7r	1.95	0.21/0.26	20131112-in-house	0734	5BZH
8	1.64	0.20/0.24	20130710-ALS-031Y-02	0730	
8a	1.23	0.19/0.22	20140315-IMCA-070Y-07	0288	5BZM
9	1.23	0.17/0.19	20131215-IMCA-070G-03	0096	5BZN
10	1.22	0.16/0.18	20131215-IMCA-070G-06	0099	5BZO
11	1.23	0.18/0.20	20131215-IMCA-070G-11	0104	5BZP
12	1.20	0.16/0.18	20131215-IMCA-070R-06	0075	5BZQ
15a	1.15	0.17/0.18	20141201-IMCA-031G-07	0597	5BZR
15b	1.50	0.21/0.24	20141208-IMCA-070B-01	0592	5BZS
16	1.50	0.24/0.26	20130710-ALS-031G-03	0001	
16g	1.23	0.22/0.23	20140621-IMCA-031B-03	0378	
B1	1.24	0.17/0.19	20150418-IMCA-070G-05	0704	5BZT