Technische Universität München

Max-Planck-Institut für Biochemie

Abteilung Strukturforschung

Biologische NMR-Arbeitsgruppe

Biochemical and biophysical characterization of CD44 and its binding partner, hyaluronic acid

and

structural investigations of the ubiquitin-like protein 5

Kaja M. Kowalska

Publications

Bista M, **Kowalska K**, Janczyk W, Dömling A, Holak TA

Roboust NMR screening for lead compounds using tryptophan containing proteins.

Journal of American Chemical Society. 2009 Jun 10; 131:7500-1

Rothweiler U, Czarna A, Weber L, Popowicz GM, Brongel K, **Kowalska K**, Orth M, Stemman O, Holak TA

NMR screening for lead compounds using tryptophan-mutated proteins.

Journal of Medicinal Chemistry. 2008 Aug 28; 51:5035-42

Bista M, Wolf S, Castro M, **Kowalska K**, Dömling A, Holak TA, Popowicz GM

Employing transient protein states in designing inhibitors of the p53-MDM2 interaction.

Proceedings of the National Academy of Science of the United States of America – submitted

Kowalska K, Beck B, Czarna A, Wang W, Day B, Dömling A, Holak TA

SAR of new Imidazoline Mdm2 antagonists.

Angewandte Chemie – manuscript in preparation

Contents

1 Introduction

1.1 CD44, the hyaluronan receptor

1.1.1 CD44, introduction

The cell surface receptor CD44, first described in 1983 (Gallatin et al., 1983), belongs to a larger group of hyaluronan-binding proteins, termed the hyaladherins. Many of these proteins contain a sequence of amino acid homolgous to B loop of cartilage link protein, called Link module, which is important in the binding of hyaluronic acid (HA) (Goetinck et al., 1987). CD44 glycoprotein is a multifunctional and multistructural cell surface molecule involved in cell proliferation, cell migration, cell differentiation, presentation of cytokines, chemokines and growth factors to the corresponding receptors, angiogenesis, and docking of proteases at the cell membrane, as well as signaling for cell survival. All these biological functions are essential to the physiological activities of normal cells, but they are also closely connected with the pathologic activities of cancer cells (Naor et al., 2002). Through alternative splicing, the CD44 is expressed in a wide variety of isoforms in many cells, and some of these variants induce a metastatic phenotype in locally growing tumor cells. There is ample evidence for the importance of CD44 expression in the progression of many tumors, as well as for its presence on cancer-initiating cells (CICs, also known as cancer stem cells) (Zoller, 2011).

1.1.2 CD44 structure and distribution

The activity of adhesion proteins such as integrins, selectins, addressins, members of the immunoglobulin superfamily, cadherins and CD44 determinate the interactions of cells with other cells or with components of the extracellular matrix (ECM), as well as their locomotion on blood vessel and extravascular tissues (Naor et al., 2002).

CD44 was cloned in 1989 and identified as a member of the cartilage link protein superfamily (Goldstein et al., 1989; Stamenkovic et al., 1989). CD44 is a single chain glycoprotein derived from a single copy of a gene located on the short arm of the human chromosome 11 and on chromosome 2 in mice, spanning ~50 kb of genomic DNA. The protein is an acidic molecule (isoelectric point = 4.2 to 5.8), its charge mainly due to sialic acid. The $t_{1/2}$ of the CD44 turnover is estimated as 8 h (Naor et al., 2002). CD44 has seven extracellular domains, a transmembrane domain and a cytoplasmic domain. The protein varies in size due to O-glycosylation and N-glycosylation and the insertion of alternatively spliced products in the extracellular domains of the molecule (Goldstein et al., 1989). The first five amino-terminal exons of CD44 are constant and encode a globular extracellular domain, the next 10 exons (6 to 15) are variant, that is up to ten variant exon products can be inserted by alternative splicing between domains 5 and 16. Exons 17-18 are again variant and exons 19 to 20 are constant (Figure 1.1.2.1).
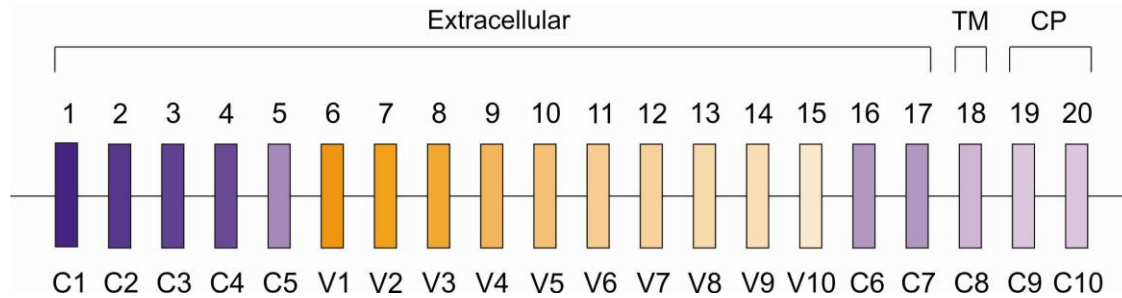
Figure 1.1.2.1 CD44 consists of several exons. The violet bars represent constant exons, which are used in every CD44 mRNA. The yellow bars represent exons that can be inserted by alternative splicing, resulting in the generation of the variable region. Note that exon v1 is not expressed in human CD44. TM, transmembrane-encoding exon; CP, cytoplasmic tail-encoding exon (Adapted from Zoller, 2011).

Exons 16 and 17 encode the C terminus part of the extracellular domain, exon 18 encodes the transmembrane domain, and exons 19 or 20 differently utilized by alternative splicing and resulting in either a short (three amino acids) cytoplasmic tail or the more abundant long one (72 amino acids). Mouse CD44 contains 20 exons, but the human one does not contain exon v1. The utilization of the 10 different exons (v1-v10) generates multiple CD44 variants with different combinations of variant exon products. Theoretically, almost 800 membrane-bound CD44 isoforms can be generated, although not all of them are expressed. To date, dozens of different isoforms have been identified; the most common one is a standard form of CD44, in which exon 5 is spliced directly to exon 16, skipping the entire variant sequences (Naor et al., 2002) . Examples of CD44 isoforms are presented in Figure 1.1.2.2.
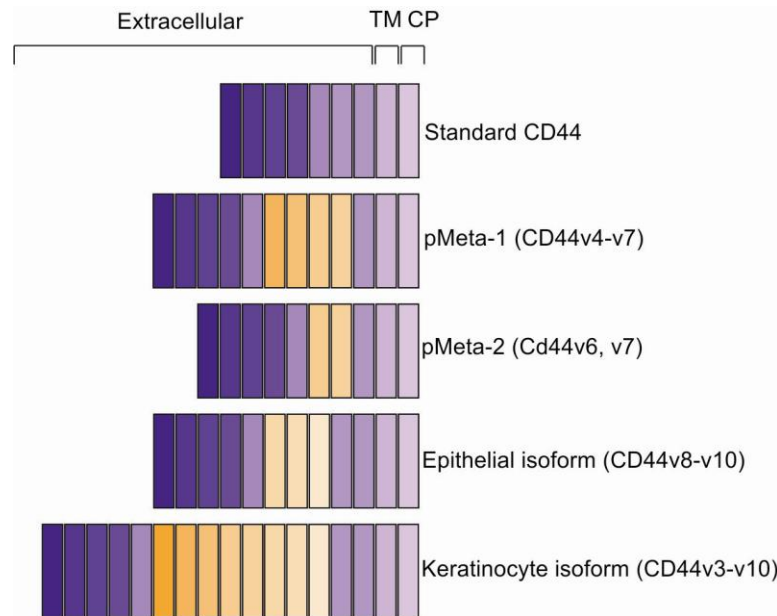
Figure 1.1.2.2 Examples of alternatively spliced CD44 proteins (Adapted from Zeller, 2011; Naor et al., 2002).

The structural variability of CD44 is further enriched by post-translational modifications. All isoforms are highly glicosylated, containing both N- and O-linked carbohydrate side chains (Brown et al., 1991), as well as glycosaminoglycan (GAG) attachments (Naor et al., 1997). Variations in the degree of glycosylation give rise to multiple molecular mass forms, resulting in ranges from ~80 kDa to ~250 kDa. Both alternative splicing and posttranslational modifications contribute to this diversity. Most of the N-linked glycosylation sites (Asn25, Asn57, Asn100, Asn110, and Asn120) are present within the N-terminal of the extracellular domain, and in the alternatively spliced variable regions. The potential O-linked glycosylation sites (Ser/Thr residues) and GAG attachments are located in the more carboxyl terminal

regions of the extracellular domain, including the membrane proximal region, and the variable fragment of the protein. Additionally, human CD44 cytoplasmic tail contains six potential phosphorylation sites (Ser) (Naor et al., 2002).

The CD44, in common with other HA binding proteins, interact with ligand via a Link module (Day and Prestwich, 2002). The Link domain is a conserved α/β-fold, approximately 100 residues in length. It comprises two antiparallel β-sheets, made from the total of six β-strands and two α-helices, stabilized by a pair highly conserved disulfide bridges, and is structurally related to the C-type lectin domain (Blundell et al., 2003; Kohda et al., 1996). However, in case of CD44, HA binding is more complex, involving the additional sequences outside the consensus Link domain just described. Some observations led to the hypothesis that CD44 contains an extended HA binding domain that is specially adapted for the regulation of ligand binding (Jackson, 2003; Jackson et al., 2001). The crystal structure of the HA-binding domain (HABD) shows that CD44 is enlarged by four additional β-strands, which contribute to flanking N- and C-terminal sequences. These extensions encode extra pair of cysteine residues, which form a third disulfide linkage (between Cys28 and Cys129) essential for stabilizing the HA binding domain (Teriete et al., 2004). X-ray and nuclear magnetic resonance studies (NMR) on three-dimensional structure of HABD demonstrated that the Link module and the N- and C-terminal extensions form a single folded unit, which is composed of three α-helices and eleven β-strands, arranged in the order of β0-β0'-β1-α1-α2-β3-β4-β5-β6-β7-β8-β9-α3 (Takeda et al., 2003; Teriete et al., 2004)

1.1.3 CD44 ligands and mechanism of HA-binding

The principal ligand of CD44 is hyaluronic acid, a linear polymer of repeating disaccharide units. However, CD44 can interact with several additional molecules such as collagen, fibrinogen, fibronectin, laminin, chondroitin sulfate, serglycin/gp600, mucosal vascular addressin, osteopontin, and the major histocompability complex class II invariant chain, as well as L-selectin and E-selectin. In many cases CD44 does not bind to its ligands unless activated by external stimuli (Naor et al., 2002). Three activation states of CD44 have been identified in cell lines and normal populations: active CD44, which constitutively binds HA; inducible CD44, which does not bind HA or binds weakly, unless activated by inducing monoclonal antibodies, cytokines, combination of cytokines, growth factors, or phorbol ester; and inactive CD44, which does not bind HA even in the presence of inducing agents (Blundell et al., 2003). The status of HA binding capacity of CD44 receptor is dictated by its N-glycosylation pattern, where active form is the least glycosylated, inactive CD44 the most glycosylated, and inducible CD44 holds the intermediate position (Lesley et al., 1995). The CD44 protein shows substantial specificity for recognition of HA over other glycosaminoglycans. Octamer is likely to be minimum size of HA that completely occupies the HA binding site in the nonglycosylated CD44 HA-binding domain (Teriete et al., 2004). HA binding induces an allosteric conformational rearrangement of the HABD, in which the C-terminal segment (T153 to Y169) becomes unstructured (Takeda et al., 2006). The crystal structure of the mouse CD44-HA_8 complex suggests that this binding specificity arises from molecular recognition of number of characteristics peculiar to

repeating polymer of GlcUA and GlcNAc (Figure 1.1.3.1A). The recognition of HA by CD44 appears to be driven by hydrogen bonds and the van der Waals forces. The lack of ionic and CH-π stacking interactions, although not unique, is unusual for this particular protein-carbohydrate interaction (Banerji et al., 2007). It distinguishes CD44 from the mode of binding observed in bacterial and bee-venom hyaluronidases (Markovic-Housley et al., 2000), as well as that predicted from models of the Link module in lymphatic hyaluronan receptor LYVE-1 (Jackson 2004) and the TSG-6 (Takeda et al., 2006). The absence of ionic interactions and the presence of a hydrophobic core in the CD44-HA complex are both desirable features in an interaction that might be targeted by structure-based designed screening.

CD44 HABD exchanges its conformation between ordered (O) and partially disordered (PD) states in either the absence or presence of the HA ligand, respectively. This interconversion between two distinct conformations has the exchange rate of hundreds of milliseconds. Under HA binding the equilibrium is shifted toward PD conformation exhibiting the disordered C-terminal segment, hence less than 10% of the total HABD exists in O state (Ogino et al., 2010). This equilibrium state of CD44 explains inconsistency regarding the conformation of the HABD in the HA-bound form raised by previous NMR and X-ray studies (Banerji et al., 2007; Takeda et al., 2006). Under physiological condition, both conformations, solved by X-ray crystallography and NMR (Figure 1.1.3.1), are present as a minor and major population, respectively. Although less than 10% of the total HABD exists as the O conformation in the HA-bound form, it is likely that sparsely populated O conformation was selectively crystallized because the PD conformation is

unfavorable for crystallization due to the flexibility of the C-terminal segment (Ogino et al., 2010). In contrast, the PD conformation in solution could be obtained in the structure determination by NMR of HA-bound HABDs (Takeda et al., 2006), since the signal from the major conformation can be predominantly observed in NMR spectra.

A   B
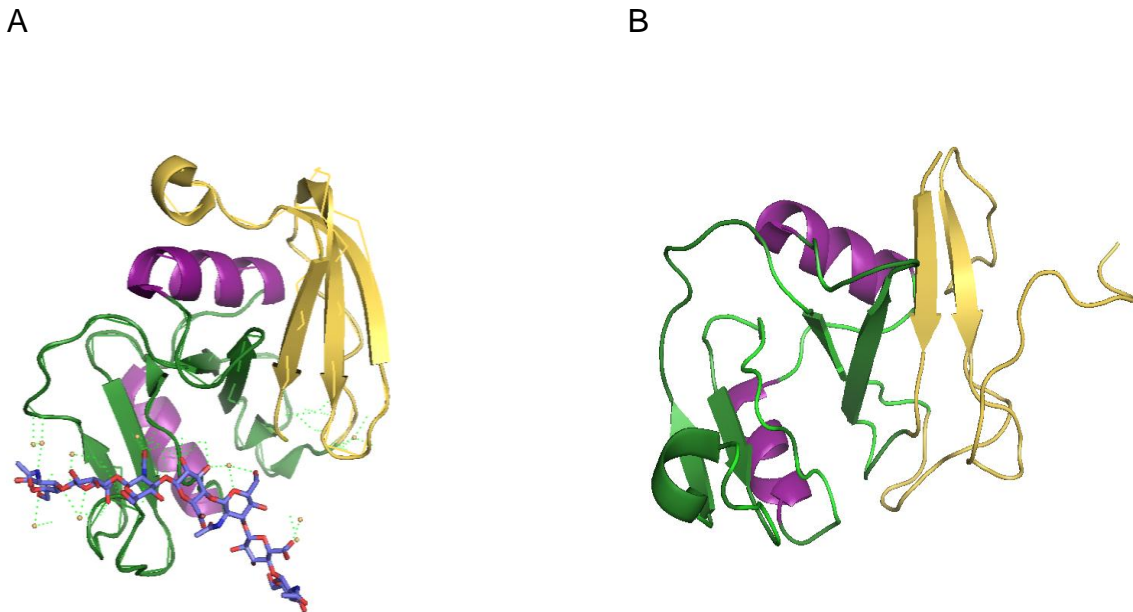
Figure 1.1.3.1 (A) Crystal structure of CD44 in the HA-bound state (PDB: 2JCR). The Link module is colored green and the extension segments are yellow. The α1 and α2 helix are colored purple. (B) NMR structure of CD44 in complex with HA hexamer (PDB: 2I83). Color representation is the same as in (A). The figures were prepared with the PyMol software.

These two distinctive conformations differ in ligand-binding affinity, where the PD form shows high affinity and O form demonstrates low affinity. During cell rolling adhesion the ligand-free state assumes O conformation, which shows higher

tethering frequency than the PD conformation. As a result of the HA binding, the HABD undergoes the conformational change to PD state, which is more resistant for the detachment force with the slower "cellular off rate". However, the CD44 HABD in the HA-bound state still exist in the equilibrium between O and PD forms, with less than 10% of O conformation. Thus, the breakage of the receptor-ligand bound at the rear edge could be facilitated at a moment, when the conformation-liganded CD44 transitions changes to the O conformation (Figure 1.1.3.2). It was shown that the time scale of the conformational exchange of the HABD is comparable to the dwell time of the ligand-receptor bound on the cell surface. These findings support hypothesis that, in rolling adhesion, the O conformation could accelerate the dissociation of the receptor-ligand bound at the rear edge. The two-state equilibrium of the HABD of CD44 between high- and low-affinity states seems to be crucial in CD44-mediated cell rolling under flow conditions (Ogino et al., 2010).
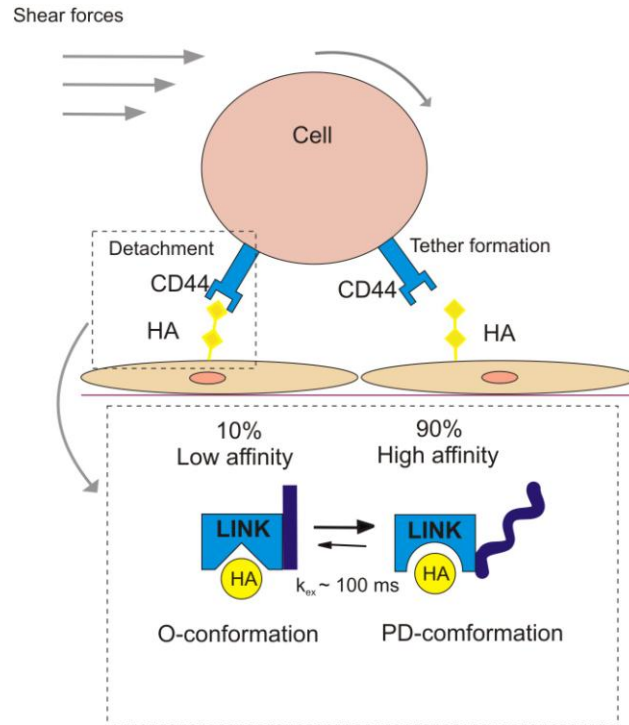
Figure 1.1.3.2 Scheme showing the significance of two-state equilibrium of CD44 HA-binding domain in the cell rolling. Detachment at the trailing edge, which occurs at the time scale of 100 ms, would be facilitated by the conformational transition from the PD (high affinity) to the O (low affinity) conformation (Adapted from Ogino et al., 2010).

1.1.4 CD44 in human cancer

Hyaluronidase and matrix metallproteinases (MMPs) are extracellular matrix degrading enzymes and are involved in tissue remodeling during development, wound healing, bone resorption, and angiogenesis (Stetler-Stevenson et al., 1993). In this context it is not surprising that tumor cells, which overexploit many normal functions, make use of the same enzymes to allow their malignant spread. Degradation of HA by hyaluronidase may allow tumor invasiveness and the induction of angiogenesis by HA fragments. HA matrix can also support CD44-

dependent tumor cell migration as well, so that its degradation might interfere with tumor spread. This double-edged function of the enzyme has yielded apparently conflicting findings (Naor et al., 2002). On the one hand, a correlation was found between elevated levels of intracellular hyaluronidase and high-grade bladder cancer, as well colon carcinoma and metastatic melanoma (Liu et al., 1996; Pham et al., 1997). On the other hand, however, it was shown that mice expressing augmented levels of hyaluronidase were resistant to tumor growth (De Maeyer and De Maeyer-Guignard, 1992). The outcome of the hyaluronidase activity is possible at variance in different tumors, depending on enzyme isotype, the size of the degraded HA, degradation intensity, and the dependency of tumor cell migration on CD44-HA interaction. For instance, extensive HA digestion by hyaluronidase may create spaces in the ECM that allow tumor invasiveness, while delicate digestion might have a different effect – disruption of the ability of HA to support cell migration. Therefore, enzymatic activity of each tumor must be individually explored and generalizations should be avoided (Naor et al., 2002). Interestingly, cell surface serine proteases and MMPs can also be used to cleave CD44 itself, thus releasing the ectodomain of this molecule. CD44 cleavage may play a crucial role in cell detachment from HA substrate and, consequently provide a mechanism for CD44-dependent cancer cell trafficking (Naor et al., 1997).

Many studies document the prevalence and diagnostic value of CD44 and its variant isoforms in human cancer. However, the precise function of CD44 in tumor has not yet been completely established. CD44 has been defined as a cancer-initiating cells (CIC) marker in many tumors entities, but it remains to be explored whether CICs

preferentially express standard or variant isoform of CD44. CD44 has biological functions that would be of value for CICs. This is mainly due to the bidirectional signaling between CD44 and its surroundings. The crosstalk between CD44 and transmembrane and cytoplasmic molecules works as a potential amplifier of a variety of signals. According to results from initial preclinical trials CICs receive multiple benefits from CD44 and strengthen the point that CIC function might be disrupted by the blockade of this one molecule (Zoller, 2011). Several physiological functions of normal cells mediated by cell surface CD44 have been "adopted" by cancer cells, frequently in an uncontrolled manner. These include the following CD44-dependent cellular functions: cell to cell or cell to matrix interactions, vasculogenesis, cell migration, presentation of growth factors, chemokines or cytokines to the adequate receptors, signaling for cell survival, and docking of proteases at the cell surface to allow cleavage of matrix components. On the other hand, the CD44 receptor can also mediate cell functions that do not favor tumor growth and progression, such as cell differentiation and apoptosis. Cell to cell or cell to matrix interaction dependent on CD44 can protect tumor cells from immune surveillance, as well as enhance their entrapment in target organs and facilitate their colonization in secondary tissues (Naor et al., 2002).

CD44 expression on malignant cells and on their normal counterparts has been detected by a variety of techniques. Enhanced expression of CD44 isoforms in malignant tissues was detected in patients with the following diseases: thyroid carcinoma, lung cancer, hepatocellular carcinoma, renal cell carcinoma, melanoma, endometrial cancer, and ovarian carcinoma. The association between a gradual

increase in CD44 isoform expression and tumor progression from less malignant to more advanced stages is another indication of CD44 involvement in the malignant process. However, it was reported that unchanged or even reduced expression of CD44 isoforms occurred in breast, gastric and bladder carcinoma. Loss of CD44 may be associated with advanced stages of the cancer, because downregulation of the protein may be related to the detachment of potential metastatic cells from the primary growth. CD44 functions as cell-cell or cell-matrix adherence molecule, therefore in such cases, the release of cells from the primary tumor is dependent on reduced expression of CD44. The level of CD44 in tumor biopsies or in tissues of resected tumors may be used for prognostication in patients with neoplastic diseases. In conclusion, CD44 is a promising targeting molecule for the therapy and diagnosis of several human cancers (Naor et al., 2002). However, many uncharted areas on the CD44 map have to be explored before CD44-based targeting of cancer cells and it was suggested that high efficiency can be achieved with a wide range of drugs as far as side effects can be prevented (Zoller, 2011).

1.1.5 Hyaluronic acid

Hyaluronic acid (HA) is a high molecular weight linear polysaccharide composed of alternating units of D-glucuronic acid and N-acetyl-D-glucosamine with $\beta(1\rightarrow4)$ interglycoside linkage (Figure 1.1.5). It is the only non-sulfated glycosaminoglycan (GAG) that is abundant in the extracellular matrix (ECM) and synovial fluid. More than 50% of HA in the body is present in intestine, skin and lung.

HA can be also found in the umbilical cord, blood and synovial fluid. HA plays pivotal role in cell motility, wound healing, angiogenesis construction of ECM (Oh et al., 2010), as well as organ structural stability, tissue organization and proper cell growth (Platt and Szoka, 2008). Its physicochemical properties, like the capacity to bind large amounts of water, the filtering effects on the molecular level and the formation of viscous gels enable tissue remodeling in normal and pathological context (Toole et al., 2005). The amount of HA in tissues depends upon a complex interplay among HA synthesis by HA synthases, extracellular degradation by hyaluronidases and HA internalization by cell surface receptors (Platt and Szoka, 2008).

Hyaluronan is produced by the hyaluronan synthases (HAS1-3), which are integral plasma membrane proteins. Newly synthesized HA oligosaccharides are transported directly onto cell surface for assembly into pericellular or extracellular matrix. HA molecules can also be bound by its cell surface receptors (CD44, CD168/RHAMM, layilin, LYVE-1, and Toll-like receptor-4) or can be retained on the cell surface by attachment to the synthase. Interactions between HA and its receptors initiate intracellular signaling transduction. The intracellular domain of the receptor interacts with cytoskeletal proteins passing the signal between HA and intracellular structures. Activation of various forms of CD44 by HA can modulate cell proliferation, aggregation, migration and angiogenesis. RHAMM (receptor for hyaluronan-mediated motility) binds HA and can activate the protein kinases, like Src, Erk or focal adhesion kinases. CD44 and RHAMM have different functions depending on the cell type (Sironen et al., 2011).
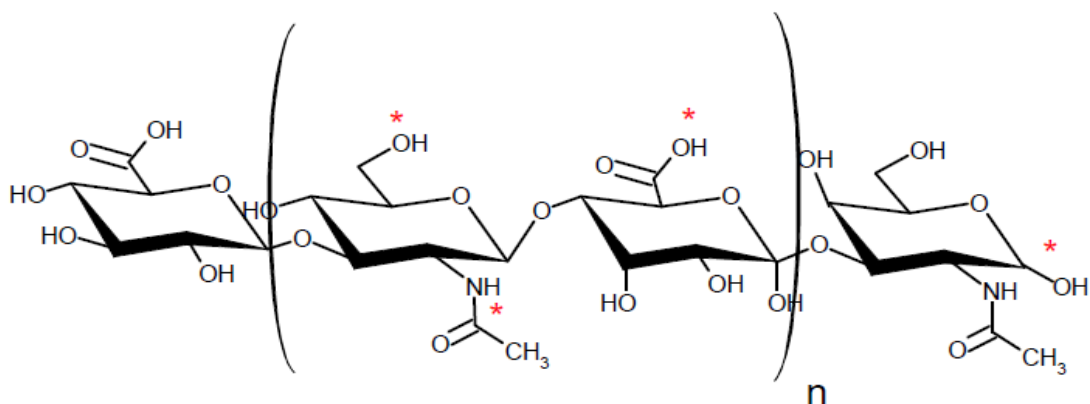
Figure 1.1.5 HA structure: polymeric repeat of N-acetylglucosamine and D-glucuronic acid. The asterisk (*) represents potential sites of chemical modifications (adapted from Platt and Szoka, 2008).

HA turnover is due to systematic clearance from the blood by the HARE receptor located on liver sinusoidal endothelial cells, cellular catabolism and removal by endocytosing LYVE-1 receptor on cells in the lymphatics (Platt and Szoka, 2008). The major portion of HA in the tissues is taken up by lymphatic system. The degraded HA enters blood circulation and is transported to the liver by receptor-mediated endocytosis, where it is catabolized (Oh et al., 2010). Catabolism of HA is performed by hyaluronidases (Hyal-1, Hyal-2, Hyal-3, and PH-20) or reactive oxygen species (ROS) (Sironen et al., 2011). A molecular weight of HA occurring in the body is very wide, ranges from 1000 to 10,000,000 Da. HA breaks down to small oligosaccharides during the metabolic pathways through the tissues, lymphatics, lymph nodes, kidney, liver and blood, in series. HA role in body depends on its MW. For example, degraded fragments of HA induce receptor-mediated intracellular signaling while high MW HA, which is typically found in loose connective tissues, has a role in maintain cell integrity and water content in ECM (Oh et al., 2010). The half-

life of circulating high molecular weight HA is about two to five minutes. In contrast, HA in skin has a half-life of over a day. Human body exploits in total about five grams of HA per day as outcome of various clearance processes (Platt and Szoka, 2008). Hyaluronan degrades to glucuronic acid and N-acetyloglucosamine in lysosomes and finally they are metabolized by hepatocytes to CO_2, H_2O and urea.

1.1.6 Hyaluronan conjugates as drug carriers

Because of various biological functions and excellent physicochemical properties, HA was widely used in arthritis treatment, ophthalmic surgery, drug delivery and tissue engineering. Especially for drug delivery applications HA exhibits a number of properties which make it successful drug carrier. HA is water soluble, non-toxic, non-immunogenic, non-inflammatory, and a completely biodegradable polysasccharide, and since it is the major ligand of receptors (CD44 and RHAMM) overexpressed in several cancers, HA can be used to target cells on which these proteins are expressed (Luo and Prestwich, 1999; Oh et al., 2010). Coupling of antitumor agents to biopolymers can provide advantages in drug solubilization, localization, stabilization, and controlled release (Maeda et al., 1992). HA can be coupled with an active cytotoxic prodrug directly to form non-toxic prodrug. It has been shown that these bioconjugates are internalized into cancer cells through receptor-mediated endocytosis, followed by release (e. g. by intracellular enzymatic hydrolysis (Luo et al., 2000) and activation of the drug and the same time, thus restoring its original function (Jaracz et al., 2005). Hyaluronan was reported to be

used for conjugation with various chemical drugs such as paclitaxel, doxorubicin, mitomycin C, and butyrate as well as with siRNA, DNA and peptides (Platt and Szoka, 2008).

1.1.7 Chemical modification for the preparation of various HA derivatives

HA has several chemical groups to which drugs can be conjugated. The carboxylate on the glucuronic acid, the N-acetylglucosamine hydroxyl, and the reducing end have all been successfully used in conjugation reactions with drugs. The structure with possible sites of chemical conjugation is shown in Figure 1.1.5. The acetyl group can be enzymatically removed from the N-acetylglucosamine and is also potential modification site (Platt and Szoka, 2008).

Currently, HA is commercially produced by extraction from microbial fermentations and cock's comb, which has a very high molecular weight (MW) up to few millions. For chemical modifications high MW HA has to be degraded to low MW in the range from several thousand to ca. 500 kDa by the treatment with enzyme hyaluronidase, heat, pressure or UV. The reasons for using low molecular weight HA are: to monitor the saccharide loading by NMR, to have a readily non-viscous solution; besides once in the plasma, LMW HA is taken up quickly by cells without further degradation and can be expelled from the body via kidneys (Leonelli et al., 2008). Then, the LWM HA can be easily chemically modified in aqueous solution or organic solvents such as dimethyl sulfoxide or hexafluoroisopropanol. HA can be designed to be coupled with various functional groups through its pendant groups. Mostly, chemical

modifications were carried out through carboxyl groups of HA with introduction of amine groups to HA using adipic acid dihydrazide (ADH), hexamethylenediamine (HMDA), or cystamine. In order to prevent crosslinking reactions, the molar amount of ADH, HMDA or cystamine should be much larger than that of carboxyl groups of HA. By the conjugation of HA-COOH with N-hydroxylsuccinimide (NHS) and periodate oxidation, HA with amine-reactive groups can be prepared in the forms HA-NHS and HA-aldehyde. HA with the thiol-reactive functional groups can be prepared by conjugation of HA-COOH with cystamine and reduction with dithiothreitol (DTT), the conjugation of HA with iodoacetoamide, and that with 3-(2-pyridyldithio) propionyl hydrazide (PDPH). In case of chemical modifications in organic solutions, tetrabutylammonium salt of HA (HA-TBA) has to be prepared. Then the HA-TBA derivative can be reacted with aminoethyl methacrylate (AEMA), aminopropyl methacrylamide (APMAm), or vinyl sulfone (VS) to prepare HA-AEMA, HA-APMAm, and HA-VS, respectively (Oh et al., 2010). The resulting chemical formulas of HA derivatives with various functional groups prepared for further applications are shown in Figure 1.1.7.

The efficiency of bioconjugates depends on the level of cytotoxic agent loading to HA. Chemical modification of HA-COOH changes its biological behavior in the body, since carboxyl groups of HA are known to be recognition sites for HA receptors and hyaluronidases (Banerji et al., 2007). More than three carboxyl groups (hexasacchaaride) in the HA molecule are related with its binding to HA receptors. It was reported that HA modifications less than 25 mol% appeared not to affect its receptor-mediated endocytosis, what makes it possible to apply slightly modified HA

derivatives to target specific intracellular delivery of biopharmaceuticals. According to the real-time bioimaging using quantum dots, HA conjugates with 35 mol% HA modifications maintaining enough binding sites for HA receptors were mainly internalized in the liver, while those with 68 mol% of modification lost much of HA characteristics and were evenly distributed to the tissues in the body (Oh et al., 2010).
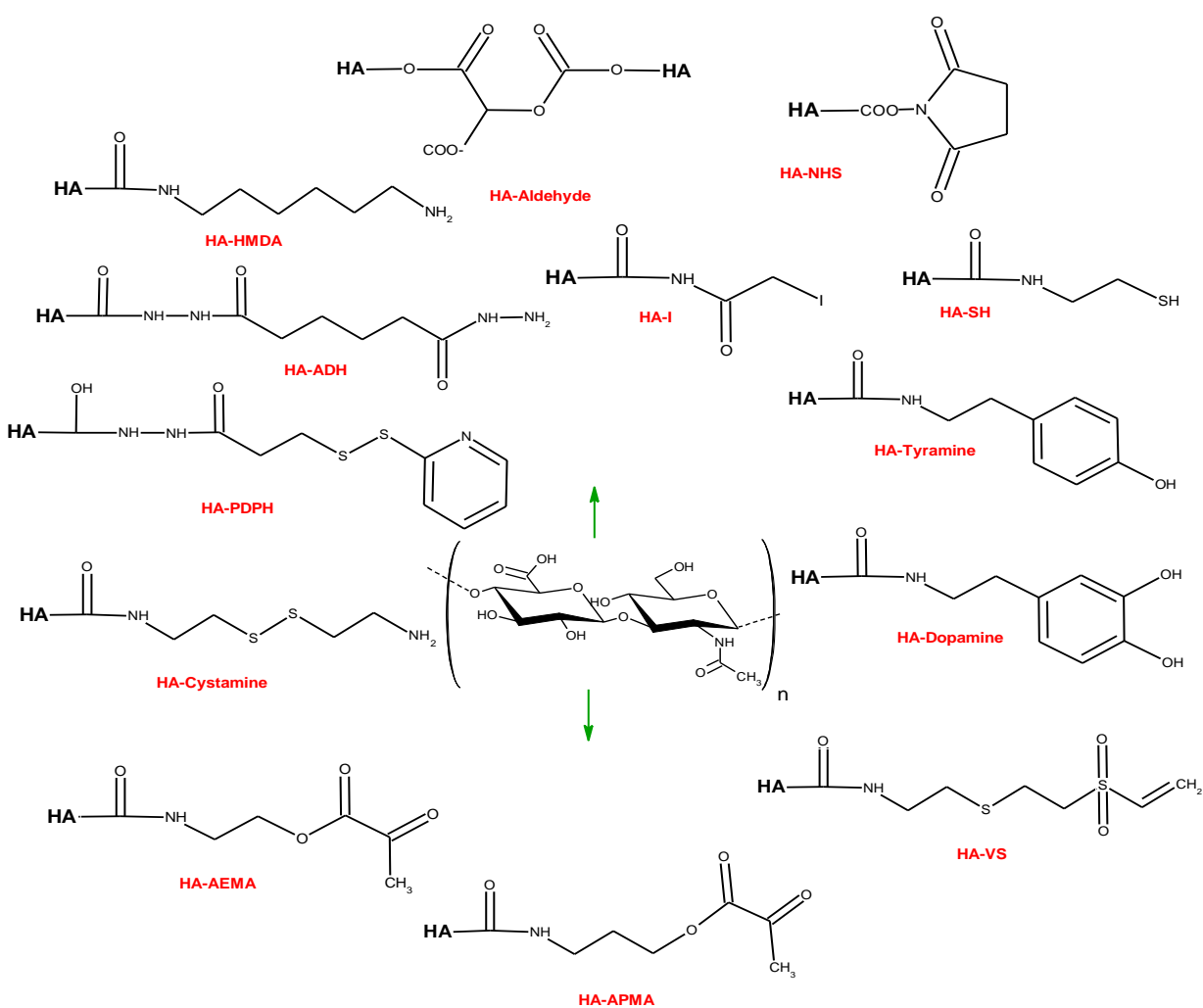
Figure 1.1.7 Chemical modification of hyaluronic acid (HA) in water or in DMSO for the preparation of various kinds of HA derivatives (adapted from Oh et al., 2010).

1.2 Ubiquitin and ubiquitin-like proteins

1.2.1 Ubiquitin-like proteins

Protein homeostasis is essential for most processes in the cell. The ubiquitin-preoteasome system (UPS) is responsible for much of the regulated cellular proteolysis, and has non-degradative functions as well. Ubiquitin can be reversibly attached to other proteins and belongs to an elaborate post-translational modification pathway. Several ubiquitin-like proteins (UBLs) have been identified, like SUMO, NEDD8 and ISG15, which structurally share the ubiquitin characteristic three-dimensional fold but are otherwise distinct (Bedford et al., 2011). Covalent modification of proteins by ubiquitin and UBLs often critically alters substrate activity by influencing metabolic stability, localization or binding behavior. The switch-like properties of UBLs are crucial for pathways that regulate, for example, protein sorting, signal transduction, DNA repair, and development (Hochstrasser, 2009). Because the UPS and UBLs conjugation pathways have many essential biological roles, their malfunctions are important factors in various human diseases, like for example numerous cancer types, viral diseases, cardiovascular diseases and some serve mental retardation (such as Angelman syndrome), neurodegenerative disorders such as Parkinson's dieases, Huntington's disease and Alzheimer's disease, and type 2 diabetes (Bedford et al., 2011; Hochstrasser, 2009). It should be possible to target these diseases by modulating components of the UPS and UBL conjugating pathways using small-molecule inhibitors. Intervention in the UPS in cancer therapy has been already demonstrated by the proteasome inhibitor

bortezomib (Velcade; Millenium). Recently, several enzymes, which are important components of UPS and UBL mechanisms, have been identified and can be additional targets for small molecule inhibitors (Bedford et al., 2011).

1.2.2 Ubiquitin, a diverse cellular signal

Ubiquitin is a 76-residue polipeptyde that is activated and attached to substrate proteins in highly controlled manner. Conjugation of ubiquitin to a target protein or to itself is regulated by sequential activity of a series of enzymes. Three types of enzymes – ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin-ligating (E3) enzymes - carry out ubiquitin-modification reactions, including assembly of polyubiquitin chains. E1s activate ubiquitin in ATP dependent reaction. E2s pick up the ubiquitin by transthiolation from E1 and conjugate it to substrates. E3s recognize and ligate the substrate to ubiquitin. In all eukaryotes modifications of many proteins are performed in a highly specific manner due to occurrence of multiple isozymes of E2 and E3 (up to several dozen E2s and many hundreds of E3s). Such modifications are often under strict temporal and spatial control (Hochstrasser, 2009). Ubiquitin is usually attached to proteins through an amide linkage between its carboxyl terminus and a primary amino group on the acceptor protein. The most common linkages are with the ε-amino group of lysine, and additionally recently it has been found that ubiquitin can be attached to cysteine, serine and threonine residues in proteins as well (Hochstrasser, 2009). The complementation of one ubiquitin cycle results in a monoubiqutilated substrate.

Additional ubiquitin molecules can be ligated to a lysine residue (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63) in a previously attached ubiquitin, allowing for seven possible homotypic linkage types and multiple possible heterotypic chains. Alternatively, ubiquitin chains can be linked head to tail to form linear chains. As consequence, cellular proteins are modified by various ubiquitin signals like monoubiquitin, multiple monoubiquitin marks and ubiquitin chains, which can be of diverse length and linkage. The attachment of a single ubiquitin to a target protein can alter protein localization and activity (by regulating lysosomal targeting, endocytosis, meiosis and chromatin modeling). Polyubiquitylation is implicated in events such as targeting to the 26S proteasome, DNA repair and immune signaling (Dikic et al., 2009). The linear (head-to-tail) ubiquitin polymer chains, the most recently described polymers of ubiquitin, are built by a specific ligase complex called the linear ubiquitin chain assembly complex (LUBAC) and are essential for the nuclear factor- κB (NF-κB) signaling (Iwai and Tokunaga, 2009). Substrates that are labeled with Lys48-linked ubiquitin chains are delivered to the 26S proteasome to be degraded into small peptides, with the ubiquitins released to be used again (Bedford et al., 2011). The polyubiquitin chain provides a generic affinity tag that result in tight binding of the proteolytic substrate to the proteasome. Multiple polyubiquitin receptors are present within the proteasome. Additionally, polyubiquitin-binding domains are found in mobile shuttling factors that direct marked proteins to the proteasome (Hochstrasser, 2009).
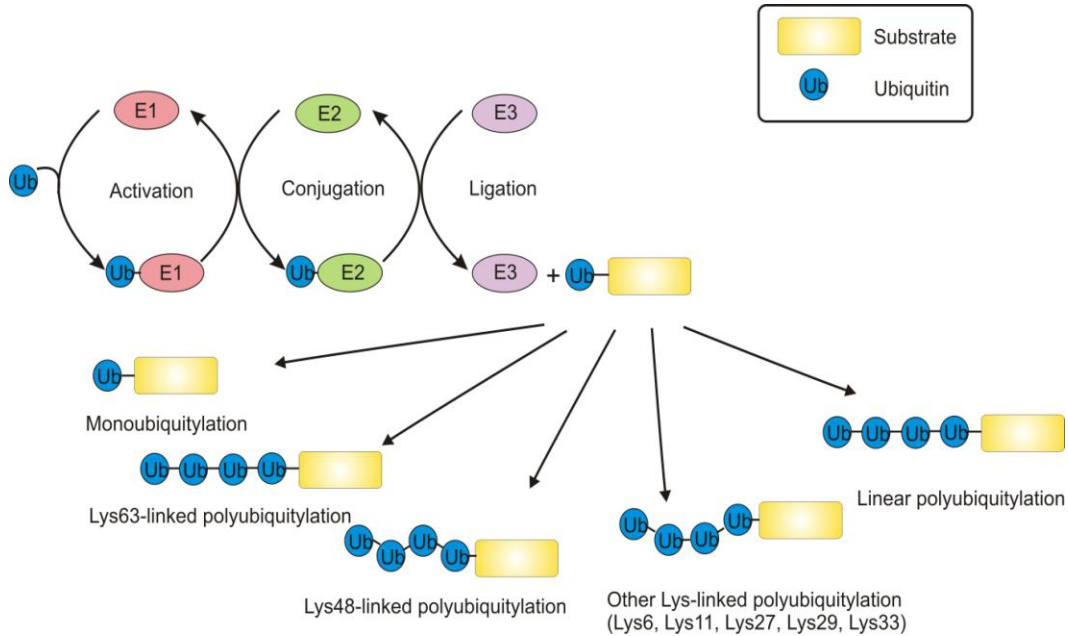
Figure 1.2.2 Basics of the ubiquitin-conjugation pathway. The activity of three enzymes is required for ubiquitylation. Enzyme E1 activates ubiquitin, E2 is a ubiquitin-conjugating enzyme, and E3 is a ubiquitin ligating enzyme, which recognized the substrate. One complete ubiquitylation cycle results in a monoubiqutylated substrate. Additionally, ubiquitin molecules can be ligated to one of Lys residues in a previously attached ubiquitin to form various Lys-linked chains (Adapted from Dikic et al., 2009).

Substrates that are polyubiquitylated with the Lys63-linked ubiquitin chains are generally not degraded but are essential components of signaling pathways. The functions of ubiquitin chains that are linked through lysines 6, 11, 27, 29, and 33 and of chains which contain 'mixed' linkages are still to be evaluated, although there is increasing evidence that the non-Lys63 chains target proteins for degradation by the 26S proteasome (Bedford et al., 2011). Deubiquitylating enzymes (DUBs) regulate the function of various ubiquitin modifications and may rescue a marked protein from

degradation by removing a degradative ubiquitin signal or by detaching or changing a non-degradative ubiquitin signal.

Covalent conjugation of UBLs to a substrate's target is also ATP dependent, involves an enzyme cascade, and usually requires a free di-glycine (GG) motif at the protruding carboxyl-terminal end of the UBL. Archetypal UBLs (ubiquitin, SUMO, NEDD8) are expressed as inactive precursors with C-terminal extensions, which are removed by UBL-specific proteases, exposing the crucial C-terminal GG motif. Enzymes of this class also mediate UBL deconjugation, thus making the UBL-dependent switch reversible (Hochstrasser, 2009).

1.2.3 Ubiquitin binding domains (UBDs)

Specific interaction between ubiquitin and UBDs are not reduced to the binding of ubiquitylated molecules to the proetasome. It has been shown in the past decade that many of the functions of ubiquitin and UBLs are mediated by association with UBDs (Hurley et al., 2006). UBDs are diverse modules in a protein that can bind, and often distinguish, different types of ubiquitin modifications. The amount of identified UBDs is constantly growing, with more than twenty different families identified and new classes of UBDs are continuously being reported. They differ both in structure and in the type of ubiquitin recognition that they use. Mostly they fold into α-helical structure, zinc fingers (ZnFs), plekstrin homology (PH) folds, or the ubiquitin-conjugating domains present in E2 enzymes. The new findings about their mechanism of molecular recognition reveal that loops rather than secondary

structural elements of UBDs can mediate ubiquitin binding and indicate that distinct tertiary folds, which are not easily distinguishable as modular domains, may create yet another set of UBDs (Dikic et al., 2009). UBDs usually bind to ubiquitin weakly (Hurley et al., 2006). Linking multiple ubiquitin moieties into a chain may have marked effects on affinity or avidity of ubiquitin for target proteins. Knowledge of ubiquitin-UBDs interactions may be used in medical applications, since several UBDs have now been linked to various human pathologies, like cancer and immune deficiencies, thus becoming interesting as putative targets for therapy (Hoeller and Dikic, 2009).

1.2.4 Consequences of UBL-protein conjugation

Conjugation of UBLs to a target protein frequently promotes interaction of the target protein with other proteins. Usually, when UBLs enhance the target's interaction with another macromolecule, they do so by participating directly in the formation of the binding interface with the target protein (Figure 1.2.4 A). UBLs modulate interactions also by attachment to target protein and induce an allosteric change in a target binding site, which enhances or inhibit binding to the target site (Figure 1.2.4 B). Most known mechanism of regulation by UBLs fall into the first category, and only the handful operate by the latter (Archer et al., 2008). Modification performed by one UBL helps recruit a factor that is different from the substrate that would be recruit when another type of UBL would be employed (Figure 1.2.4 C). These modifications can be mutually exclusive and may potentially

involve the same attachment site. Another way, by which UBLs can also exert their function, is through steric hindrance, as the attached UBL can simply block the binding of the proteins. There are relatively few well established examples of this intermolecular mechanism (Figure 1.2.4 D).
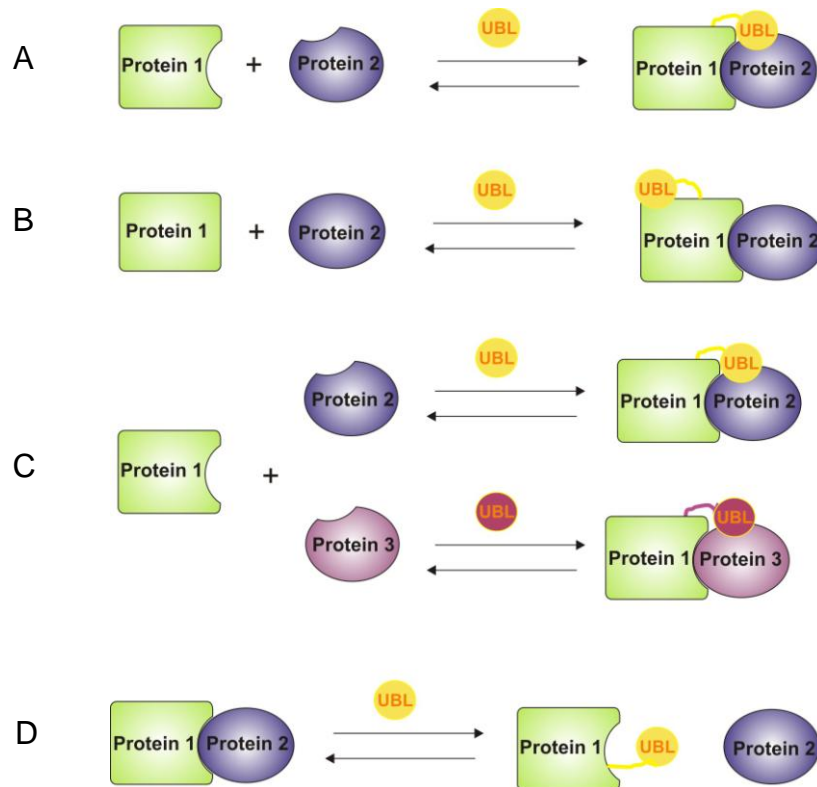
Figure 1.2.4 General function of UBL tagging. (A) Conjugation, which facilitates protein association providing an additional binding site. (B) UBL conjugation induces conformational change that enhances or inhibits binding to a target site. (C) (D) UBL conjugation, which blocks an interaction between two proteins. (Adapted from Hochstrasser, 2009).

1.2.5 Ubiquitin-like protein 5 (UBL5)

UBL5 is a widely expressed human protein that is strongly conserved across phylogeny. Orthologs of UBL5 occur in every eukaryotic genome characterized to date. This strong conservation, from yeast to mammals, suggests an important function for UBL5 (McNally et al., 2003). UBL5 was initially identified as a highly expressed gene in human iris (Friedman et al., 2001). The protein contains 73 amino acids with a molecular weight of 8.5 kDa and pI 8.6. UBL5 (known also as a Beacon) was reported to be involved in feeding behavior and development of type 2 diabetes and obesity in the Israeli sand rat *Psammomys obeseus* (Collier et al., 2000). Reports indicate that UBL5 may interact with some cyclin-like kinases (CLK1, 2 and 4) (Kantham et al., 2003). These kinases have been reported to be involved in the activity regulation of PTP1B (Moeslein et al., 1999), a protein involved in the development of obesity and diabetes. Based on sequence homology and structure prediction algorithms, it was predicted that the structure of UBL5 should be similar to ubiquitin (Friedman et al., 2001). However, sequence similarity is weak, and the protein does not comprise the characteristic di-glycine residues at the C terminus that are required for ubiquitin-like modifiers to conjugate to their targets. The structure of UBL5 resolved by nuclear magnetic resonance shows that secondary structures of the two proteins are almost identical (Figure 1.2.5). The one notable difference between the two structures is the unstructured C terminus found in ubiquitin that extends beyond the end of the highly structured part of the protein. The additional C-terminal residues important for conjugation in ubiquitin-like protein modifiers are, however, missing (McNally et al., 2003). It was suggested that if UBL5

behaves like a ubiquitin modifier (Jentsch and Pyrowolakis, 2000), then the mechanism of protein conjugation should be distinct from that of other ubiquitin-like modifiers because of the lack of the unstructured C terminus and C-terminal di-glycine motif, which was found in other proteins of this class (McNally et al., 2003).

A   B
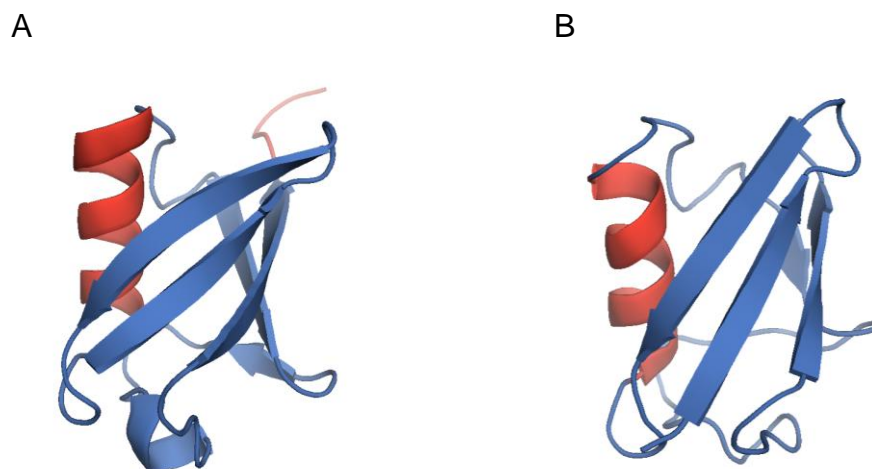
Figure 1.2.5 (A) Ribbons depiction of the crystal structure of ubiquitin (PDB: 1UBQ). (B) Ribbons depiction of UBL5 resolved by NMR (PDB: 1POR). The α-helixs and unstructured C terminus of ubiquitin are colored in red. The figures were prepared with the PyMol software.

The yeast ortholog Hub1 (homologous to ubiquitin) from *Saccharomyces cerevisiae* has been reported to have ubiquitin-like modifier activity (Jentsch and Pyrowolakis, 2000). Hub1 has been linked to various physiological functions including cell cycle progression and polarized growth, mRNA splicing and the mitochondrial unfolded protein response (Mishra et al., 2011). Hub1, as well as UBL5, are unique in lacking a protruding C-terminal GG motif. Instead, both proteins contain a C-terminal tail with the double tyrosine (YY) motif, followed by a non-conserved amino acid residue

(McNally et al., 2003; Ramelot et al., 2003). It was found that Hub1 forms covalent conjugates similar to ubiquitin and proposed that Hub1 is synthesized as a precursor and matured by processing the C-terminal YY motif (Dittmar et al., 2002; Mishra et al., 2011). However, no specific Hub1 processing, conjugation, deconjugation enzymes have been identified. Further studies excluded that Hub1 functions as a covalent modifier. In fact, Hub1 was reported to bind proteins non-covalently and independently of ATP, and the YY motif was shown to be nonessential (Luders et al., 2003; Yashiroda and Tanaka, 2004).

1.2.6 Role of Hub1 in alternative splicing

Hub1 is structurally closely similar to ubiquitin and equally strongly conserved and ancient, yet the functions of two proteins are completely different (Mishra et al., 2011). Recently, Hub1 was found to be important factor in alternative splicing and splice-site usage. Alternative splicing substantially increases the gene product repertoire and is a major source of cell differentiation. It is estimated that majority of human pre-mRNA undergoes alternative splicing (Keren et al., 2010). Hub1 modifies spliceosome through binding to Snu66, spliceosomal (U4/U6.U5) tri-small nuclear ribonucleoprotein particle (snRNP). However, how this interaction affects the splicing is unclear. The minimum polypeptide sequence of Snu66 is 18-19 amino acids long, and was called HIND (Hub1-binding domain). The X-ray crystal structure of the Hub1-HIND complex showed that Hub1 modifies the substrate through non-covalent manner and revealed a new binding paradigm unseen in interactions of ubiquitin and

ubiquitin-binding proteins with their binding partners (Mishra et al., 2011). Most ubiquitin-binding modules of ubiquitin substrates bind to the hydrophobic surface of ubiquitin centered on Ile44 (Dikic et al., 2009), which is almost exactly on the site oppositing the HIND-binding face of ubiquitin fold. The Hub1-HIND interaction comprises a salt bridge and several hydrophobic contacts, and is stronger than most ubiquitin interations with ubiquitin receptors (Mishra et al., 2011). Hub1 in *S. cerevisiae* is not required for general splicing and the usage of canonical 5' slices sites but is required for the usage of certain non-canonical splice sites.

2 Materials and laboratory methods

2.1 Materials

2.1.1 *E. coli* strains and plasmids

Cloning strains

DH5α   Invitrogen (Holland)

TOP10 F´   Invitrogen (Holland)

Protein expression strains

BL21 RIL (DE3) Codon Plus   Stratagene (USA)

Plasmids

pET 16b   Novagen (Canada)

pET 19b   Novagen (Canada)

pET 28a(+)   Novagen (Canada)

pGEX-2T   GE Healthcare (Sweden)

2.1.2 Cell growth media and stocks

For 1 l LB medium: 10 g tryptone

                                                      5 g yeast extract

                                                      10 g sodium chloride

pH was adjusted to 7.0. For the preparation of agar plates the medium was supplemented with 15 g agar.

Minimal medium (MM) for uniform enrichment with ^{15}N:

For 1 l of medium: 0.5 g NaCl

                                                      1.3 ml trace elements solution

                                                      1 g citric acid monohydrate

                                                      36 mg ferrous citrate

                                                      4.02 g KH_2PO_4

                                                      7.82 g K_2HPO_4 x $3H_2O$

                                                      1 ml Zn-EDTA solution

                                                      1 g $^{15}NH_4Cl$

pH was adjusted to 7.0 with NaOH. The medium was autoclaved, and upon cooling,

separately sterilized solutions were added in following amounts:

25 ml glucose

560 µl thiamin

antibiotics

2 ml $MgSO_4$

Defined medium for selective labeling of proteins

For 1 l of medium:

400 mg Ala, Glu, Gln, Arg, Gly

255 mg Asp

125 mg Met

125 mg cytosine, guanosine, uracil

100 mg Asn, Leu, Lys, His, Pro, Thr

100 mg Tyr

400 mg Ile, Val

50 mg Phe, thymine, thymidine

1.6 g Ser

10 mg $CaCl_2$

2 g NaAc

10 g K_2HPO_4

1 g citric acid

1.3 ml trace element solution

36 mg ferrous citrate

1 ml Zn-EDTA solution

1 g NH_4Cl

pH was adjusted to 7.0 with NaOH, and the mixture was autoclaved. Separately sterilized stock solutions were added to the medium in the following amounts:

25 ml glucose

560 µl thiamin

antibiotics

2 ml $MgSO_4$

50 mg Cys, Trp, nicotinic acid

0.1 mg biotin

Stock solutions

Ampicillin: 100 mg/ml of ampicillin in deionized H_2O, sterilized by filtration, stored in aliquots at -20°C until used. Final concentration: 120 µg/ml.

Chloramphenicol was dissolved in ethanol (0.34 g/10 ml) to the end concentration of 34 mg/ml. Final concentration: 34 µg/ml.

Kanamycin: 60 mg/ml of kanamycin in deionized H_2O, sterile filtrated and stored in aliquots at -20°C until used. Final concentration: 60 µg/ml.

IPTG: a sterile filtered 1 M stock of IPTG in distilled water was prepared and stored in aliquots at -20°C until used.

Glucose: 20% (w/v) in deionized H_2O, autoclaved.

Thiamin: 1%, in deionized H_2O, filtrated.

$MgSO_4$: 1 M, in deionized H_2O, filtrated.

Zn-EDTA solution:	5 mg/ml EDTA
	8.4 mg/ml $Zn(Ac)_2$

Trace elements solution:	2.5 g/l H_3BO_3
	2.0 g/l $CoCl_2$ x H_2O
	1.13 g/l $CuCl_2$ x H_2O
	9.8 g/l $MnCl_2$ x $2H_2O$
	2.0 g/l Na_2MoO_4 x $2H_2O$

pH lowered with citric acid or HCl.

2.1.3 Protein purification buffers

Buffers for immobilized metal-chelate chromatography (IMAC) in native conditions:

Lysis buffer: 50 mM NaH_2PO_4

                                                 300 mM NaCl

                                                 10 mM imidazole

                                                 pH 8.0

Wash buffer: 50 mM NaH_2PO_4

                                                 300 mM NaCl

                                                 20 mM imidazole

                                                 pH 8.0

Elution buffer: 50 mM NaH_2PO_4

                                                 300 mM NaCl

                                                 250 mM imidazole

                                                 pH 8.0

Buffers for IMAC in denaturing conditions:

Binding buffer: 6 M guanidinium chloride

                                                         100 mM NaH_2PO_4

                                                         10 mM Tris

                                                         10 mM β-mercaptoethanol

                                                         pH 8.0

Wash buffer: 6 M guanidinium chloride

                                                         100 mM NaH_2PO_4

                                                         10 mM Tris

                                                         10 mM β-mercaptoethanol

                                                         pH 6.3

Elution buffer: 6 M guanidinium chloride

                                                   100 mM NaH$_2$PO$_4$

                                                   10 mM Tris

                                                   10 mM β-mercaptoethanol

                                                   pH 4.5

Refolding buffers:

Refolding buffer for CD44:   400 mM arginine

                                                   100 mM Tris-HCl

                                                   2 mM EDTA

                                                   5 mM red GSH

                                                   0.5 mM ox GSH

                                                   1 mM PMSF

                                                   pH 8.0

Buffers for purification of proteins with GST-tag:

Binding buffer: 10 mM Tris

                                                  150 mM NaCl

                                                  1 mM EDTA

                                                  pH 8.0

Elution buffer: 50 mM Tris

                                                  20 mM red GSH

                                                  pH 9.0

Gel filtration chromatography buffers:

PBS: 140 mM NaCl

                                                  2.7 mM KCl

                                                  10 mM Na_2HPO_4

                                                  1.8 mM KH_2PO_4

                                                  5 mM DTT

                                                  pH 7.3

Crystallization buffer: 10 mM Tris

                                                 100 mM NaCl

                                                 5 mM β-mercaptoethanol

                                                 pH 8.0

HA Buffer: 0.1 M NH_4HCO_3

2.1.4 Buffer for the DNA agarose gel electrophoresis

50xTAE (for 1 l):

40 mM Tris-acetate   242 g Tris

1 mM EDTA   100 ml of 0.5 EDTA (pH 8.0)

Glacial acetic acid   57.1 ml

2.1.5 Reagents and buffers for the SDS-PAGE

Anode buffer (+):   200 mM Tris pH 8.9

Cathode buffer (-):   100 mM Tris pH 8.25

                                             100 mM tricine

                                             0.1% SDS

Separation buffer: 1 M Tris pH 8.8

                                             0.3% SDS

Stacking buffer: 1 M Tris pH 6.8

                                             0.3% SDS

Separation acrylamide: 48% acrylamide

                                             1.5% bis-acrylamide

Stacking acrylamide: 30% acrylamide

                                             0.8% bis-acrylamide

Pouring polyacrylamide gels:

Separation gel:	1.675 ml H_2O
	2.5 ml separation buffer
	2.5 ml separation acrylamide
	0.8 ml glycerol
	25 µl APS
	2.5 µl TEMED
Intermediate gel:	1.725 ml H_2O
	1.25 ml separation buffer
	0.75 ml separation acrylamide
	12.5 µl APS
	1.25 µl TEMED
Stacking gel:	2.575 ml H_2O
	0.475 ml stacking buffer
	0.625 ml stacking acrylamide

12.5 µl 0.5 M EDTA, pH 8.0

37.5 µl APS

1.9 µl TEMED

5 x SDS-PAGE sample buffer: 0.225 M Tris-HCl, pH 6.8

50% glycerol, 5% SDS

0.05% bromophenyl blue

Protein visualization

Coomassie-blue solution: 45% ethanol

10% acetic acid

Destaining solution: 5% ethanol

10% acetic acid

2.1.6 Reagents and buffers for Western blots

Transfer buffer:	10 mM CAPS pH 11
	10% methanol
Wash buffer:	10 mM Tris
	150 mM NaCl
	0.05% Tween 20
	pH 8.0
1st antibody solution	1:2000 diluted in the wash buffer
2nd antibody solution	1:5000 diluted in the wash buffer

Substrate for alkaline phosphatase BCIP (Sigma); dissolve 1 tablet in 10 ml of water.

2.1.7 Enzymes and other proteins

BamHI	Fermentas (Lithuania)
BglII	Fermentas (Lithuania)

EcoRI   Fermentas (Lithuania)

NdeI   Fermentas (Lithuania)

DpnI   Fermentas (Lithuania)

Pfu DNA Polymerase   Fermentas (Lithuania)

Phusion HF DNA Polymerase   Finnzymes (Finland)

Taq DNA Polymerase   Fermentas (Lithuania)

T4 DNA Ligase   New England BioLabs (USA)

FastAP   Fermentas (Lithuania)

Dnase I   Roche (Germany)

Hyaluronidase   Roth (Germany)

Mouse monoclonal anti-His antibodies   Santa Cruz Biotech (USA)

Goat anti-mouse antibodies   Santa Cruz Biotech (USA)

2.1.8 Kits

QIAquick PCR Purification Kit   Qiagen (Germany)

QIAprep Spin Miniprep Kit   Qiagen (Germany)

QIAquick Gel Extraction Kit   Qiagen (Germany)

pET LIC cloning Kits   Novagen (Canada)

2.1.9 Protein and nucleic acids markers

1kb DNA Ladder   Fermentas (Lithuania)

Prestained Protein Ladder   Fermentas (Lithuania)

2.1.10 Chromatography equipment and columns

ÄKTA explorer 10   Amersham Pharmacia (Sweden)

Peristaltic pump P-1   Amersham Pharmacia (Sweden)

Fraction collector RediFrac   Amersham Pharmacia (Sweden)

Recorder REC-1   Amersham Pharmacia (Sweden)

UV flow through detector UV-1   Amersham Pharmacia (Sweden)

HiLoad 16/60 Superdex s200pg   Amersham Pharmacia (Sweden)

HiLoad 10/30 Superdex s75pg   Amersham Pharmacia (Sweden)

HiLoad 16/60 Superdex s30pg   Amersham Pharmacia (Sweden)

NiNTA-agarose   Qiagen (Germany)

Buthyl Sepharose 4 Fast Flow   Pharmacia (Sweden)

BioloLogic LP System   Biorad (USA)

2.1.11 Other equipment and reagents

Amicon 8050   Millipore (USA)

Ampicilin sodium salt   Roth (Germany)

Amonium chloride (^{15}N)   Cambridge Isotope Lab. (USA)

Chloramphenicol   Roth (Germany)

Kanamycin   Roth (Germany)

Complete Protease Inhibitor Cocktail   Roche (Germany)

Dialysis membrane, MWCO 3.500   Spectrum Laboratories (USA)

Isopropyl β-D-1-thiogalactopyranoside (IPTG)   Fermentas (Lithuania)

Sybr safe DNA gel stain   Invitrogen (Holland)

Tricine   Roth (Germany)

Trizma base   Sigma (USA)

Triton X-100   Roth (Germany)

TopVision™ LE GQ Agarose   Fermentas (Lithuania)

Hialuronic acid sodium salt   Sigma (USA)

All other reagents were purchased in Merck (USA)

2.1.12 Equipment for organic synthesis

Integral 100Q Multidimensional Biospecific HPLC System   PerSeptive Biosystems

Hei-VAP Value rotary evaporator   Heidolph

Two step rotary vacuum pump   Edwards

Source 5RPC ST4,6/150   Pharmacia

2.2 Experimental procedures and principles

2.2.1 General remarks on constructs' design

Optimizing protein constructs is important for X-ray crystallography and NMR. The choice of a construct may decide its solubility, stability, and later determine success or failure of crystallization attempts. Disruption of secondary structure can interrupt the folding and cause an inclusion bodies-directed protein. Alternatively, flexible parts can be subjected to subcutaneous cleavage, contributing to protein instability. Furthermore, flexible fragments of proteins are known to inhibit crystallization and are responsible of obtaining crystals of low quality. Determination of approximate boundaries of protein's domains employs limited proteolysis, protein

sequencing, mass spectrometry, NMR spectroscopy, bioinformatics tools, as well as the information obtained from the previously published literature.

2.2.2 Choice of the expression system

E. coli is the simplest and most popular expression system, which offers several advantages like high yield of production, fast growth and low costs. A large number of prokaryotic expression vectors currently available, possessing various restrictions sites, and encoding tags, and fusion proteins (Makrides, 1996), makes a choice of a suitable vector an easy task. We frequently use pET or pGEX series for the expression heterologoues proteins in *E. coli.*

The pET system was mostly used in this work. This system is based on the phage T7 RNA polymerase, provides high, tightly controlled protein expression, yet still promoting high-level of transcription and translation. The protein coding sequence of interest is cloned downstream of the T7 promoter (not recognized by host polymerase) and protein expression is induced by the IPTG addition, which elicit a chromosomally integrated T7 RNA polymerase expressed from the *lacUV5* promoter. The system provides several levels of control due to additional plasmids, which express T7 lysosyme, a natural T7 RNA polymerase inhibitor enabling a high level of control over a basic (i.e. un-induced) expression of transgenes.

Bacterial cells differ substantially from eukaryotic cells, what influences expression efficiency of the heterologous proteins derived from higher organisms. Each

organism carries its own bias in the usage of the 61 available amino acid codons, and frequencies of codon usage in E. coli is distinct from that of the Eukaryota. In particular, Arg codons AGA, AGG, CGA, CGG, Gly codon GGA, Ile codon AUA, Leu codon CUA, and Pro codon CCC are rarely used. Insufficient tRNA pools can lead to translational stalling, premature translation termination, translation frameshifting and amino acid misincorporation (Kurland and Gallant, 1996). It was shown that the yield of a protein, whose genes contain rare codons, can be dramatically improved when the cognate tRNA is increased within the host (Rosenberg et al. 1993). This is typically accomplished by inserting the wild type tRNA gene into *E. coli* strains, either at the genomic or on a plasmid level. Strains with artificially introduced copies of genes encoding tRNAs are commercially available (Rossetta from Invitrogen; BL21 RIL from Stratagene).

Prokaryotic cells are unable to perform many of the port-translational modifications characteristic for higher eukaryotic cells such as disulfide bond formation, glycosylation folding, and proteolytic processing. Therefore, expression of eukaryotic proteins in *E. coli* leads to much poorer results, often ending up as inactive inclusion bodies. Utilization of eukaryotic, systems like mammalian cells expression or Baculovirus system, solves this problems but is connected with low efficiency and high costs.

2.2.3 DNA techniques

2.2.3.1 Preparation of plasmid DNA

Plasmid isolation from *E. coli* was carried out using the QIAprep Spin Miniprep Kit from Qiagen. The procedure is based on standard alkaline lysis of bacteria in a presence of strong detergent, SDS and RNAse, followed by neutralization and renaturation with acetate. A crude cell was loaded onto a silica membrane column, washed with an ethanol-containing buffer, and eluted in small volume of water, yealding up to 10 µg, depending on the plasmid copy number per cell.

2.2.3.2 PCR

The polymerase chain reaction (PCR) was performed to amplify desired DNA fragments, introduce mutations, restriction sites, and Stop codons. Primers were designed according to standard principles regarding the length, GC-content, melting temperature, and occurrence of secondary structures. All used primers are listed in Tables in respective parts of the thesis.

PCR reaction was carried out in different conditions for three different DNA polymerases:

	Melting temp.	Synthesis temp.
Phusion HF	98°C	72°C
Pfu DNA Polymerase	95°C	72°C
Taq DNA Polymerase	95°C	70-75°C

The annealing temperature varied depending on the primer's melting temperature. The primer stock solutions were always 100 µM and working solutions 10 µM. Usually 2 µl of each primer was used for PCR. Amount of the template was most often 0.1 µg and dNTP 20 nmol for each nucleotide. Synthesis time was adjusted to the product length, calculating 1 kb/min for Pfu DNA and Taq DNA polymerases and 1 kb/15s for Phusion HF polymerase. The synthesis step in the last cycle was extended to 5-10 min to assure complete product length.

2.2.3.3 Digestion with restriction enzymes

Usually, 1-2 units of each enzyme were used for the digestion of 1 µg of plasmid DNA. The reaction was performed in a buffer supplied by manufacturer at the optimal temperature (37°C) for 2-3 h. Because ends of digested DNA were cohesive, 5'-ends of the vector were dephosphorylated using the thermosensitive

alkaline phosphatase (Fast AP) to avoid possibility of plasmid recirculation that can occur when double digestion is not 100% efficient. FastAP treatment was performed with 1 unit of enzyme per 3 µg of plasmid DNA at 37°C for 1h.

2.2.3.4 Purification of restriction digested and PCR products

To purify PCR products and digested DNA from primers, nucleotides, enzymes, buffers, agarose, SYBR-Gold DNA dye and other contaminants, the QIAquick PCR Purification Kit was used. The QIAquick system uses a simple bind-wash-elute procedure. The binding buffer (containing high concentration of salt) was added to the PCR mixture or enzymatic reaction. When gel extraction was performed, the slice of the agarose gel containing DNA was melted in the binding buffer at 50°C. Nucleic acids were absorbed to a silica membrane on a spin column while all impurities were washed away. DNA was eluted with small volume of water or the 10 mM Tris pH 8 buffer.

2.2.3.5 DNA electrophoresis

To verify the presence of DNA and its length, agarose gel electrophoreses were run. For this purpose 1% gel was prepared by diluting an appropriate amount of agarose in the TAE buffer supplemented with the SYBRGold fluorescent dye (Invitrogen). Samples mixed with the 6x DNA Loading Dye (Fermentas) were loaded

on a gel and run (at 100 V for about 30 min) along with the 1kb DNA ladder. The results were analyzed under the UV illumination.

2.2.3.6 DNA quantitation and sequencing

DNA concentration was determined by spectrophotometric measurements at wavelength of 260 nm using the Pharmacia Biotech Ultraspec 3000 spectrometer. Sample concentration was calculated using equation:

$$A_{260} \times 50 \times \text{dilution factor} = \mu g \text{ DNA/ml}$$

Protein contaminations were estimated by measurements at 280 nm wavelength. Pure DNA sample have ratio $A_{260} : A_{280}$ between 1.8 and 2.

To check constructs correctness, plasmids were sequenced at the Core Facility of the Max Planck Institute of Biochemistry in Martinsried.

2.2.4 Transformation of competent *E. coli* cells

2.2.4.1 Making chemically competent cells

Many factors were shown to improve efficiency of transformation. Such factors like: prolonged incubation with $CaCl_2$, addition of cations such as Mg^{2+}, Cs^{2+} or Rb^{2+}, treatment with dimethyl sulfoxide, polyethylene glycol, dithiothreitol, and

hexaminecobalt in presence of both monovalent and divalent cations, were reported to increase yields of transformation (Chung et al., 1989).

Preparation of chemically competent cells was performed by rinsing bacteria harvested during the log-growth phase with the ice-cold 100 mM $MgCl_2$, followed by resuspension in 100 mM $CaCl_2$ supplemented with 15% glycerol. Subsequently cells were dispensed into aliquots, flash frozen in liquid nitrogen and stored at -80°C.

2.2.4.2 Transformation of chemically competent cells

Transformation of competent *E. coli* cells was performed using the heat shock method. For each transformation, 5 µl of ligation reaction or ca. 1 µl of plasmid DNA was added to 50 µl of competent cells. The mixture was incubated for 20 min on ice, followed by a heat shock at 42°C for 1 min, 2 min short cooling on ice and addition of 300 µl of the LB medium. The cells were than incubated for 1 h at 37°C for sufficient time to be given for expression of antibiotic resistance genes. After incubation cells were plated on the LB-agar plates supplemented with appropriate antibiotic and incubated overnight at 37°C.

2.2.5 Protein chemistry methods and techniques

2.2.5.1 Protein expression in *E.coli*

Expression in E. coli provides proteins either in a soluble form or in inclusion bodies. The aim is to obtain maximum amount of a well-folded protein in a soluble fraction. This requires optimization of expression conditions by changing temperature and time of bacteria cultivation, the induction OD, and concentration of inducer (IPTG). Often proteins are expressed as insoluble inclusion bodies but can be refolded back to their native forms. Refolding requires various strategies and was performed in this study according to known protocols.

2.2.5.2 Sonication

Sonication is a simple method used for the disruption of cells by ultrasounds. The high frequency sound waves induce formation of cavitations in the solution, the collapse of which produces sheer forces to break cells. The advantage of this method for cell disruption over freezing and thawing, lysozyme or French press, is comparatively low viscosity of the lysate due to the nucleic acid fragmentation.

Sonication was performed on ice, in 3-4 steps of 3 min each, with 5 min intervals between steps. Pulse mode: output control 8, 60% duty cycle.

2.2.5.3 Protein purification

Proteins expressed in soluble and insoluble forms contained N-terminal polyhistidine tags and therefore could be purified on the Ni-NTA affinity chromatography. The polyhistidine tail of the protein, containing imidazole side chains, chelates reversibly divalent cations (Ni^{2+}, Co^{2+}) attached to the resin. Nickel ions are chelated by nitrilotriacetic acid (NTA), which has four chelation sites for nickel and is covalently coupled to Sepharose resin. Protein can be subsequently dissociated from Ni-NTA resin by applying the elution buffer containing high concentration of imidazole (100-250 mM).

Purification protocols differed substantially between denaturing and native conditions. In case of denaturing conditions, the pellet obtained after centrifugation (9000×g, 30 min) was washed with the PBS containing detergent (Triton X-100) and solubilized in the binding 6 M guanidinium chloride buffer overnight at $4^{o}C$. Afterwards the suspension was centrifuged (20000×g, 30 min) to remove plasma membranes and cell debris. The resulting supernatant was incubated with Ni-NTA resin (Qiagen) for 1 h at $4^{o}C$, subsequently loaded on an empty column, and washed with the washed buffer. The protein was eluted with the buffer of lower pH, than pooled and concentrated. Subsequently, the refolding by rapid dilution was performed. For this purpose, the protein was stepwise diluted in a suitable refolding buffer in a 1:100 volume proportion. The refolding mixture was left with stirring at $4^{o}C$ for 2 days. Finally, the mixture was concentrated and subjected to gel filtration.

In the case of native conditions, the supernatant obtained after centrifugation was mixed with Ni-NTA or glutathione resin preequilibrated with appropriate buffer and incubated for 1-2 h at 4°C with gentle agitation. Subsequently, mixture was loaded on a column, and washed with the wash buffer, and protein was eluted with a buffer containing high concentration of imidazole or in cases of proteins containing the GST-tag, with the buffer containing reduced glutathione. The fractions containing desired protein were pooled, concentrated and subjected to gel filtration.

2.2.5.4 Gel Filtration Chromatography

In the final step of protein purification, gel filtration was performed. The gel filtration column separates proteins according to differences in their sizes. They pass through the gel containing pores of different sizes. Larger particles cannot enter as many pores as smaller and they are eluted earlier.

Before protein purification, column HiLoad 16/60 Superdex s30 (protein separation range < 10 kDa), Superdex s75 (protein separation range 75 – 5 kDa) or Superdex s200 (protein separation range 250 – 20 kDa) connected to the ÄKTA explorer 10 was equilibrated with the gel filtration buffer. The protein sample was concentrated to a very small volume (4-6 ml) using Amicon 8050. The separation was performed at a flow rate 0.7 ml/min in an appropriate buffer. Protein fractions of 1.5 ml were collected according to measurements of A$_{280}$.

2.2.5.5 Protein electrophoresis under denaturing conditions (SDS-PAGE)

The SDS polyacrylamide gel electophoresis was routinely used at various stages of purification to analyze protein purity and identity. Due to small size of the expressed proteins, tricine gels were prepared (Schagger and von Jagow, 1987) with stacking, intermediate and separation layers. Prior to loading on a gel guanidinium HCl-free, protein samples were mixed with 5 × sample buffer in a 4:1 (v/v) ratio and incubated at $100^{o}C$ for 5 min. Samples after the Ni-NTA chromatography under denaturing conditions were prepared in a following fashion: 20 µl of the protein solution was diluted with 400 µl of 20% trichloroacetic acid (TCA) and incubated on ice for 20 min. The sample was centrifuged for 15 min and the resulting pellet was washed by vortexing with ice-cold ethanol. After centrifugation and ethanol removal, pellet was resuspended in a sample buffer and loaded on gel. Electrophoresis was performed at 100 V for the separation gel, 120 V for an intermediate gel, and 150 V for the stacking gel. After electrophoresis, the gel was stained in a Coomassie-blue solution and visualized by incubation in the destaining solution. Both processes were greatly accelerated by brief heating with microwaves of the gel submerged in an appropriate solution.

2.2.5.6 Western Blot

Western blot is a technique for detecting specific proteins out of a number of proteins. In this study the semi-dry type of Western Blot was performed. The assay begins with running a desired sample on SDS-PAGE. Simultaneously a

nitrocellulose membrane and the Watmann paper was cut to the size of the SDS-PAGE gel and soaked in the transfer buffer. All was assembled as a 'sandwich' over the electroblot in following order: three wet Watmann papers, wet nitrocellulose membrane, SDS-PAGE gel, and three wet Watmann papers. The transfer was run at 100 V for 45-60 min. Afterwards, the membrane was incubated in the blocking solution for 1 h, at room temperature with constant shaking. After blocking, the membrane was washed three times with the wash buffer and incubated for 1 h in the 1st antibody solution. The membrane was washed three times with the wash buffer and incubated in the 2nd antibody solution for 1 h at room temperature. Again the membrane was washed three times and the Western blot was developed by incubating in the substrate (BCIP) solution for 30 s.

2.2.5.7 Determination of protein concentration

If needed, protein concentration was measured spectrophotometrically at 280 nm or using the colorimetric Bradford assay. Absorption at 280 nm was measured and converted to a protein concentration on basis of theoretical extinction coefficients. It has been shown that it is possible to estimate the molar extinction coefficient $E\lambda$ of a protein from its amino acid composition (Gill and Hippel, 1989) based on molar extinction coefficients of tryptophan, tyrosine and cystine (cysteine residues do not absorb appreciably at wavelengths >260 nm, while cystines do). At a given wavelength λ the extinction coefficient of a protein can be calculated using the following equation:

$$E\lambda(Prot) = Numb(Tyr)\times Ext\lambda(Tyr) + Numb(Trp)\times Ext\lambda(Trp) +$$

$$Numb(cystine)\times Ext\lambda(cystine)$$

The absorbance can be computed using the formula:

$$A\lambda \ (Prot) = E\lambda(Prot) \ / \ Molecular \ weight$$

In order to measure concentration of protein labeled with rhodamine, additionally absorbance at 555 nm was measured and put to the following formulas:

$$Protein \ concentration \ (M) = [A_{280} - (Amax \times CF)]/ \ E\lambda prot \times dilution \ factor$$

The Bradford assay was performed by diluting 5 μl of the protein in 1 ml (10 × diluted stock) of the Bradford reagent (BioRad). After mixing in a plastic cuvette, the absorption at 595 nm was measured. The protein concentration was calculated on the basis of the calibration curve made for the bovine serum albumin of a known concentration.

2.2.5.8 NMR spectroscopy

One-dimensional (1D) proton NMR spectra were recorded to investigate proteins' folding correctness. Two-dimensional (2D) NMR HSQC spectra provided information on protein folding, structure and ligand binding.

All NMR spectra were recorded at 300 K on a Bruker DRX 600 MHz spectrometer equipped with a cryoprobehead. Typically, NMR samples contained up to 0.3 mM of protein in the PBS or in PBS pH 7.4 supplemented with 5 mM DTT, 0.02% NaN$_3$ and

10% D$_2$O. For the ^{1}H-^{15}N HSQC spectrum (Mori et al., 1995), a total of 2048 complex points in t_2 and 256 t_1 increments were acquired. Water suppression was carried out using the WATERGATE 5 sequence (Piotto et al., 1992). NMR data were processed using the Bruker program Xwin-NMR version 3.5 and UCSF Sparky (T. D. Goddard and D. G. Kneller). NMR binding experiments NMR ligand binding experiments were carried out typically on 500 µl of the protein sample, at a concentration of 0.2-0.4 mM, by titrating the sample with 10 mM stock of the peptide. Normalized chemical shift perturbations were calculated according to equation:

$$\Delta\delta = \sqrt{0.04\left(\delta_N^{bound} - \delta_N^{free}\right)^2 + \left(\delta_H^{bound} - \delta_H^{free}\right)^2}$$

2.2.5.9 Crystallization and X-ray data collection

Crystallization trials were performed using proteins purified on the preparative gel filtration columns in the buffer typically containing 10 mM Tris pH 8.0, 100 mM NaCl and preferably 5 mM β-mercaptoethanol. The protein was concentrated to 10-15 mg/ml, centrifuged at high rpm briefly, and set up for crystallization screening. Sitting drop vapor technique in 96-well format was used. A typical screen was composed of the following suites: Index (Hampton Research), Classics, pHClear I, pHClear II, PEGs, Anions, Cations (Qiagen) at 4°C and 20°C. The most promising conditions were optimized through manipulation of the crystallization reagent composition and drop size, by using both, sitting drop and hanging drop techniques.

X-ray datasets of the UBL5 (1.9 Å) were collected on the PXII beamline of the Swiss Light Source at the Paul Scherrer Institut (Villigen, Switzerland). Data were indexed, integrated and scaled with the XDS and XSCALE programs (Kabsch, 2010a; Kabsch, 2010b). The structures were determined by molecular replacement using the Molrep program from the CCP4 suite (Collaborative Computational Project 1994).

2.2.5.10 Circular dichroism

CD spectra were recorded on the JASCO J-715 spectrometer. Peptide of concentration of 10 µM in suitable buffers was used in a 0.1 mm cuvette and scanned 4 times in order to reduce random error and noise. The spectrum of the buffer was recorded and subtracted to isolate the changes solely from the peptide. The spectra were recorded using the following parameters: response 5 sec, scanning speed 20 nm/min, bandwidth 1.0 nm, step resolution 0.1 nm, sensitivity 5 mdeg.

2.2.5.11 Microscale thermophoresis

Microscale thermophoresis was used to measure the binding interactions between proteins and peptides (Jerabek-Willemsen et al. 2011). Titration series with constant peptide concentration and varying protein concentrations was prepared in a final solution of 0.05 % Tween-20 and 10 µM BSA in PBS. The final peptide concentration was 50 nM. Approximately 2 µl of each sample was loaded in fused silica capillary covered with hydrophilic polymer. An infrared laser diode was used to create a 0.12 K/µm temperature gradient inside the capillaries [Furukawa FOL1405-RTV-617-1480, wavelength λ ¼ 1;480 nm, 320 mW maximum power (AMS Technologies AG)]. Fluorescein was excited with an diode and was measured with a 40x SUPRASIL synthetic quartz substrate microscope objective, numerical aperture 0.8 (Partec). The Hill equation (n ¼ 2) was fit to the data to determine the EC50 value for each sample; the EC50 value is the concentration at which half of the protein sample is bound to its ligand.

2.2.6 Chemistry methods and reactions

2.2.6.1 Estrification reaction

Estrification of the carboxylic group was carried out as described previously by (Neises and Steglich, 1978). To a stirred solution of 200 µmol of WK2-3A in 0.5 ml DMF was added 3 mg DMAP and 20 mg 1,6-hexanodiol. 15 mg DCC was added to the reaction mixture at 0°C and after 5 min the reaction was maintained for 3 h at

20°C. The reaction mixture was evaporated in vacuum and the residue was dissolved in solution of acetonitrile (50%) and purified by the gel reverse phase chromatography (HPLC).

2.2.6.2 Preparation of N-hydroxylsuccinimide derivative

The N-hydroxylsuccinimide derivative of WK2-3A compound was carried out according to the procedure described by Luo and Prestwich (1999). To the stirred solution of compound 1 (4 mg) and succinic anhydride (3 mg) in CH$_2$Cl$_2$ (2 ml) at room temperature, the dry pyridine (16 µl) was added. The reaction was allowed to proceed for 3 days at room temperature and then concentrated in vacuum. The rest was dissolved in a solution of acetonitrile-water (1:1) and was purified by reverse phase chromatography.

2.2.6.3 Preparation of LMW hyaluronan

To solution of 10 mg of the high molecular mass HA (Sigma) in the PBS pH 6.5 buffer (4 mg/ml) was added 1.5 mg hiarunidase (300U/mg, Roth). The digestion was carried out at 37°C, 180 rpm stirring for 1 h and then deactivated for 20 min at 98°C. The mixture was subsequently centrifuged and loaded on the gel filtration column (Superdex s 75) and separated in 0.1 M NH$_4$HCO$_3$. The desired fractions of the LMW hyaluronan were pooled, flash frozen and lyophilized (Luo and Prestwich, 1999).

2.2.6.4 Preparation of hyaluronic acid coupled with adipic dihydrazide

The synthesis of HA-ADH was performed according to two different protocols:

1. The low molecular weight HA was dissolved in water to give a concentration of 3.5 mg/ml and then the 30-fold molar excess of ADH was added into the solution. The pH of the reaction was adjusted to 6.8 with 0.1 N NaOH/ 0.1 N HCl. One equivalent of ECDI was added in solid form followed by one equivalent of 1-hydroxybenzotriazole in DMSO-H_2O (1:1). The pH of reaction mixture was maintained at 6.8 and the reaction was allowed to proceed overnight (Auzenne et al., 2007).

2. The LMW HA was dissolved in water to give a concentration of 3.5 mg/ml. A 5-fold excess of ADH was added into the solution and the pH of the reaction was adjusted to 4.75 by the addition of 0.1 N HCl. Next, one equivalent of EDCI was added in the solid form and the pH of the reaction was maintained at 4.75 and proceeded overnight (Luo and Prestwich, 1999).

Both reactions were quenched by the addition of NaOH to adjust the pH to 7.0. The mixtures were then transferred to the pretreated dialysis tubing (MW cut off 1000) and dialyzed exhaustively against 100 mM NaCl, and 25% ethanol, and finally H_2O. The solutions were filtrated through the 0.2 μm cellulose acetate membrane, flash frozen and lyophilized (Auzenne et al., 2007; Luo and Prestwich, 1999).

3 Results and Discussion

3.1 CD44

3.1.1 Constructs, expression and purification conditions

The CD44 gene was fished out from the human cDNA of the breast cancer cell line MDA-MB-468. Various constructs were prepared and cloned into one of the pET vectors (Novagen) using different restriction nucleases. Primers used in preparation of the constructs are shown in Table 3.1.1.

Table 3.1.1 List of primers used in the work

No.	Sequence	Name
1	ATGGACAAGTTTTGGTGGCACGCAG	full-length for
2	TTACACCCCAATCTTCATGTCCACATTC	full-length rev
3	GGAATTCCATATGCTGGCGCAGATCGATTTGAATAT	CD44 19 for
4	GGAATTCCATATGGCGCAGATCGATTTGAATAT	CD44 20 for
5	GGAATTCCATATGCAGATCGATTTGAATATAACC	CD44 21 for
6	CATGCCATGGCGCAGATCGATTTGAATATAACC	CD44 20 noHis
7	TGGAAGATCTTCATCTGCTATTCTCCTTTCTGGACATAGCG	CD44 162 rev
8	TGGAAGATCTTCACACGTCATCATCATAGGG	CD44 178 rev
9	TGGAAGATCTTCAGCTCACGTCATCATCAGTAGGG	CD44 179 rev
10	TGGAAGATCTTCAGCTGCTCACGTCATCATCAGTAG	CD44 180 rev

For: forward, rev: reverse, restriction sites underlined, stop codons in italics

Expression and purification of a hyaluronan-binding domain of CD44 was performed according to the modified protocol established by (Banerji et al., 1998). For all experiments we chose the CD44 amino acid 21-178 construct (CD44 21-178) in the pET19b, because of its relatively good expression level. Constructs' residue coverage together with their properties are listed in Table 3.1.2.

Table 3.1.2 List of constructs produced in the work.

No.	Name	Properties	Vector	Primers used
1	CD44 (19-178)	low expression	pET19b	3,8
2	CD44 (20-178)	low expression	pET19b	4,8
3	CD44 (21-178)	moderate expression	pET19b	5,8
4	CD44 (20-178) noHis	low purity	pET19b	6,8
5	CD44 (21-162)	no expression	pET19b	5,7
6	CD44 (21-179)	moderate expression	pET16b	5,9
7	CD44 (19-180)	low expression	pET19b	3,10

CD44 was obtained only in insoluble (inclusion bodies) fractions. The protein was purified by the NiNTA chromatography under denaturing conditions and refolded in a standard refolding buffer. The buffer additives normally fall into two categories: aggregation suppressors and refolding enhancers. The buffer contained arginine, which like proline, mild detergents or polyethylene glycol, acts as an aggregation

suppressor through binding to the hydrophobic regions of the folding intermediate. Crucial for functional refolding is also adequate redox system, which enhances the correct disulfide bond formation, being the limiting step during folding process. Reduced and oxidized thiol reagents, comprising a suitable redox couples, allow disulfide reshuffling i.e. wrong disulfide bridges will be reduced again because they are not protected by the correct structural context. One of the most popular redox systems that comprises the reduced and oxidized glutathione (GSH/GSSG) was used in the refolding protocol.

The CD44 construct was expressed in the E. coli BL21-CodonPlus RIL (DE3) and the protein was detected within 4 h at 37°C after induction with 0.4 mM IPTG. 5 h after induction, the cells were harvested by centrifugation and resuspended in the phosphate-buffered saline (PBS). The cell suspension was incubated on ice and disrupted by sonication. The cell lysate was centrifuged and the resulting inclusion body pellet was washed three times in ice-cold Triton wash (0.5% Triton X-100, 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.02% NaN_3), and subsequently washed three times with the same buffer but without detergent. After final centrifugation the body pellet was solubilized in 10 ml of the buffer comprising 6 M guanidnium chloride and a reducing agent through vigorous stirring overnight at 4°C. The solution was cleared by centrifugation and applied onto the NiNTA resin. Protein was bound to the resin by agitation for 2 h at 4°C. Afterwards, the slurry was applied on the column and washed twice with guanidinium chloride of lower pH, firstly supplied with β-mercaptoethanol, and secondly, with a wash buffer deprived of the reducing agent. Elution of the His-tagged protein was performed by the guanidinium chloride solution

of pH 4.0. Under these conditions the His-tagged protein can no longer bind nickel ions and will dissociate from the resin. The low pH prevents the creation of disulphide bridges through sulphydryl group protonation.

Pellet resuspension
10 ml PBS/1l culture

↓

Cell disruption
Sonication, done on ice

↓

Centrifugation
4000 rpm, 10 min

↓

Pellet solubilization
6M guanidinium chloride, overnight at 4°C

↓

Centrifugation
22000 rpm, 25 min

↓

NiNTA chromatography

↓

Refolding
Rapid dilution technique

↓

Concentration
Amicon, 10 kDa MWCO

↓

Size exclusion chromatography
p75 prep grade
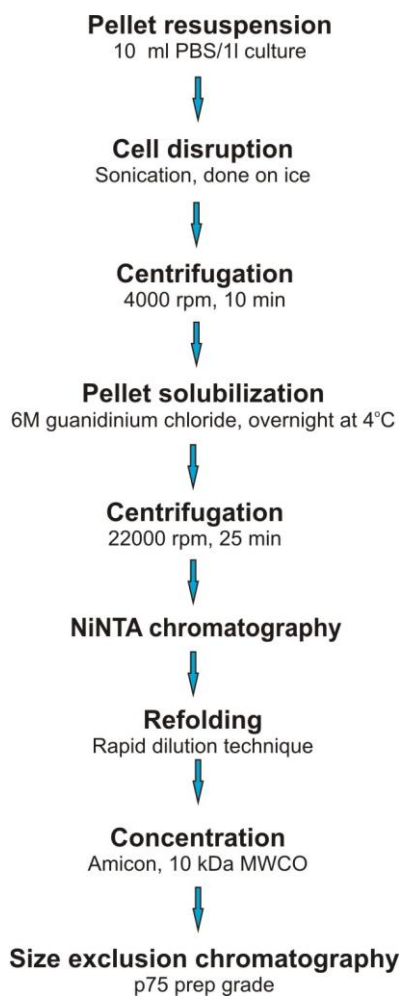
Figure 3.1.1 Flow chart of the purification scheme for CD44 constructs.

Refolding was performed by rapid dilution technique and the protein was injected stepwise into the buffer containing 100 mM Tris-HCl (pH 8.0), 400 mM L-arginine, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione, 0.5 µg/ml

leupeptin, taking care of keeping a constant end concentration of guanidinium chloride (200-fold dilution) and refolded at 4°C for 48 h. After completed refolding, protein was concentrated on Amicon and applied onto the gel filtration column (HiLoad 26/60 Superdex 75 pg) equilibrated with an appropriate buffer in order to isolate refolded monomers.
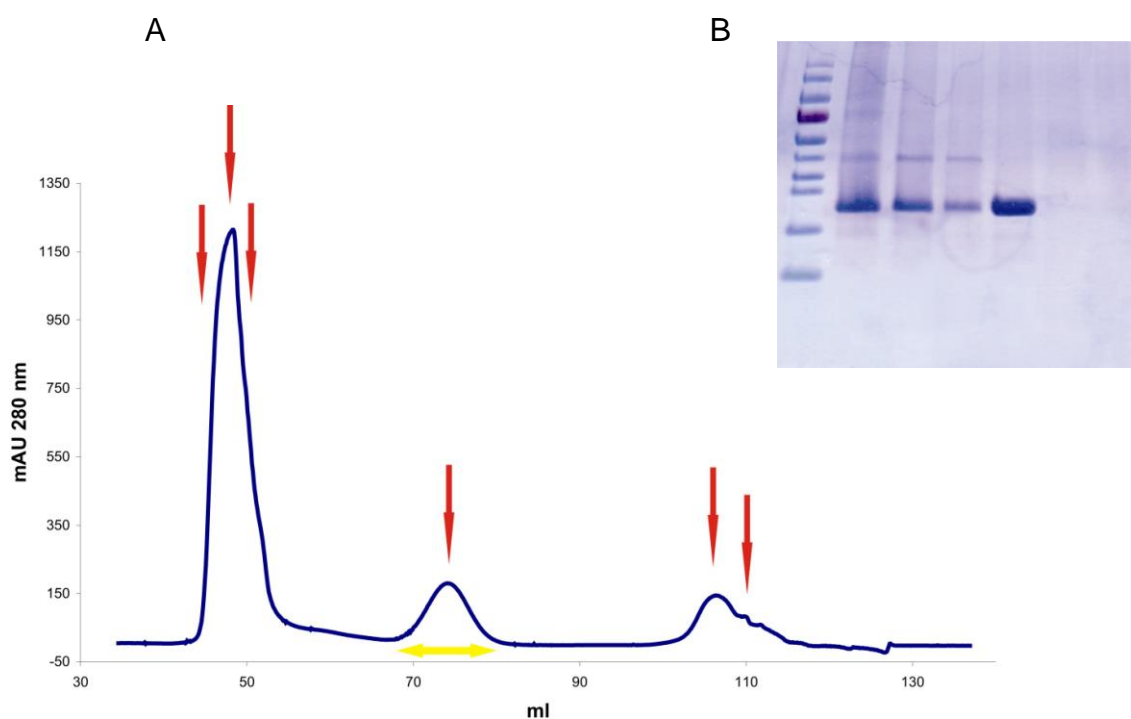
Figure 3.1.2 (A) Gel filtration of CD44 after refolding. Red arrows indicate fractions selected for electrophoresis; yellow arrow indicates pooled fractions of the desired monomer of CD44. (B) SDS-PAGE image of chosen fractions. Lane 1 – protein marker, lanes 2-7 selected fractions.

As show on Figure 3.1.2 A, the majority of the refolded material comprised higher molecular aggregates. A fraction corresponding to the monomer molecules was analyzed by mass spectroscopy and yielded a value of 20147.6, which is in

agreement with the calculated molecular mass of the protein lacking the N-terminal

methionine and having three disulfide bonds.

3.1.2 Structural analysis and functional properties

The structural integrity of the CD44 construct was confirmed using one-

dimensional proton NMR spectroscopy.
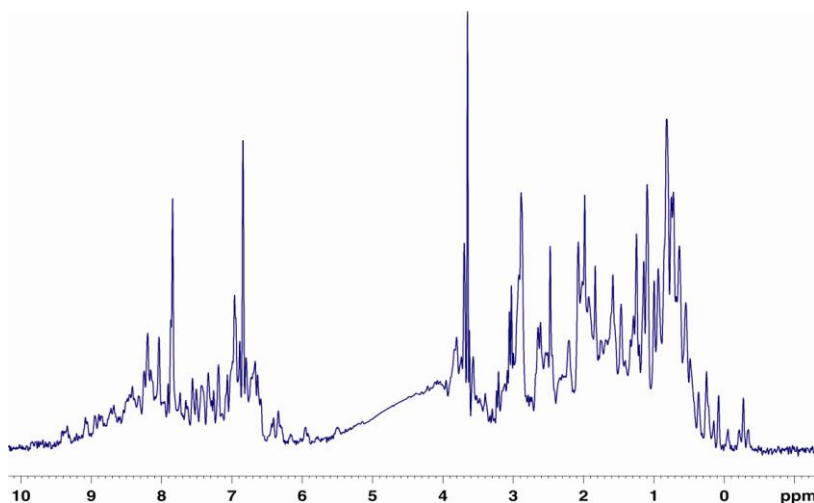
Figure 2.1.3 NMR spectrum of CD44 (21-178).

The ^{1}H NMR spectrum in Figure 1.1.3, which was acquired at neutral pH, shows

sharp, well-dispersed peaks, consistent with a homogenous, folded protein. The final

yield of the refolded monomeric CD44 is 2 mg/L of the bacterial culture, while the

majority of the material is lost as disulfide-linked aggregates (dimers, trimers, and

higher order species) during the refolding process. As a final test for the correct folding of the purified CD44 monomer, we analyzed the hyaluronan binding capacity of the protein. We used NMR spectroscopy to check the interaction of hyaluronan with CD44. NMR measurements consisted of monitoring changes in NMR chemical shifts and line widths of the backbone amide resonances of ^{15}N-enriched samples upon addition of an unlabeled ligand (Meyer and Peters, 2003). The ^{15}N-uniformly labeled CD44 was produced and mixed with the short-chain HA_8 oligosaccharide, in molar ratio 1:3, respectively. The ^{15}N-^{1}H HSQC spectrum of the mixture was recorded on a 600 MHz spectrometer.
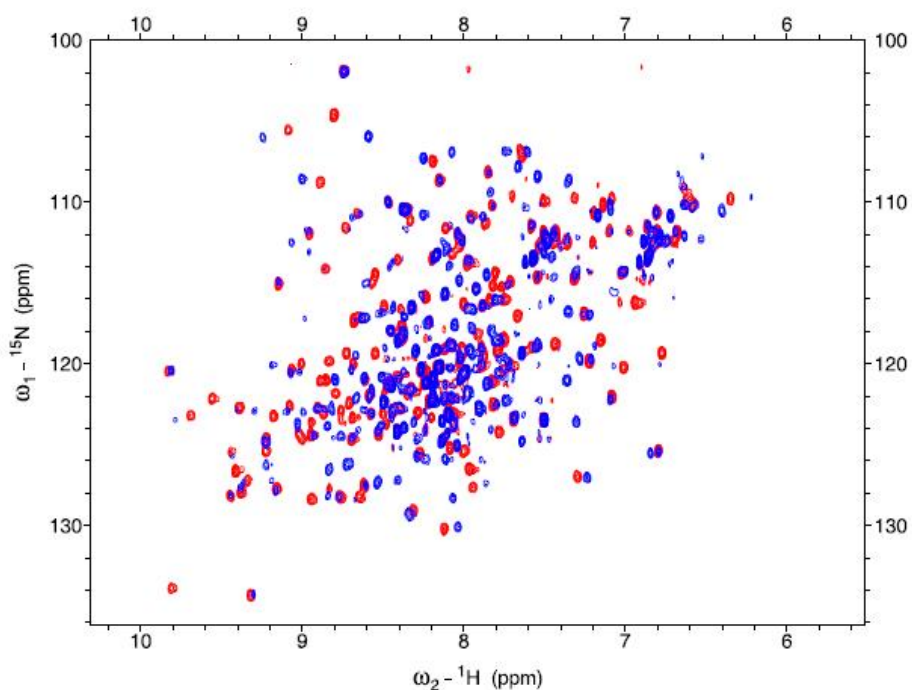
Figure 3.1.4 Superimposed ^{15}N-^{1}H HSQC spectra of the hyaluronan binding domain of CD44 (21-178). The reference spectrum is shown in red, and the one in blue is after titration with the HA_8 oligosaccharide.

In the spectrum, we can clearly see the interaction of the ^{15}N-labeled CD44 with HA_8 oligosaccharide (Figure 3.1.4). We detected changes in NMR chemical shifts of almost all amino acids upon addition of HA_8. These results show that the recombinant domain of CD44 in our study is functional and active.

3.1.3 Preparation of hyaluronic acids of defined lengths

Hyaluronan molecules from *Streptococcus equi* (Sigma) were partially degraded by the ovine testicular hyaluronidase (Roth). This enzyme degrades HA to generate a series of even-numbered HA molecules with N-acetylglucosamine moiety at the reducing terminus (Kreil, 1995). To a solution containing 100 mg of HA in 30 ml of 0.1 M sodium acetate, 0.1 M NaCl, pH 6.0, 20 mg (300 U/mg) of hyaluronidase was added, and enzymatic digestion was performed at 37°C. The incubation time of hyaluronidase varies according to the sizes of the HA oligosaccharides to be obtained. The reaction was stopped after 1 h by boiling for 20 min. The sample was centrifuged at 16,000 rpm for 30 min, and the supernatant was applied onto the Superdex s30pg column, and eluted with 0.1 M NH_4HCO_3. The fractions corresponding to the UV-positive peaks at 206 nm were lyophilized. Subsequently, samples were dissolved in distilled water and analyzed by the ESI-TOF mass spectrometry. The average molecular masses as well as formulas of the molecules arising from HA_4 to HA_{14} are shown in Table 1.1.3.

Table 1.1.3 Theoretical average molecular masses of HA oligosaccharides

HA	M (Da)	Formula
4	776.66	$C_{28}H_{44}N_2O_{23}$[GlcUA-GlcNAc]$_2$
5	951.77	$C_{36}H_{57}N_3O_{27}$[GlcUA-GlcNAc]$_2$-GlcUA
6	1155.98	$C_{42}H_{65}N_3O_{34}$[GlcUA-GlcNAc]$_3$
7	1331.09	$C_{50}H_{78}N_4O_{39}$[GlcUA-GlcNAc]$_3$-GlcUA
8	1535.30	$C_{56}H_{86}N_4O_{45}$[GlcUA-GlcNAc]$_4$
9	1710.41	$C_{64}H_{99}N_5O_{45}$[GlcUA-GlcNAc]$_4$-GlcUA
10	1914.62	$C_{70}H_{107}N_5O_{56}$[GlcUA-GlcNAc]$_5$
12	2293.94	$C_{84}H_{128}N_6O_{67}$[GlcUA-GlcNAc]$_6$
14	2673.26	$C_{98}H_{157}N_7O_{78}$[GlcUA-GlcNAc]$_7$

As we can see from Figure 3.1.5, a series of well separated and symmetrical peaks were eluted from the column. These were analyzed by MS and some found to correspond to the HA molecules. Individual oligosaccharides on this column were well resolved up to HA$_8$, while oligosaccharides larger than HA$_8$ showed clusters of several sizes, as analyzed by the ESI-TOF mass spectrometry. For example, fraction 2 was shown to contain significant amounts of HA$_{10}$, HA$_9$, HA$_8$, and HA$_7$, whereas, fraction 3 is predominantly composed of HA$_8$.
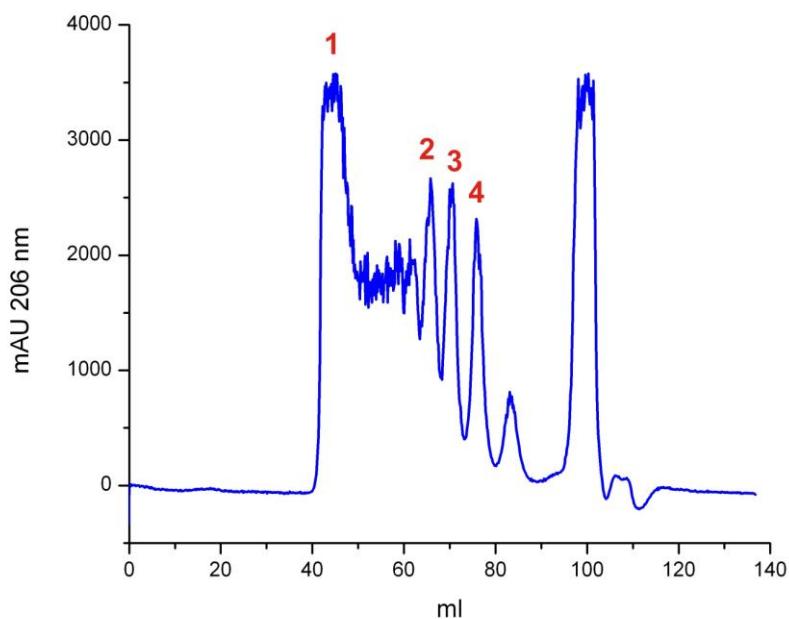
Figure 3.1.5 Elution profile of the HA digest. The labeled fractions under UV absorbing peaks were pooled and analyzed by the ESI-TOF mass spectrometry.

ESI-MS was used in the positive mode to analyzed purified oligomers, and acquired masses of molecules in charged states are shown in Table 3.1.4. In addition to the even-numbered oligomers (i.e. HA_6, HA_8 and HA_{10}), odd-numbered oligomers, such as HA_7 and HA_9, were obtained. These fragments arise from the loss of the GlcNAc residue from the reducing end of HA_8 and HA_{10} respectively, as previously reported (Mahoney et al., 2001). These species have molecular masses that are consistent with saturated 7- and 9-mers having nonreducing and reducing GlcUA residues.

Table 3.1.4 ESI-MS analysis of peaks from size exclusion chromatography of digested HA

Peak No.	Oligosaccharide(s) present	Theoretical masses (Da)	Experimental masses (Da)		
			$[M+H]^+$	$[M+2H]^{2+}$	$[M+Na]^+$
1	Cluster of several sizes, identification not possible				
2	HA_{10}	1914.62	1914.42		1937.40
	HA_9	1710.41			1735.86
	HA_8	1535.30	1535.32		1557.30
	HA_7	1330.09			1354.36
3	HA_8	1535.30	1535.47	768.24	1557.45
4	HA_6	1155.98	1156.35	578.68	

In positive ion mass spectra, the major species were generally a mixture of metal ions adducts, such as $[M+Na]^+$, and monovalent and divalent cations. As shown in the Table 3.1.4, the experimental masses obtained are in reasonable agreement with the theoretical molecular weights. Any slight differences are believed to arise from the method of calibration. Figure 3.1.6 shows a typical positive ion spectrum of the HA_8 oligosaccharide, which was obtained for peak 3 purified by size exclusion chromatography. This molecule was used to characterize protein-HA interactions with a hyaluronan-binding domain of CD44.
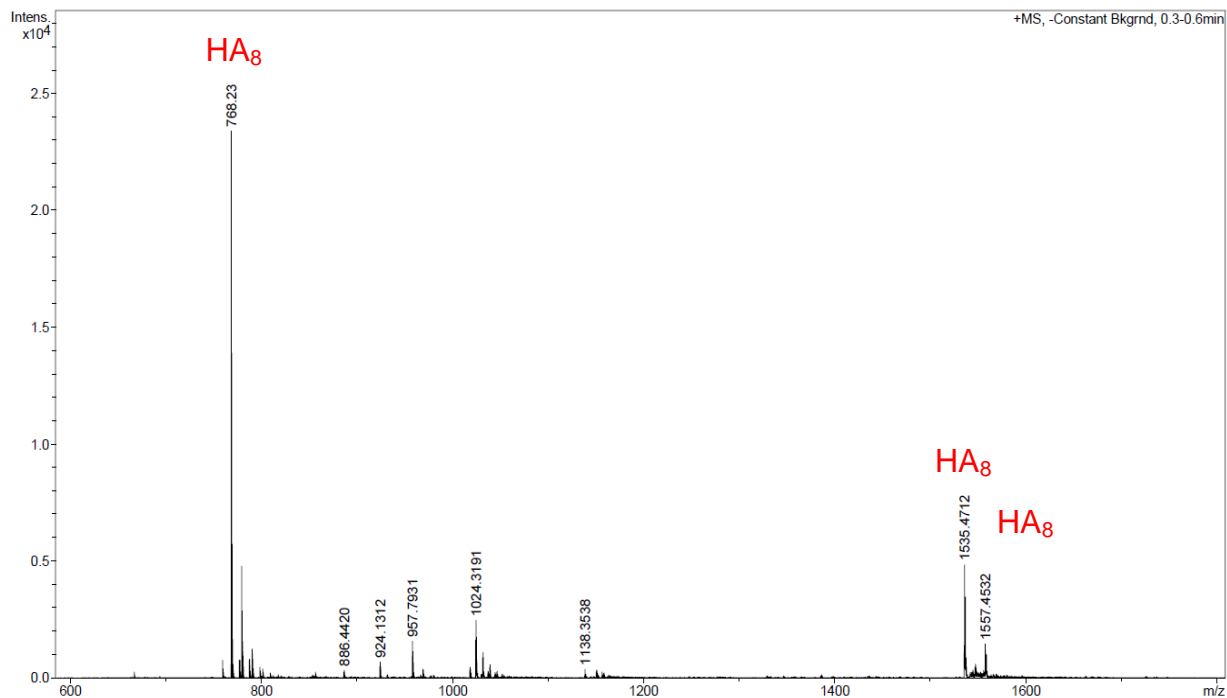
Figure 3.1.6 The positive ion ESI-MS spectrum of HA8. Charged species of the formula [M+H]+, [M+2H]2+, and [M+Na]+ are formed, where M is the theoretical molecular weight.

Purification of low-molecular-weight HA oligomers were thus successful. If needed, purity of the samples can be improved by re-running the samples on the ion exchange chromatography. For experiments in cell biological and structural studies, it is essential to obtain well-characterized and highly purified HA oligosaccharides - so that erroneous results are avoided (Camenisch and McDonald, 2000). For example, production of HA oligomers free from chondroitin sulfate (ChS) has been shown to be of importance in studies on role of HA in malaria disease (Beeson et al., 2000). ChS (as well as HA) is a ligand recognized by the Link module of CD44 (Kawashima et al., 2000). Contaminations of this glycosaminoglycan can be avoided by the use of the medical-grade or bacterial HA, which contains very low levels of

chondroitin sulfate, rather than the relatively impure samples from, for example, from the human umbilical cord (Mahoney et al., 2001). In our studies we used HA from *Streptococcus equi,* which is supposed to contain very small amounts of this contaminant. In addition, the HA preparation should be treated with deoxyribonuclease 1 to remove DNA contaminations, which also lead to false results in cell biology assays (Filion and Phillips, 2001). The purity factor is important in our further studies with hyaluronan used as a drug carrier.

3.1.4 Screening for CD44

The CD44 protein, including some of its variants, is vital to many cell functions. In animal models targeting CD44 or its ligand can reduce or even eradicate the malignant activity (Naor et al., 2002). CD44 is involved in a crosstalk with the tumor surroundings, it coordinates activity within the signal transduction pathways (through oncogene receptors), and it stabilizes multidrug-resistant proteins genes. It is thus expected that a blockade of CD44 should exert multi-functional therapeutic effects (Zoller, 2011).

The aims of the project were small molecular weight inhibitors of the cancer relevant protein. This would be a major advancement to the cancer research field since CD44 is believed to have a major impact on cancer cell growth and no compounds inhibiting this protein are currently known. In order to develop small molecules that target CD44, we carried out the in silico screening with the CD44 as the target protein. For the screening a PDB structure of the mouse CD44 in complex with HA_8

(Banerji et al., 2007) was used. Water molecules and sugar moieties, not interacting directly with the binding cleft of CD44, were removed. The binding cleft was defined as a 10 Å sphere radius around the hyaluronan ring bound in the cleft (NAG A1177). The ZINC database of purchasable compounds was used for docking. The ligands were prepared from their SMILES descriptor using Maestro Ligprep/Epik modules and stored as a single-file SDF library. The filtered library contained 2×10^7 ligands. Ligands with stereo centres were represented as multiply molecules. Protonation state of the ligand was calculated for pH 7.0. The library was docked to the receptor using the eHits Software (Zsoldos et al., 2007) working in a standard precision mode. No additional restraints were imposed during the docking.
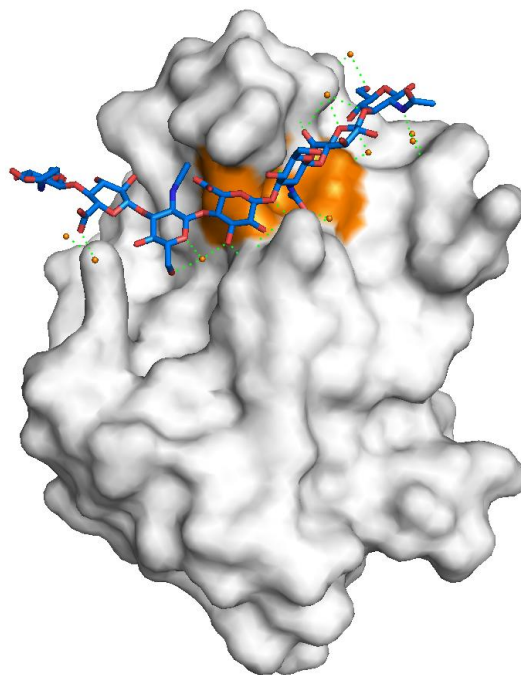
Figure 3.1.7 Surface of CD44 bound to the HA_8 oligosaccharide (type B crystal complex, taken from the PDB file 2JCQ). Orange color indicates docking surface selected for the in silico screening.

The scored docking results were sorted by highest scores and best 768 ligands were used for further filtering. The filtering process accepted only ligands filling the hydrophobic pocket formed by residues Cys81, Tyr83, Ile92, Cys101. Subsequently, only ligands forming at least two hydrogen bonds to the protein were accepted. The filtering process yielded ligands that were inspected manually. Initially, we selected 39 ligands for in vitro testing. The solubility of the chosen compounds was tested in conditions used typically for the NMR screening (10% DMSO, PBS buffer). Finally, we selected 27 "soluble" chemical substances. The formulas of the tested low-molecular-weight compounds are shown in Figure 3.1.8.

27 compounds were tested for binding to CD44 by performing "SAR-by-NMR" method (Shuker et al., 1996; Stoll et al., 2001) - a series of binary titrations of isotopically enriched ^{15}N-CD44 with a compound. The method relies on the use of chemical shift perturbations in 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled proteins upon addition of ligands or peptides/proteins and works best for proteins of small size (i.e., less than 20 kDa). Strong binding of a compound to its target is indicated by splitting of peaks in HSQC spectrum, whereas a shift of peaks shows weaker binding (Wüthrich, 1986).

Figure 3.1.8 Chemical formulas of compounds selected from the in silico screening and tested against CD44.
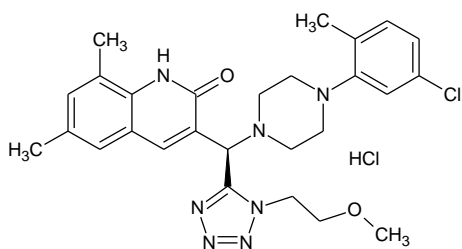
BAS 04419989
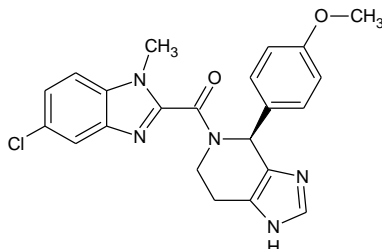
SYN 22991305
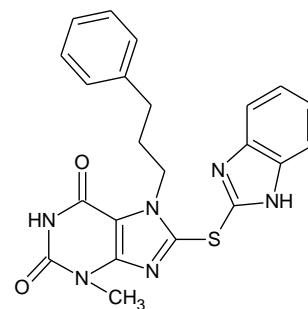
BAS 07401821
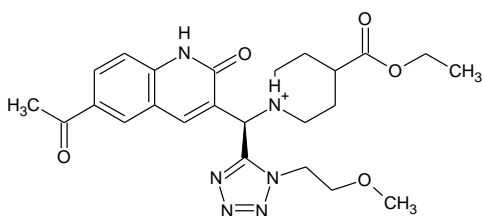
BAS 04421964
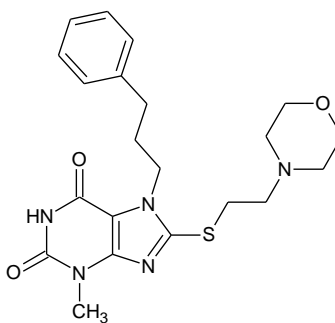
BAS 04914731
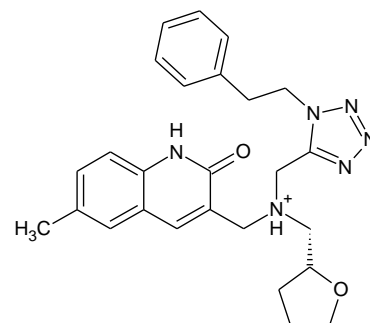
ASN 05303779
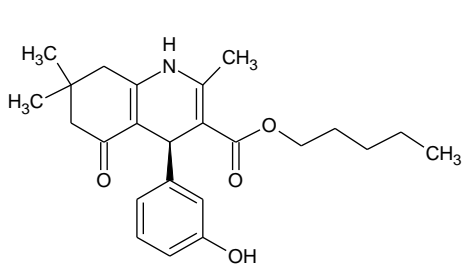
BAS 00410644
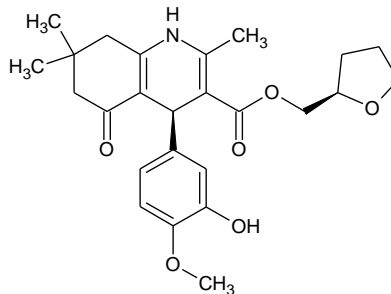
BAS 01404102
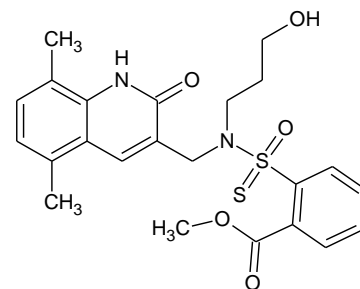
ASN 04193774
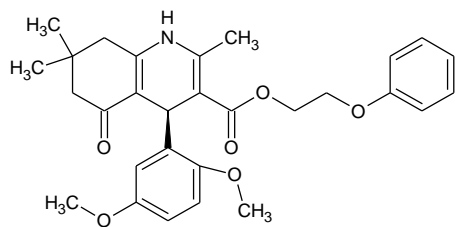
BAS 00915672
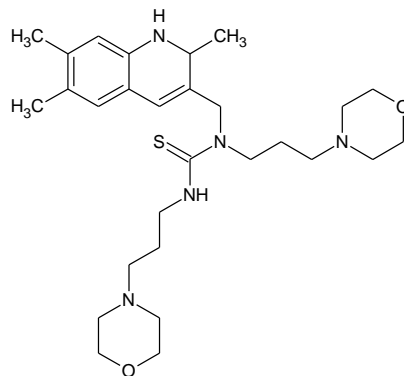
ASN 03365342
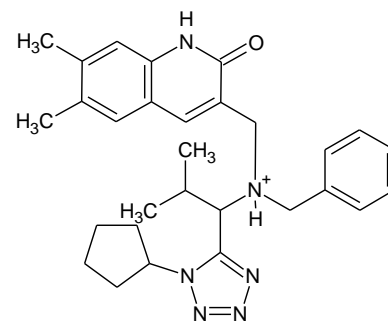
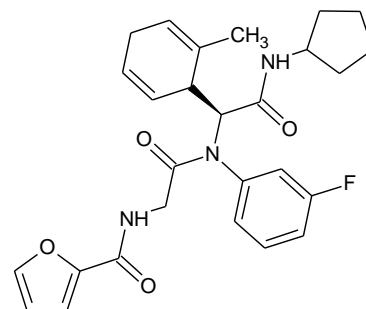ASN 05379832
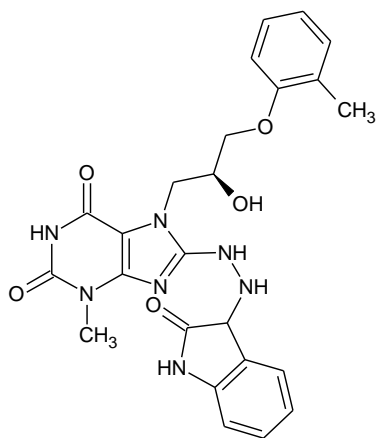
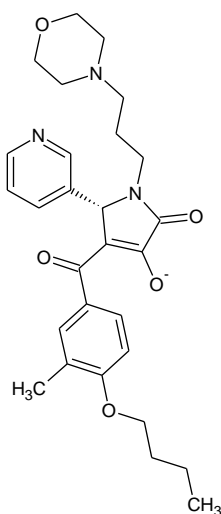BAS 01122590
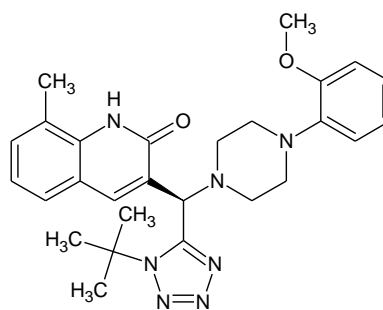
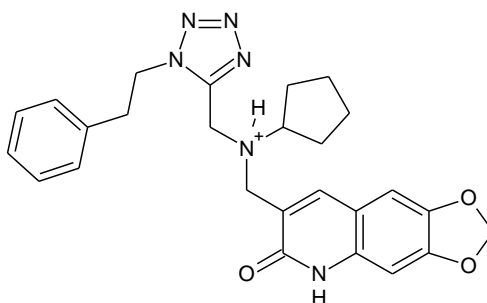ASN 04419549
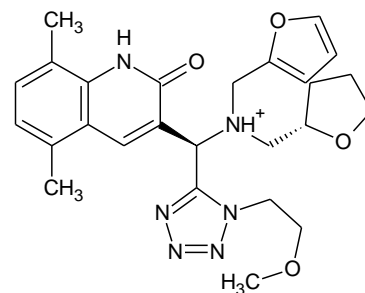
ASN 04420663
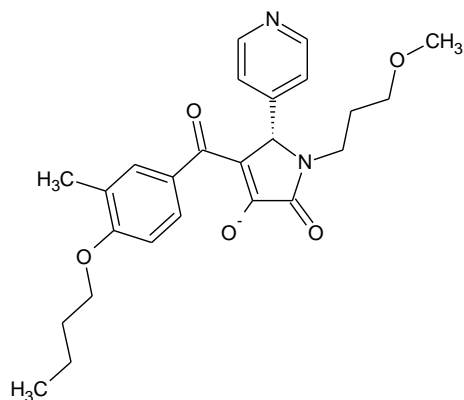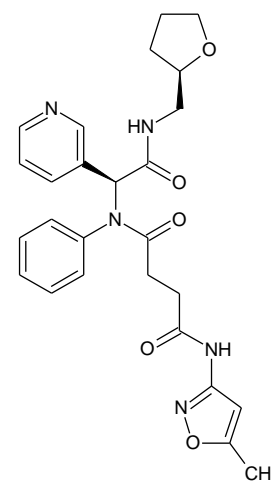
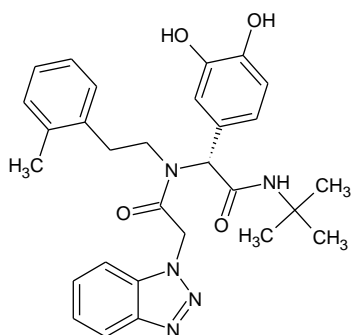BAS 03787957
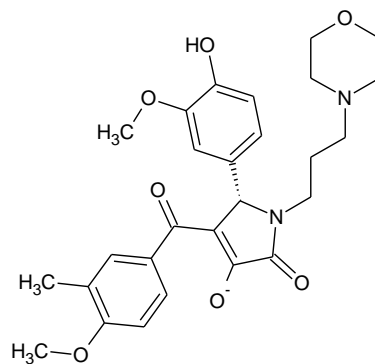
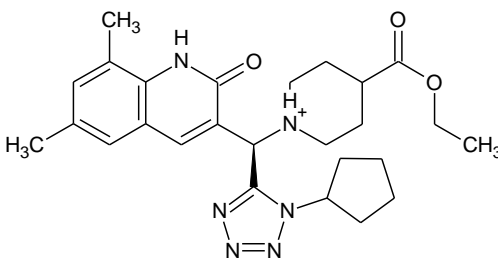BAS 03787662
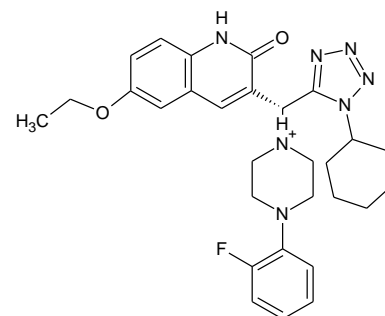
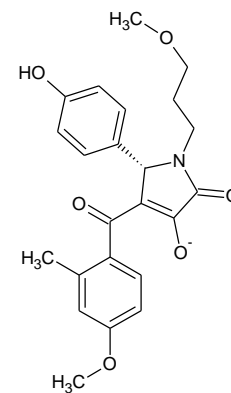ASN 03775618
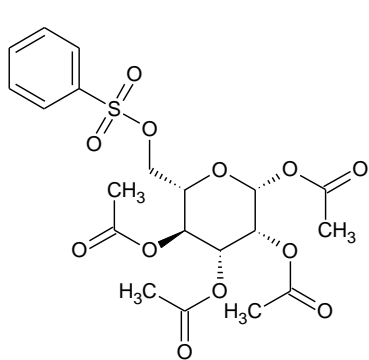
BAS 00231362
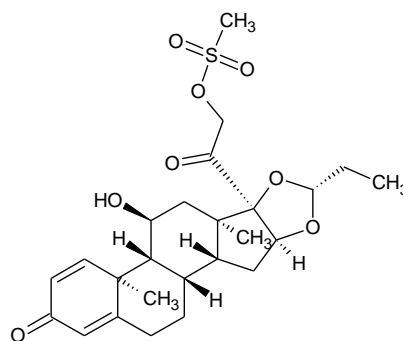
BAS 00670884

ASN 05583775

BAS 03788016

ASN 05303817

ASN 05571597

ASN 04364055

ASN 04188104

BAS 03539828

The NMR assay, unfortunately, did not indicate any binding to the CD44 protein. No shift perturbations patterns were observed after titration with the selected compounds. Figure 3.1.9 shows a typical example of the HSQC spectrum showing no interaction between CD44 and a compound.
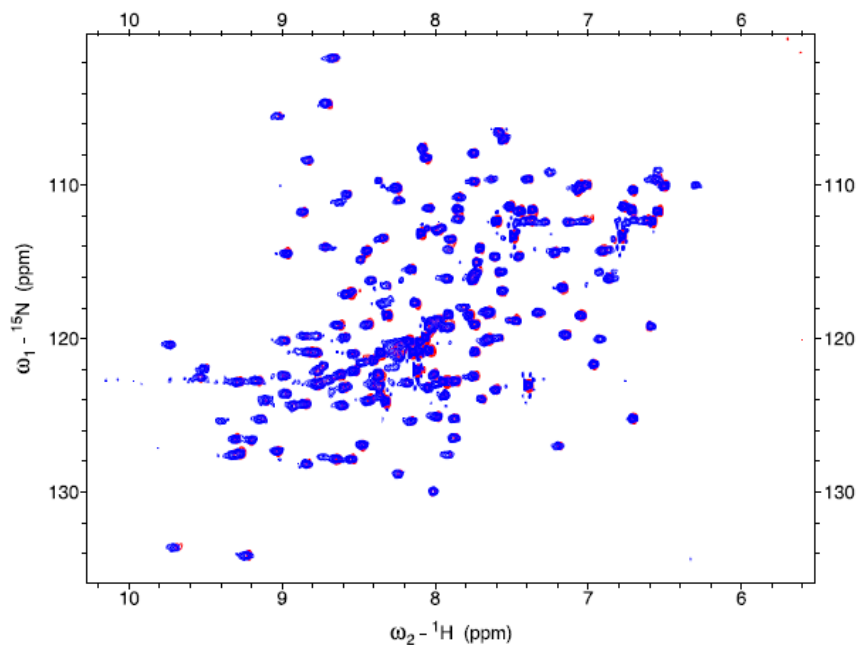

Figure 3.1.9 Example of the screening of CD44 with a low-molecular weight compound. NMR indicates no interaction; superimposed are the ^{15}N-^{1}H HSQC spectra of CD44 mixed with ASN03775618 in ratio 1:3. The spectrum of free CD44 is shown in red and the spectrum of CD44 mixed with the compound is shown in blue.

3.2 Preparation of substrates for the HA-WK2-3A bioconjugate

3.2.1 Synthesis of the WK2-3A-NHS ester

Bioconjugation of hyaluronic acid to small antitumor agents provides advantages of drug solubilization, localization, stabilization, and controlled release (Luo and Prestwich, 1999). The immunoneutrality of HA makes it an excellent building block for development of novel biocompatible and biodegradable materials used in tissue engineering and drug delivery systems (Luo and Prestwich, 1999). In order to improve solubility and specificity of the potential anticancer MDM2-binding molecule, WK2-3A (Popowicz et al., 2010), (provided by Prof. Alexander Doemiling, University of Pittsburgh, USA), we tried to link it to polymers of HA. The aqueous solubility of the compound is quite low because the protein-protein interaction of p53 and MDM2 is very hydrophobic and naturally results in very hydrophobic initial antagonists.

WK2-3A (Figure 3.2.1) was first coupled with a linker, 1,6-hexanodiol, in an esterification reaction. The linker provides a hydroxyl group for further modifications, otherwise the intrinsic carboxyl group of WK2-3A itself would be difficult to activate or any modification could destroy the compound.
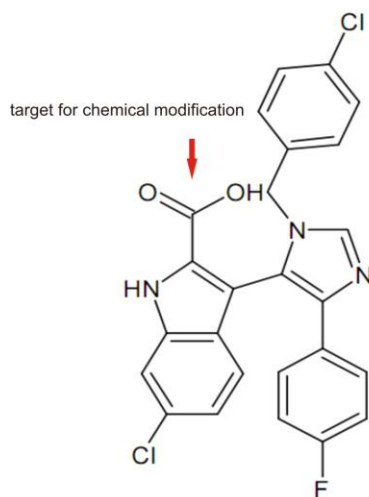
Figure 3.2.1 WK2-3A (MW 480.32 g/mol)

Estrification reaction gave 4 mg of compound 1 (20%) (Figure 3.2.2), which was analyzed by mass spectrometry (Figure 3.2.3).
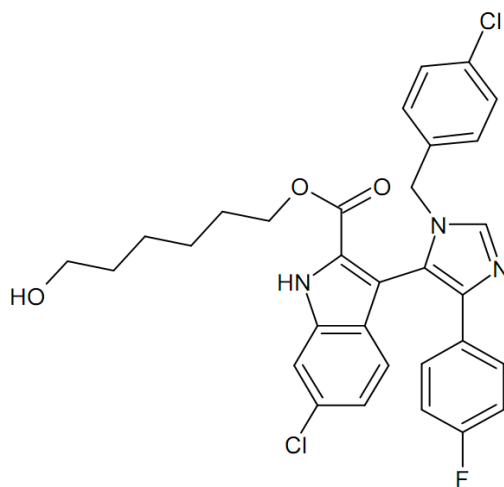

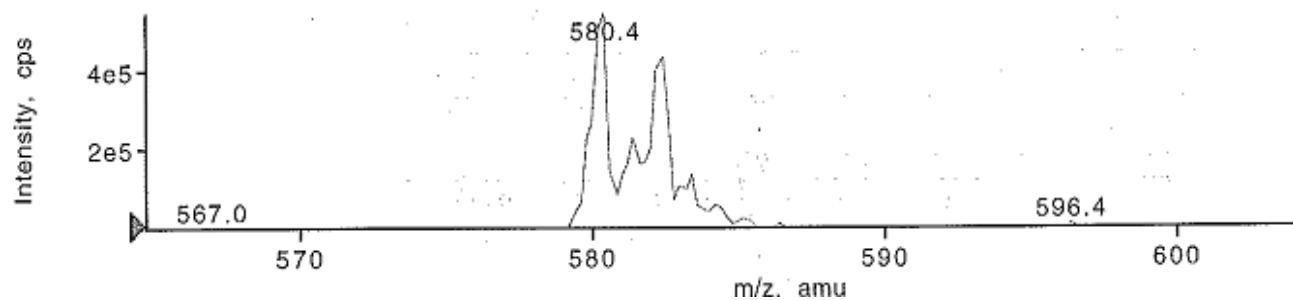Figure 3.2.2 Compound 1 (580.4 g/mol).

Figure 3.2.3 Mass spectrum of purified compound 1.

The second step was the synthesis of the N-hydroxylsuccinimide derivative of WK2-3A carried out according to the procedure described by Luo and Prestwich (1999) resulting in compound 2 (Figure 3.2.4 and Figure 3.2.5).
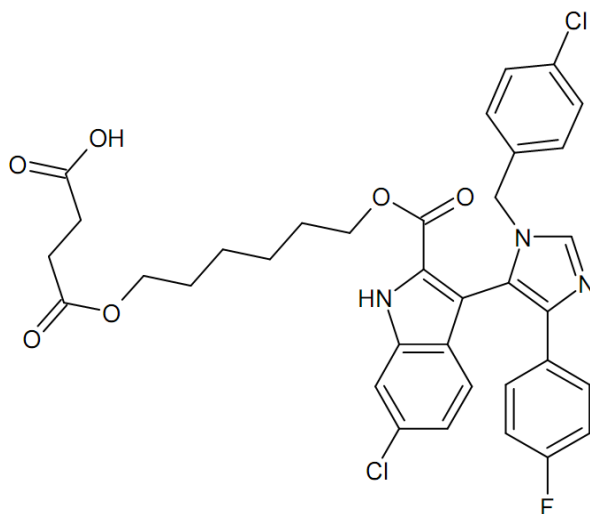

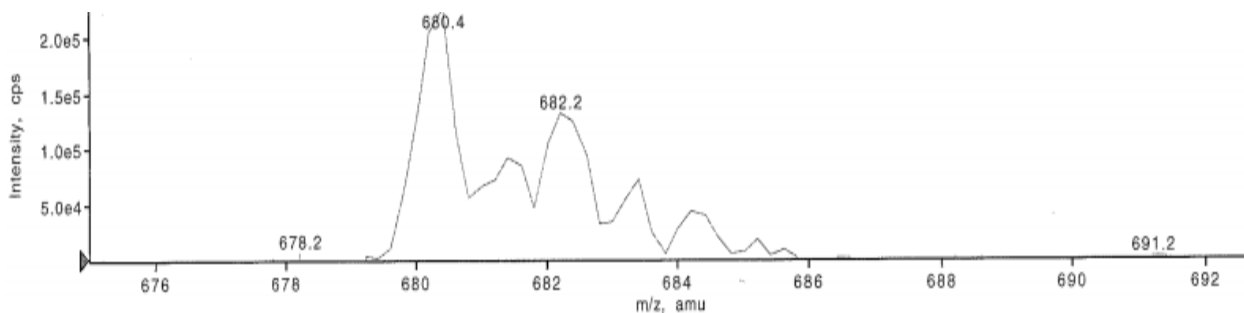Figure 3.2.4 Compound 2 (680.5 g/mol)

Figure 3.2.5 Mass spectrum of the purified compund 2.

The yield of the reaction was too low to continue the conjugation. The procedure was repeated several times but, unfortunately, we could not obtain enough of compound 2 (Fig. 3.2.4). In the future we are going to use other approaches to this synthesis. Firstly, the WK2-3A molecule can be activated with the thionyl chloride (Allen et al, 1963), subsequently connected to the linker such as buthyldiamine (Schotten, 1884; Baumann, 1886), and finally coupled to the hyaluronan derivative. Alternatively, we could try to connect directly succinic anhydride to the carboxyl group of WK2-3A in mild conditions to avoid destroying the molecule. If successful, the prepared molecule could be used for further modifications, and coupled with hyaluronan, and then tested against different cell lines.

3.2.2 Preparation of low molecular weight HA coupled with adipic dihydrazide

The use of a mild and versatile hydrazide method for preparation of the chemically modified HA derivatives allows attachment of reporter molecules, cross-linkers, drugs, and any combination of these moieties. Covalent attachment of adipic

89

dihydrazide (ADH) to the carboxylic acid groups of HA provides a controlled loading of the pendant hydrazide groups arrayed along the hyaluronate backbone, used for the conjugation of an antitumor agent.
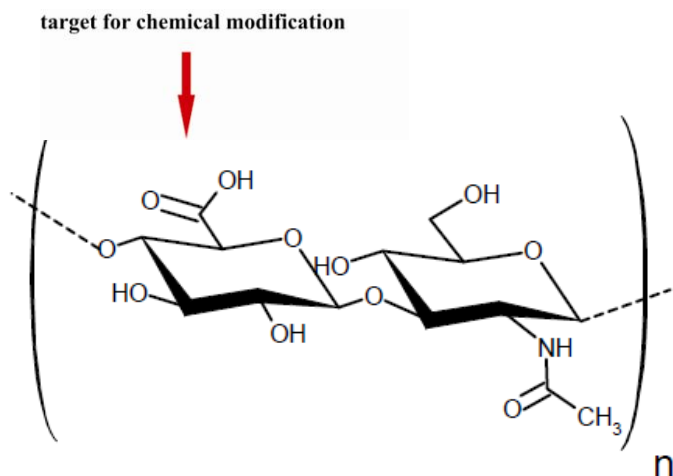

Figure 3.2.6 Disaccharide repeat unit of HA.

Use of LMW HA for coupling anit-tumor molecules has many advantages. First, it is possible to perform reducible chemical modifications and to monitor the extent of modifications by simple proton NMR methods. Second, LMW HA can be cleared from the body by renal ultrafiltration. Third, LMW bioconjugates should be injectable, nonviscous solutions and these molecules should suffer minimal further degradation in plasma and would be quickly taken up by cells. Finally, future efforts to crosslink the material into a hydrogel could also be easily controlled (Luo and Prestwich, 1999).
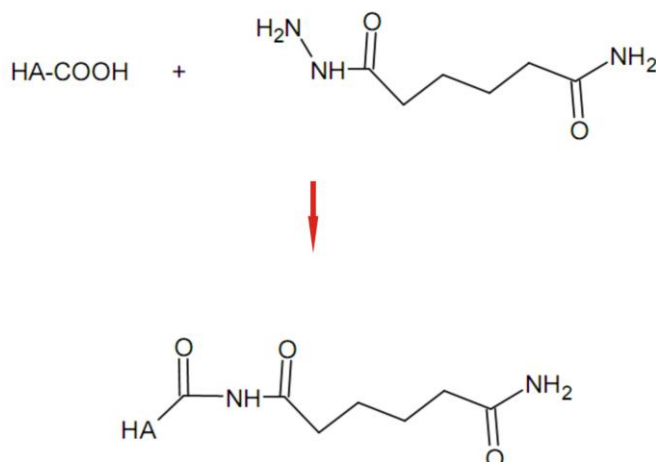
Figure 3.2.7 Synthesis of HA-ADH

The loading of ADH on the polymer backbone was determined by ^{1}H NMR spectroscopy studies with D_2O as a solvent. Analysis of the ^{1}H NMR spectra of HA-adipate suggested that adipic dihydrazide was covalently bound to the HA backbone without any appreciable crosslink formation. This conclusion was drawn by examining the changes in the NMR resonance of the methylene protons of the hydrazide linkers. The mixture of the low molecular weight HA with adipic dihydrazide in the absence of EDCI showed two multiplets in a ratio 1:1. The upfield multiplet comprised the internal methylene (H-2, H-3) protons of ADH, while the downfield multiplet contained the H-1 and H-4 methylene protons. Upon covalent binding to the HA backbone, downfield multiplet resolved into separate resonances. The H-1 methylene protons were in the ADH linkage become distinguishable from the H-4 terminal methylene protons (Figure 3.2.8). We did not observe the splitting of the upfiled multiplet (H-2, H-3), as previously reported (Pouyani and Prestwich,

1994). This was probably caused by not sufficiently optimized shimms of the magnet.
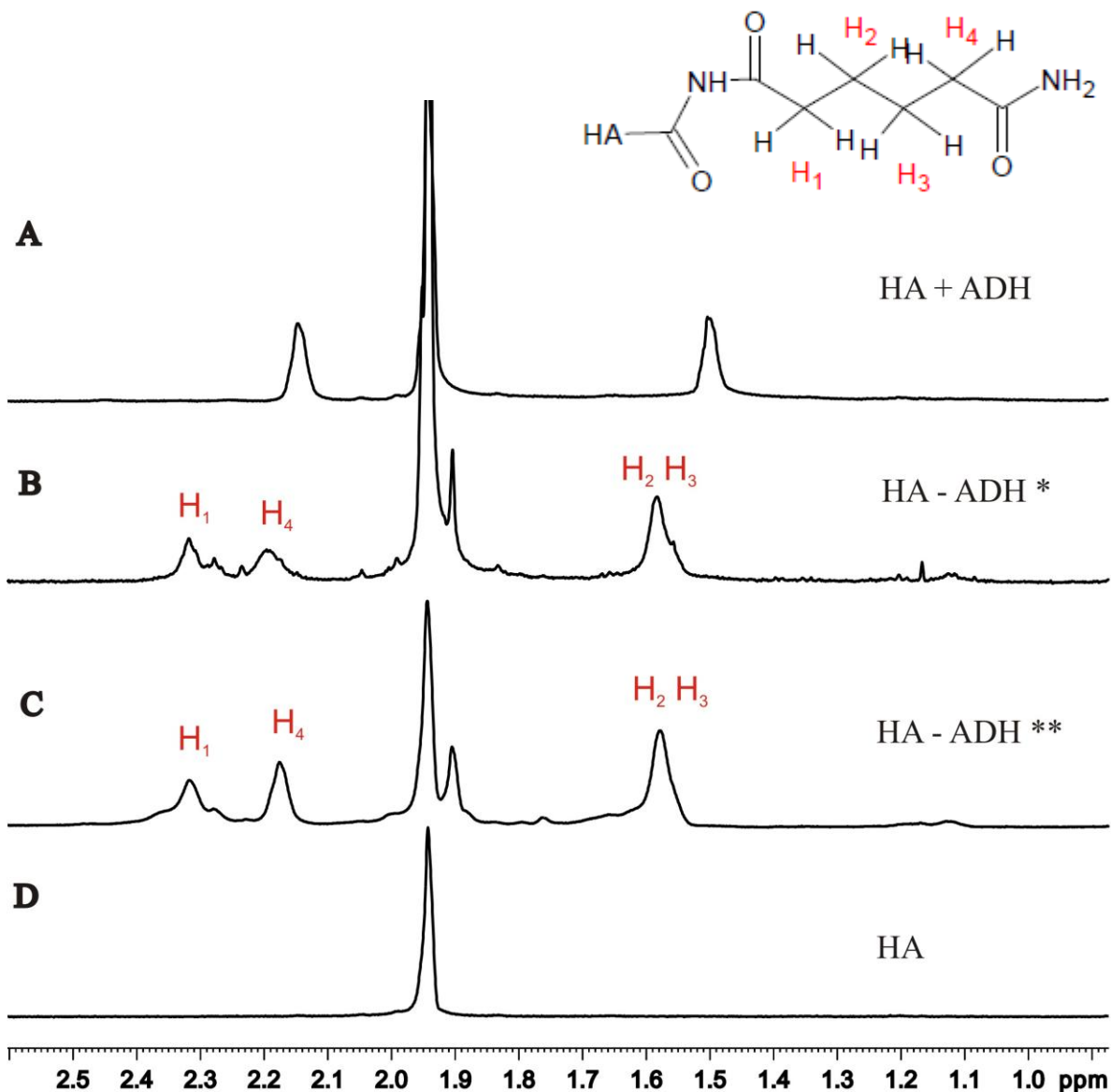

Figure 3.2.8 ^{1}H NMR spectra in D_2O of: (A) HA mixed with ADH. (B) HA covalently bound with ADH in the reaction carried out at pH 6.8 with the 30-fold excess of ADH. (C) HA covalently bound to ADH in the reaction at pH 4.75 with the 5-fold excess of ADH. (D) HA.

The extent of substitution of HA with ADH is important and should be estimated. It has been reported that the receptor-mediated cellular uptake of HA-conjugates depends on the degree of modification. HA modifications less than 25 mol% could be taken up by the cells with HA receptors without any effect on endocytosis, while highly modified conjugates, with the 68 mol% ADH content, were not internalized into the cells and evenly distributed in the body after injection (Oh et al., 2010). The degree of coupling was determined by integration of the linker methylene NMR signals (H-4) using as an internal standard the methyl resonance (δ = 1.88-1.98 ppm) of the acetamido moiety of the N-acetyl-D-glucosamine residue of a known fragment of HA (Kreil, 1995; Pouyani and Prestwich, 1994). NMR integration of spectrum B (Figure 3.2.8, reaction at pH 6.8, 30-fold excess of ADH) shows that 8% coupling of ADH to the low molecular HA occurred under these conditions. Reaction carried out at pH 4.75 with the 5-fold excess of ADH resulted in the 35% attachment of adipic dihydrazide to HA. ^{1}H NMR integration confirmed that different ADH loading occurred with different ratios of the reactants and different pHs of the reactions.

Both products could be used for further coupling with potential anti-cancer molecules and for testing the effect of degree of loading on the cellular uptake.

3.3 Ubiquitin like protein 5

3.3.1 Expression and purification of UBL5

The construct of UBL5 (comprising residues 1-73) in pET28a (gift from Tim Ammon, Department of Molecular Cell Biology, the MPIB) was expressed in high amount in *E. coli* strain BL21(DE3) in the LB medium. After induction at OD_{600} 0.7 the cells were kept at $25^{o}C$ overnight. Afterwards the cells were spinned down, resuspended in the lysis buffer and disrupted by sonication. The clarified solution was applied on the Ni-NTA resin, and after washing with the wash buffer, the protein was eluted with the buffer containing high concentration of imidazole. The His-tagged protein was subjected to the gel filtration in a suitable buffer for crystallization or NMR analysis, and eluted in form of a monomer (Figure 3.3.1 A). Purity of the protein was confirmed by SDS-PAGE electrophoresis (Figure 3.3.1 B) and mass spectroscopy. The folding of the recombinant protein was checked with NMR spectroscopy (Figure 3.3.2).
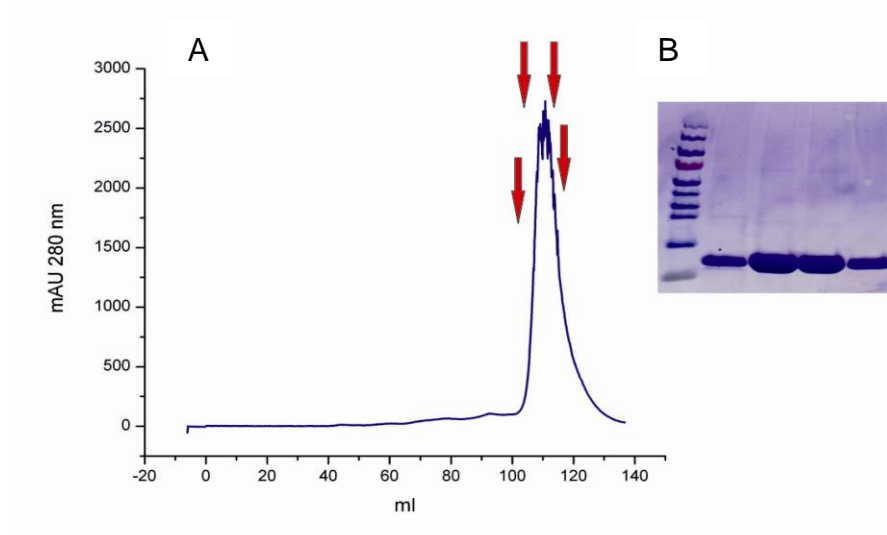
Figure 3.3.1 (A) Gel filtration of UBL5. Red arrows indicate fractions selected for electrophoresis. (B) The SDS-PAGE image of chosen fractions.
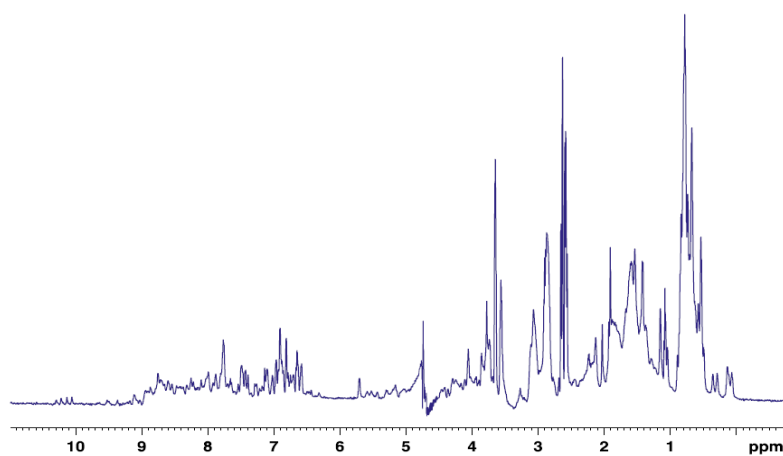

Figure 3.3.2 1D NMR proton spectra of the UBL5 protein.

The human 19 amino acids long HIND peptide was synthesized in the Microchemistry Core Facility of the Max Planck Institute for Biochemistry. The

secondary structure of the peptide was analyzed by circular dichroism spectroscopy and was shown to adopt a helical structure in solution (Figure 3.3.3).
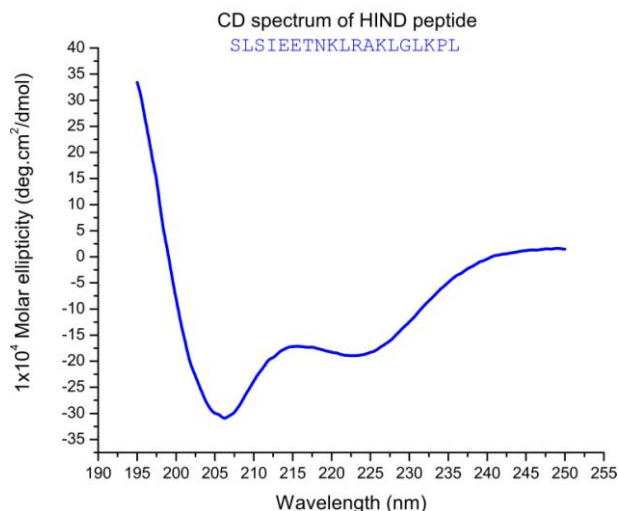
Figure 3.3.3 The CD spectrum of the free CD. Molar CD spectrum of the human 19 aa HIND shows a spectrum typical for an helical protein with two characteristic minima for α-helix at about 208 nm and 222 nm.

3.3.2 Interaction of UBL5 with the HIND peptide

In order to confirm interaction between the UBL5 protein and HIND peptide ^{1}H-^{15}N HSQC NMR experiments were carried out. The ^{15}N-uniformly labeled UBL5 was produced and mixed with the HIND peptide at a molar ratio 1:3 and the HSQC spectrum was recorded. After the addition of the HIND peptide, changes in NMR chemical shifts for several amino acids were observed (Figure 3.3.4).
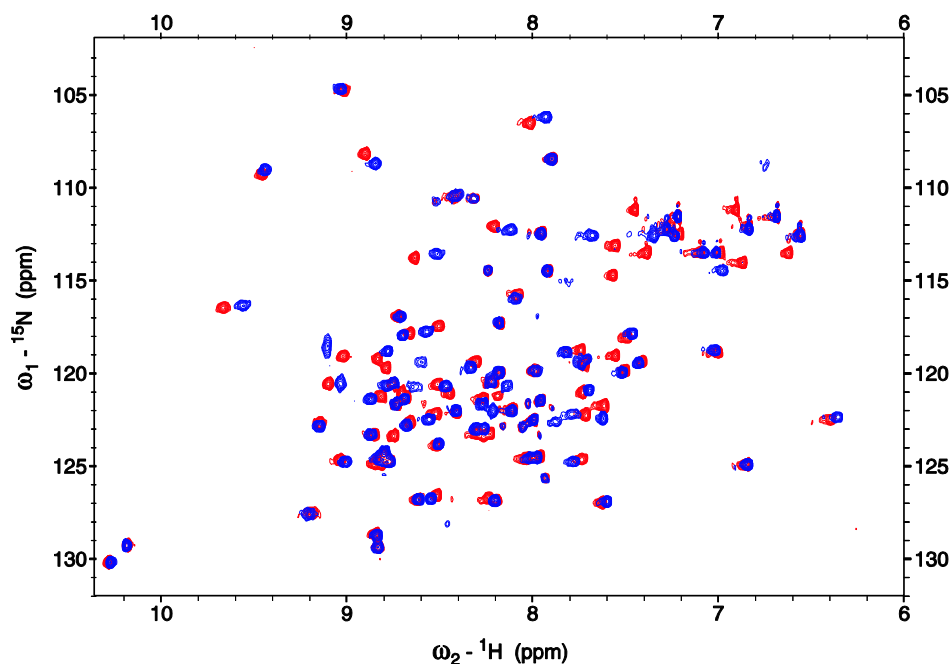
Figure 3.3.4 Superimposed ^{1}H-^{15}N HSQC spectra of UBL5 alone (red) and the mixture of UBL5 together with the HIND peptide at a molar ratio of 1:3 (blue).

3.3.3 Crystallization and structure determination

The purified UBL5 was mixed with the HIND peptide at a molar ratio 1:3 and separated by gel filtration in the crystallization buffer. The complex was concentrated to 10-13 mg/ml and crystallized at 20°C and 4°C. UBL5 was screened for crystallization using 96-well sitting drop plates with commercially available suites (Qiagene's Classics, PEGs, PACT, Cryos, and Hampton Research Index) as well as some in-house screens (crystal platform magic 1 and 2, complex screen 1 and 2), using a sitting vapor diffusion method. The screening gave rise to crystals in three different conditions (Figure 3.3.5, Table 3.3.1).

Table 3.3.1 Crystallization conditions.

Condition	Composition	Temp
Classics 44	0.1 M HEPES sodium salt pH 7.5, 1.4 M tri-sodium citrate	20°C
Classics 90	0.2 M sodium acetate, 0.1 M Tris-HCl pH 8.5, 30% (w/v) PEG 4000	20°C
Index 73	0.2 M sodium chloride, 0.1 M Tris-HCl pH 8.5, 25% (w/v) PEG 3350	20°C

In order to obtain crystals of better quality, the already established conditions were optimized. The best quality crystals for the X-ray data collection were picked up from the drop containing 0.1 M Tris-HCl, pH 9.0, 0.15 M sodium acetate, 20% (w/v) PEG 4000 (optimized Classics 90). Crystals were harvested, soaked in cryo-solution comprising of the mother liquor with 30% glycerol and were flash frozen in liquid nitrogen. The data set up to 2.2 Å was collected at the Swiss Light Source (Villigen, Switzerland) on the PXII beamline. The crystals belonged to the space group $P2_12_12$ and contained six complexes of the UBL5-HIND per an asymmetric cell. The collected data were integrated, and scaled by the XDS and XSCALE programs

(Kabsch, 1993). The structure was determined by molecular replacement using the Molrep program from the CCP4 suite (Collaborative Computational Project, Number 4, 1994). Due to expected structural similarity with the yeast homolog Hub1, the structure of the Hub1-HINDII complex (3PLV) was used as a probe for molecular replacement. Afterwards, the model was refined by Refmac5, improved and revised by MiFit. The Arp/wARP (Cohen et al., 2004) program was used to add solvent atoms. The resulting model was inspected and finished manually with the MiFit program. The final electron density map was of high quality; there were, however, no interpretable densities for couple of solvent exposed side chains, which were subsequently removed from the model. The final R-factor of the structure of the complex is 21.5% and R-free 26.6%. Data collection and refinement statistics are summarized in Table 3.3.2.
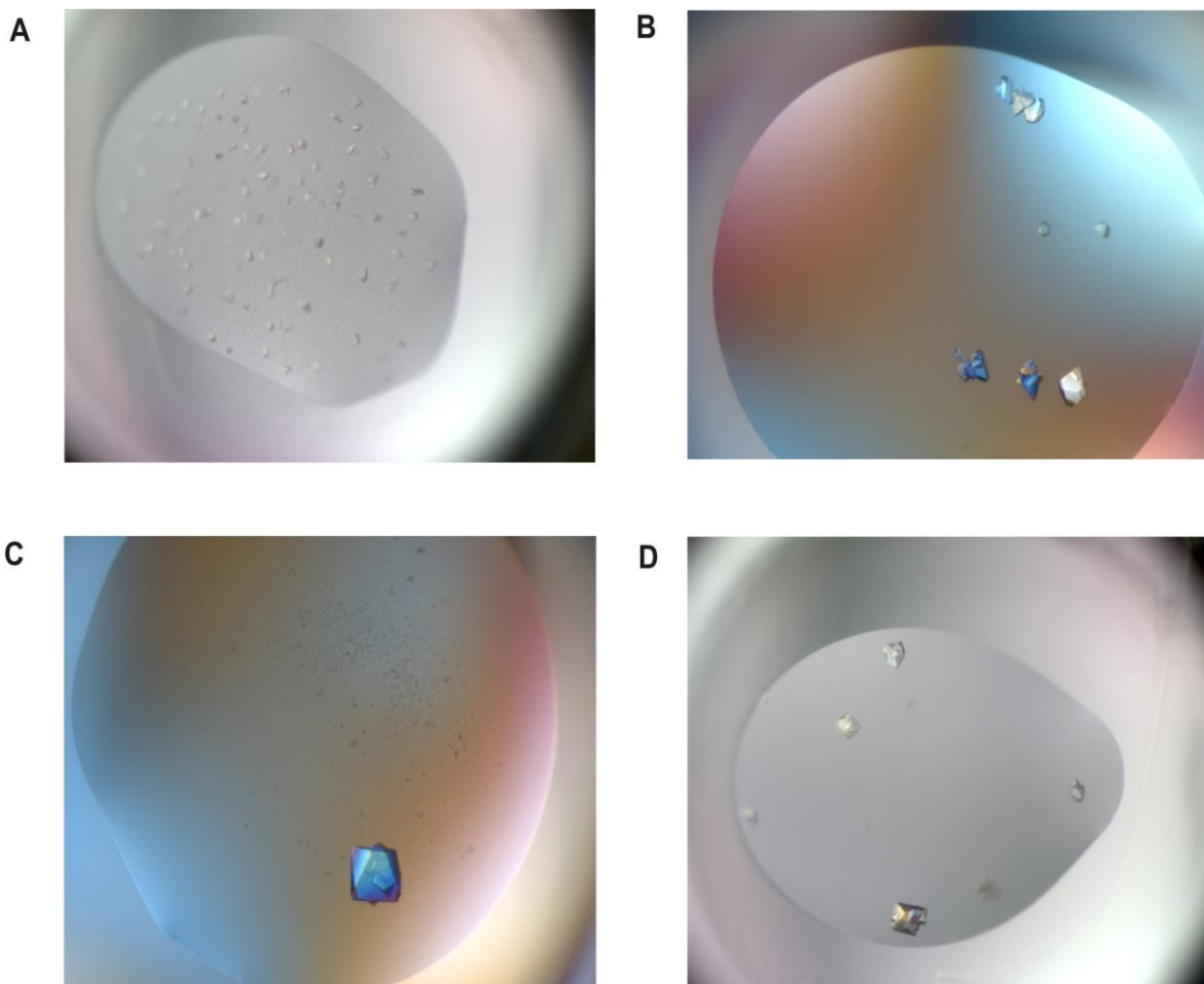
Figure 3.3.5 Crystals obtained in the commercially available factorials: (A) the UBL5-HIND complex in Classics 44, (B) and (C) the UBL5-HIND complex in Index 73, (D) the UBL5-HIND complex in Classics 90.

3.3.4 Structure of the UBL5-HIND complex

3.3.4.1 The overall structure

We have solved the structure of the HIND-bound complex of UBL5 (Figure 3.4.6). The overall fold of the human UBL5-HIND complex is essentially similar to the one described for the yeast Hub1-HINDI(II) complex (Mishra et al., 2011) and is highly similar to ubiquitin (Ramage et al., 1994) and SUMO (Bayer et al., 1998). The main chain fold of UBL5 does not change upon binding to the peptide (McNally et al., 2003). The structure of the UBL5-HIND complex comprises a half-open β-barrel completed with two flanking α-helices. The secondary structure elements have the ββαβαβ pattern. The N-terminal β-sheet of the protein covers residues Met1-Asp8 and is connected by a tight turn to another strand of the antiparallel β-strand comprising residues Gly11-Asn19. The first α-helix follows (Ile24-Thr35) and is connected to a third short sheet (Ile42-Lys45) through a small α-helical turn. The second α-helix covers residues Leu57-Tyr60 and is connected to the final β-sheet of the protein (Asn67-Tyr72). The β-strands are parallel in following pairs: β1 with β3 and β2 with β4 and they comprise a half-open barrel. The HIND peptide forms a 11-residue helix (Ile3-Leu13), which interacts with the β2 sheet and helix α1 of the protein. Interaction is mediated through a strong salt bridge formed by Arg11 of HIND and Asp22 of UBL5 accompanied by several hydrophobic contacts formed by aliphatic fragments following residues of HIND: Leu2, Ile4, Thr7, Leu10, Arg11 (Cβ and Cγ), Leu14, Leu16, and Leu19.

On UBL5 the interface is formed by: Met1, Val16, Lys17 (Cβ, Cγ, Cδ), Cys18, Asn19

(Cβ, Cγ), Lys29 (Cβ , Cγ, Cδ), Val30, and Leu33.

Table 3.3.1 Data collection and refinement statistics for the UBL5-HIND complex.

Space group	$P\ 2_12_12$
Cell constants (Å)	
a	87.51
b	103.63
c	67.00
Resolution range (Å)	50 - 2.0
Wavelength (Å)	1.0381
Observed reflections	306394
Unique reflections	41837
Whole resolution range:	
Completeness (%)	99.7
R_{merge}	6.4
$I/\sigma(I)$	21.9
Last resolution shell:	
Resolution range (Å)	2.1 - 2.0
Completeness (%)	99.8
R_{merge}	34.8
$I/\sigma(I)$	4.48
Refinement	
No. of reflections	35644
Resolution (Å)	20 - 2.0

R-factor (%)	21.5
R_{free} (%)	26.6
Average B (Å^2)	30.6
R.m.s bond length (Å)	0.01
R.m.s angles (^{o})	1.43
Content of asymmetric unit	
No. of complexes	6
No. of protein residues/atoms	570/5376
No. of solvent atoms	206
Ramachandran statistics	
Most favored regions (No./%)	474/95.4
Additionally allowed regions (No./%)	17/3.4
Generously allowed regions (No./%)	0/0
Disallowed regions (No./%)	6/1.2

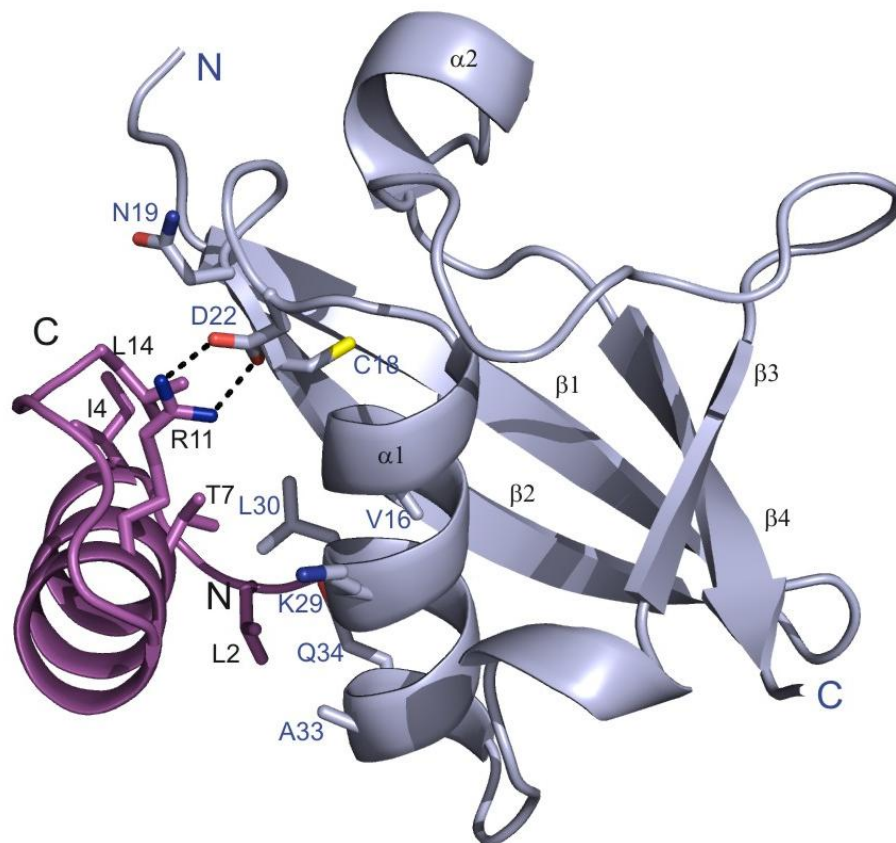Figure 3.4.6 Crystal structure of UBL5 in complex with the HIND peptide shown as a ribbon plot. The interaction interface is formed by a salt bridge between Arg11 and Asp22 of UBL5, accompanied by a patch of hydrophobic interactions.

3.3.4.2 Comparison with the Hub1-HINDI complex

The superposition of the UBL5-HIND and Hub1-HINDI(II) complexes shows them to be very similar (PDB code 3PLV; root mean squared deviation (r.m.s.d) of

0.716 Å for the main chain heavy atoms) (Figure 3.3.7).The largest difference is in a

α2-β3 loop containing Trp47 and Tyr48, which are present in human UBL5 but not in

the yeast homolog (Figure 3.3.8). The main chain of the loop is slightly shifted

outwards the protein core and side chains of these hydrophobic residues,

unexpectedly, are not buried within the protein but they point out to the solvent.

Analysis of the structure with the program PROCHECK shows that Trp47 and Tyr48

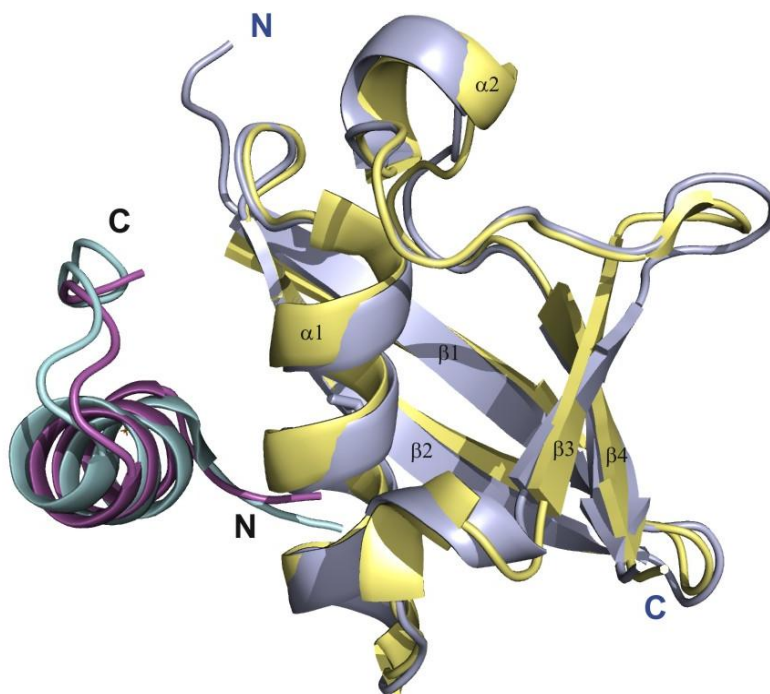lie in the disallowed regions of the Ramachandran plot.

Figure 3.3.7 Comparison of the UBL5-HIND complex with the Hub1-yHINDII complex. The UBL5-HIND interaction (blue, purple) superimposed with Hub1-yHINDII (yellow, cyan).

While comparing the main chain tracing, we can notice that strand β2 of UBL5 is

slightly shifted towards the α1 helix and its β3 sheet bends a little bit more to the

outside of the protein core. In addition, the side chain of His63 points out to different direction than in the yeasts' structure. Other minor differences include residues Lys11, Asp21 and Glu69 of UBL5, whose side chains are systematically shifted in comparison to the respective amino acids of the yeast Hub1.
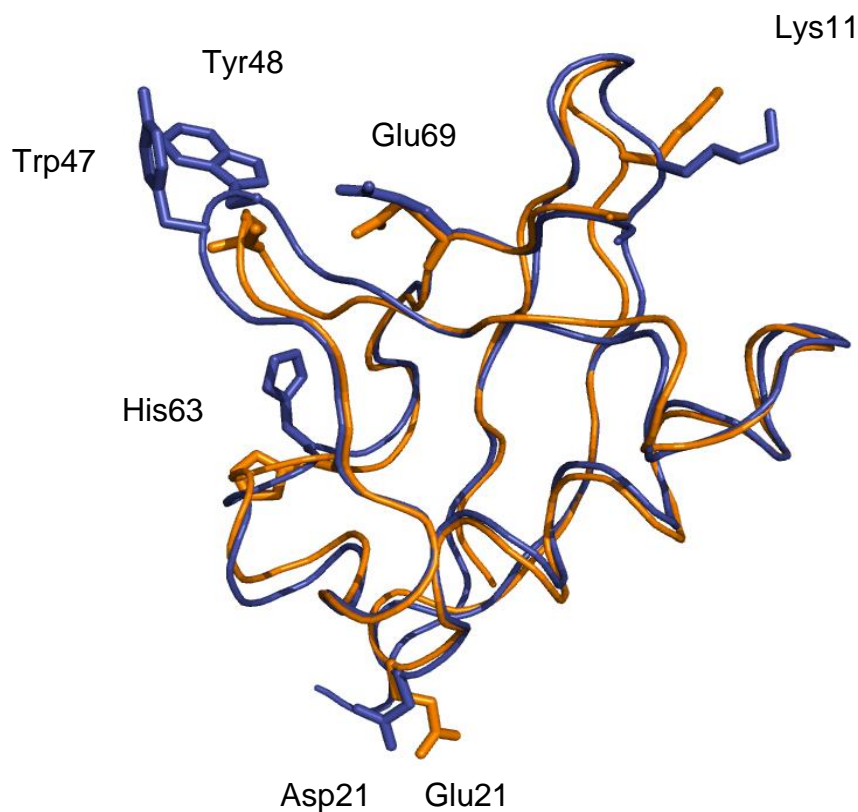
Figure 3.3.8 Alignment of the peptide-bound form of UBL5 (blue) and Hub1 (orange). The side chains of distinctive residues are depicted and labeled.

3.3.4.3 Protein-protein interactions

Analysis of orientation of the molecules in the asymmetric unit revealed astonishing observations which do not occur in the yeast Hub1. The UBL5-HIND

complexes are packed within the crystals as to form a filament that extends over the length of the crystal. The filament is arranged into a helix with six complex molecules in a turn and with the helical pitch of around 60 Å. The next complex starts the next turn and is positioned the same way as the first one (Figure 3.3.9).

A   B

Figure 3.2.9 Filament of the UBL5-HIND complex found in the crystals has a pitch of 60 Å. (A) The UBL5-HIND complex is colored differently within each crystallographic equivalent unit. (B) Axial projection of the filament.

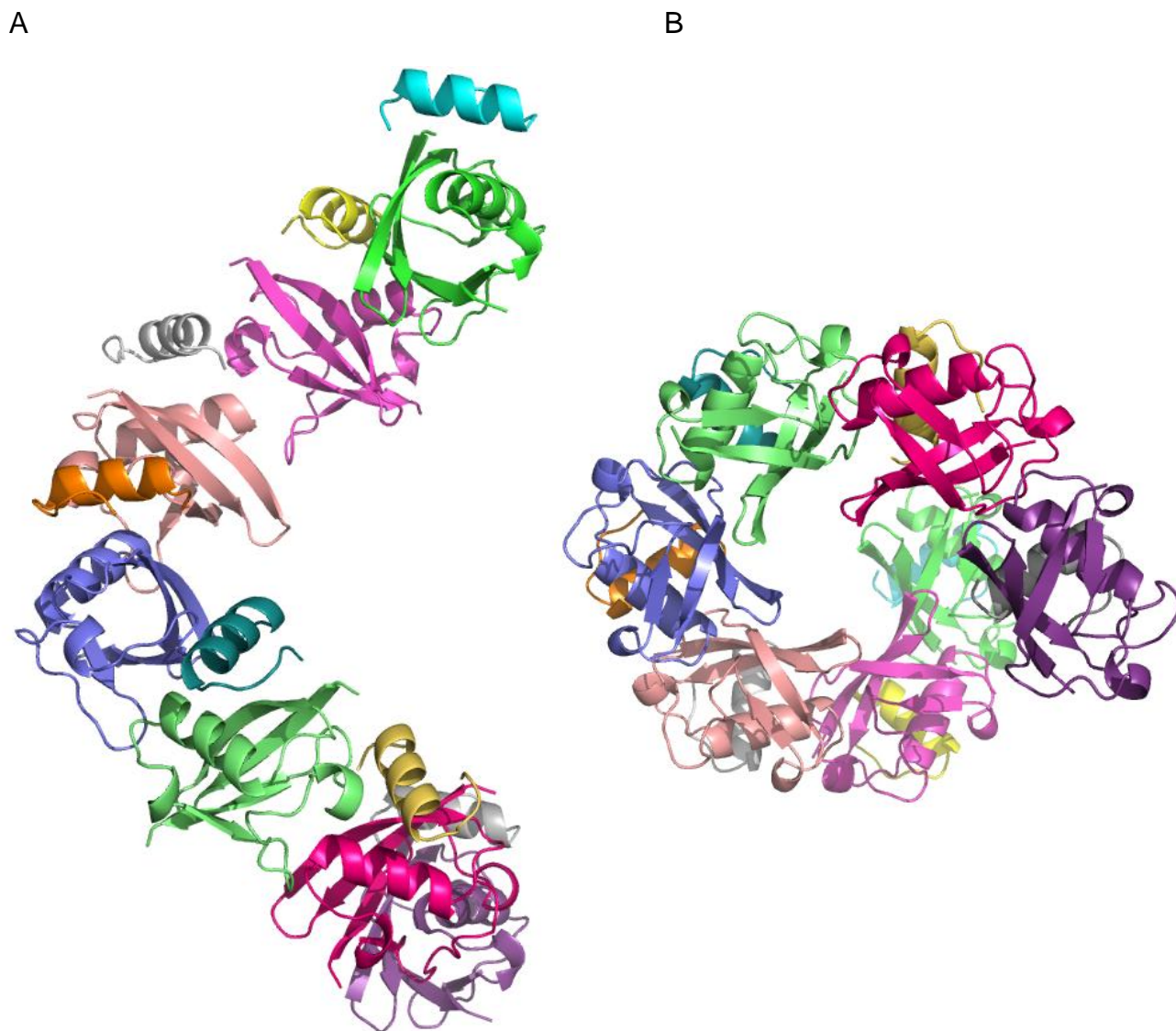To describe this interaction, we will refer to the molecules as "the lower" and "the upper" according to the presentation in Figure 3.3.10. The lower molecule interacts via the α1-β3 loop and the proximal part of β2. Major contribution to the interface (nearly half of the buried area) is provided by the peptide. The "upper" molecule interacts through helix α2 and its surrounding loops. The prominent role of hydrophobic residues from the human-specific α2-β4 loop is evident in mediating the interaction between two UBL5 molecules.

The deep binding pocket on the "lower" molecule recruits directly Trp47 and Tyr48 residues. Interactions between molecules comprise hydrophobic interactions between indole ring of Trp47 and aliphatic carbon Cβ of the sidechain of Lys11. The phenylalanine ring of Tyr48 is also found to be involved hydrophobic interactions with Lys11. Therefore, the side chain of Lys11 has a well-defined electron density and possesses the stable conformation, different from the yeast's Lys11. This triggers slight shift of the main chains of Lys11 and neighboring residues towards the core of the protein (Figure 3.3.8). At the same time the hydroxyl group of Tyr48 creates a hydrogen bond with the side-chain carboxyl of Asp8. Additionally, the loop conformation is stabilized by a hydrogen bond between Glu69 and the NH group of the Trp47 main chain. Furthermore the binding interface includes hydrophobic contacts overlapping on the HIND peptide surface.

A   B
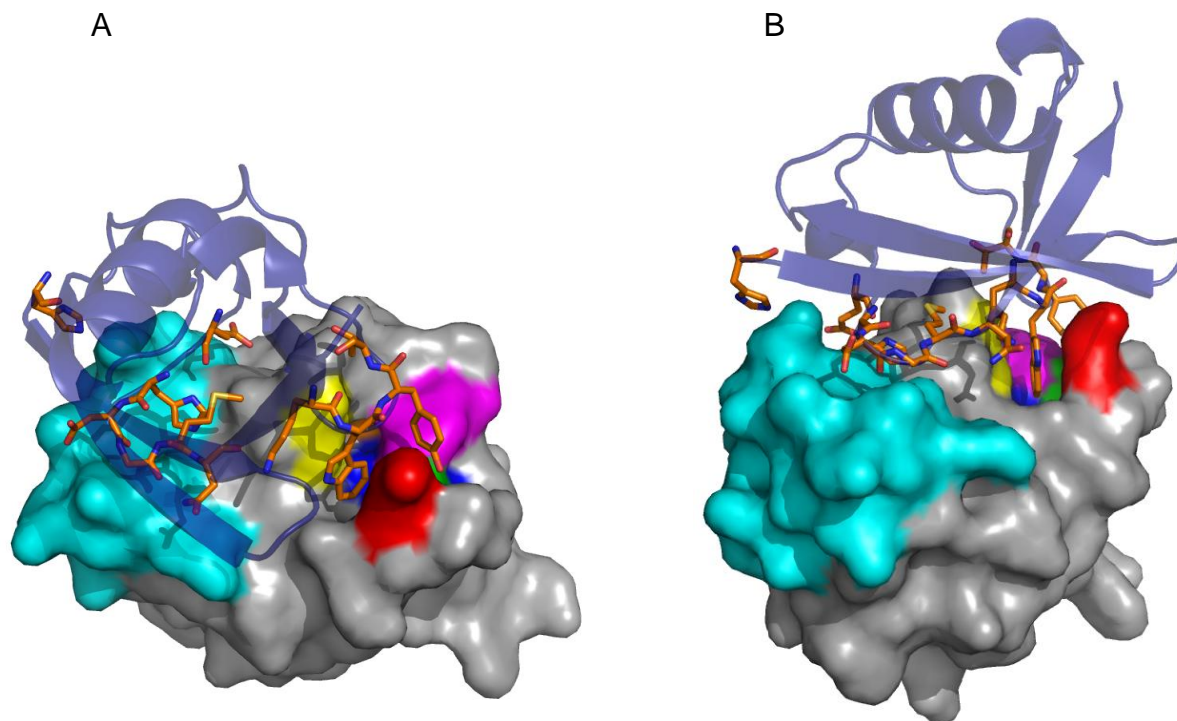
Figure 3.3.10 Surface representations of interaction between two molecules of the UBL5-HIND complex. Trp47 and Tyr48 of the "upper" molecule (dark blue) point out to the binding pocket (colored in red, pink, blue, and yellow) of the second UBL5 molecule (grey) and create hydrophobic interactions (A). The binding interface is accompanied by hydrogen bonds and hydrophobic interactions, which overlap on peptide (cyan) (B).

3.3.5 Affinity measurements

NMR spectroscopy was used to estimate the dissociation constants and compare them with the Hub1-yHINDs interactions. The yeast homolog binds yHIND peptides with affinity of ~ 1 µM (Mishra et al., 2011) and the interaction can be expected to undergo "slow" chemical exchange in NMR (Cavanagh et al., 2007).

NMR measurements consisted of monitoring changes in chemical shifts and line widths of the backbone amide resonances of the ^{15}N-enriched UBL5 samples upon addition of the unlabeled HIND peptides (Meyer and Peters, 2003; Pellecchia et al., 2002). We used the following peptides in this round of the study: the human HIND, yeast HINDI and yeast HINDII. As the assignment of the NMR spectra was not carried out up to now (and not intended), couple of the unassigned, most affected cross-peaks were chosen to track the binding. Examples of the perturbed resonances are labeled A, B and C, and described further in detail. Because of low solubility of the yeast HINDII, all titrations involving the peptide were performed in 10% DMSO.
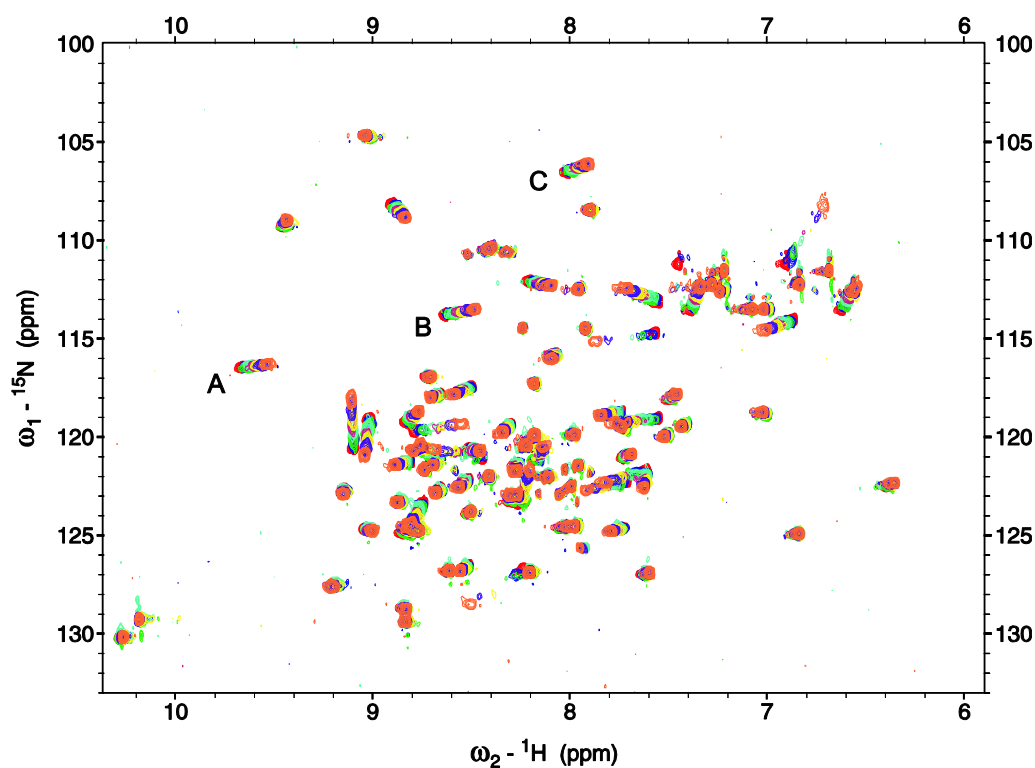
Figure 3.3.11 Titration of the ^{15}N-labeled UBL5 with the HIND peptide. The ^{15}N-labeled UBL5 was titrated with increasing concentrations of the human HIND peptide and ^{1}H-^{15}N HSQC spectra were recorded at each step of ligand addition. Examples of affected resonances are labeled A, B and C.

Figure 3.3.11 shows the titration of the ^{15}N-labeled UBL5 with the human HIND peptide. Many resonances are influenced and the changes were seen at the same UBL5 residues for titrations performed with the human HIND, yeast HINDI and yeast HINDII. The spectra showed a continuous movement of NMR peaks upon addition of increasing amounts of the peptide (Figure 3.3.12). This indicates that binding of all HIND peptides to UBL5 is in the "fast exchange", which points to the weak binding and enables determination of K_D values by NMR. In order to estimate dissociation constants of the interactions: UBL5-HIND, UBL5-yHINDI and UBL5-yHINDII, we applied the Pythagoras equation to the weighted chemical shifts as a function of ligand concentration (Stoll et al., 2001). Couple residues were chosen to calculate K_D values. Data were fit using a single binding site model.

1 A

2 A

1 B

2 B

1 C

2 C

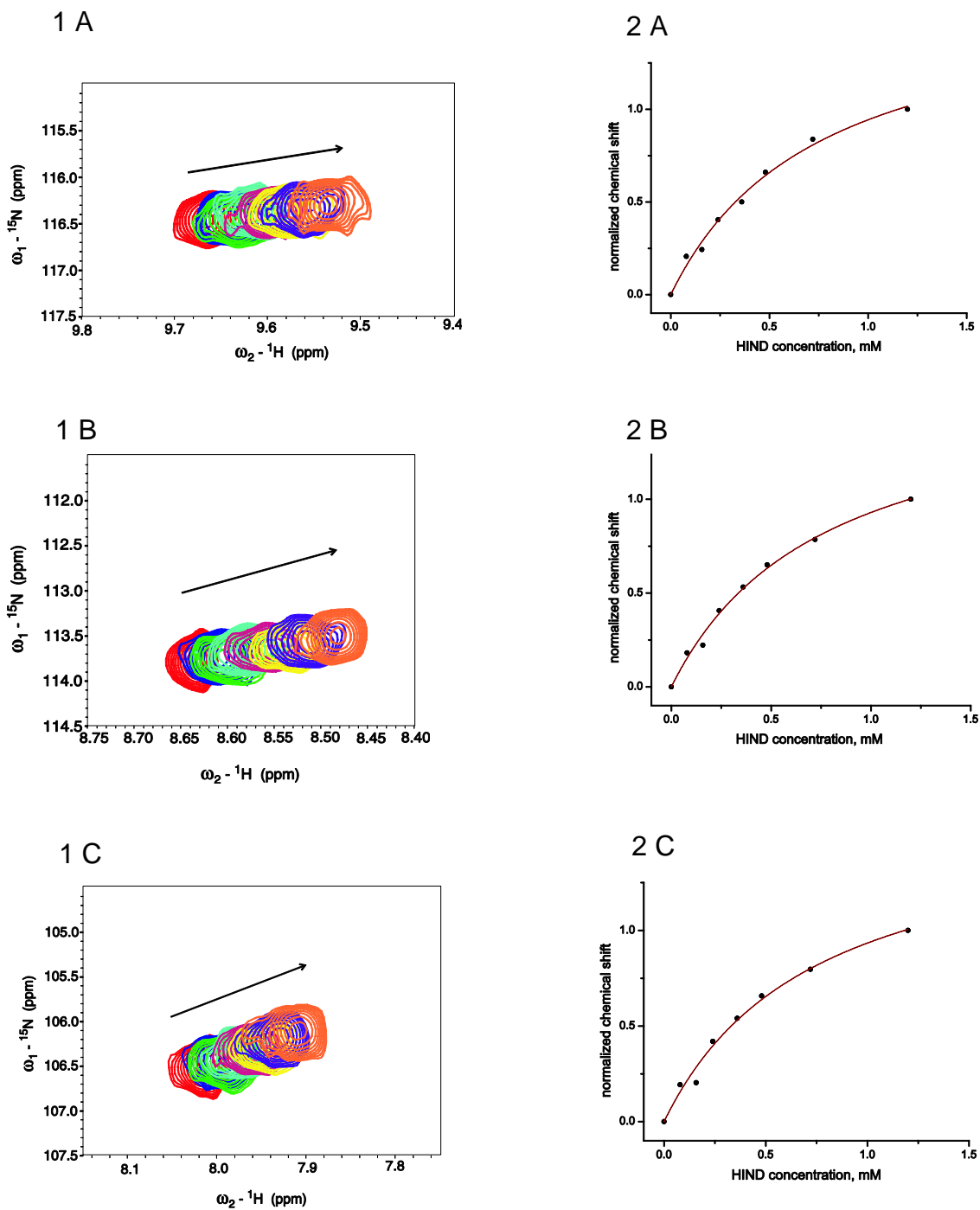Figure 3.3.12 Three examples of the "affected" crosspeaks of spectrum in Figure 3.3.11 (1A, 1B, 1C). All seven steps of titration are shown: the starting (red, untitrated), at a UBL5:HIND molar ratio of 0:0.33 (blue), 0:0.66 (green), 1:1 (light blue), 1:1.5 (purple), 1:2 (yellow), 1:3 (violet), and 1:5 (orange) (2A, 2B, 2C). Titrations with the HIND peptide are shown as chemical shift perturbations versus the ligand concetration.

Table 3.3.2 NMR titrations results for UBL5 and the HIND peptides. K_D values were determined by fitting single binding site models.

Interaction of UBL5 with:	Peptide sequence	NMR average K_D (mM)
human HIND	SLSIEETNKLRAKLGLKPL	0.33 ± 0.13
yeast HINDI	LSIEETNEIREKLGMKPI	0.39 ± 0.15
yeast HINDII	LSIEETNELRASLGLKLI	0.54 ± 0.20

Calculated K_Ds show weak binding between UBL5 and human HIND, and the cross-interaction between UBL5 and yeast HINDI and HINDII are of comparable affinities (Table 3.3.2). These values indicate 300 fold weaker binding than between the yeast homolog Hub1 and the yeast HINDI and HINDII, as reported by Mishra et al. (2011) in ITC experiments. ITC measurements of UBL5 with the HIND peptides were performed but unfortunately aggregation of the protein at higher concentrations made it impossible to get reliable results.

The ^{1}H-^{15}N HSQC titration of Hub1 with yeast HINDI and HINDII and human HIND were also performed to check binding affinities and confirm low micromolar dissociation constant obtained by ITC experiments (Mishra et al., 2011). As expected interactions: Hub1-yHINDI, Hub1-yHINDII but also Hub1-HIND undergo "slow" chemical exchange.
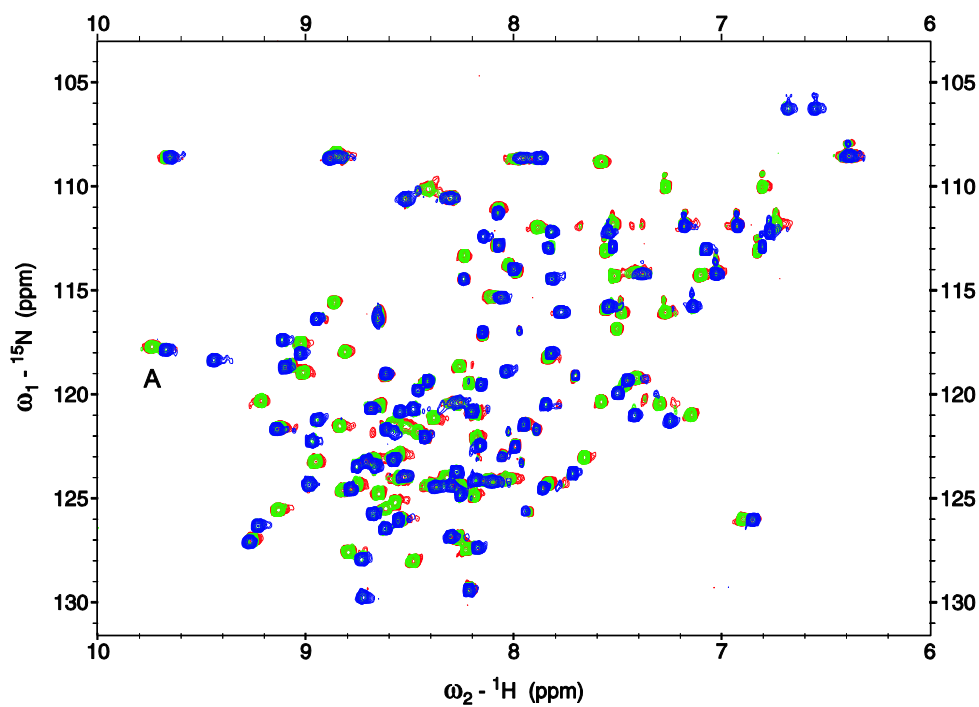
Figure 3.3.13 Overlay of ^{1}H-^{15}N HSQC spectra of Hub1 titrated with the human HIND. Three steps of the titration at specified molar ratio are shown: reference (red), at Hub1:HIND molar ratio of 1:0.33 (green) and 1:2 (blue).

The example of one of the experiments is shown in Figures 3.3.13 and 3.3.14. Peak A presented in the HSQC spectrum (Figure 3.3.13) perfectly illustrates the "slow" chemical exchange with two forms of peaks originating from the peptide "bound" and "free" forms of Hub1, where the intensity of the peak is proportional to the concentration of each of the forms.
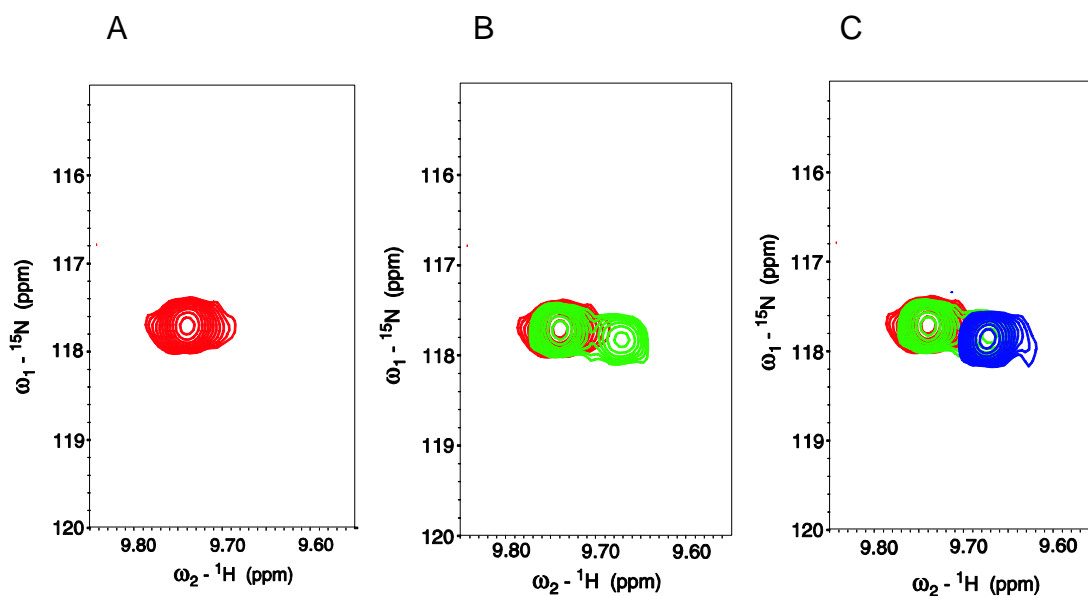
Figure 3.3.14 The cross peak (at 117.7 ppm ω_1-^{15}N; 9.74 ppm ω_2-^{1}H Figure 3.3.13) was chosen to demonstrate tight binding between Hub1 and the human HIND peptide in the "slow" chemical exchange NMR timescale. (A) First step of the titration (reference, red) shows the peak originating from the free form of an amino acid of Hub1. (B) Second step of titration at Hub1:HIND molar ratio of 1:0.33 (green) demonstrates two peaks, one corresponds to unbound form (overlay with reference in red) and smaller one, originates from the bound form. (C) The last step of experiment at Hub1:HIND molar ratio of 1:2 shows a single peak (blue) in complete saturation with the ligand.

When fast chemical exchange occurs, NMR titrations do not give precise affinity values and show only the upper limit of dissociation constants (Wüthrich, 1986). To determine accurate K_Ds, we used the newly published technology: the microscale thermophoresis (MST) (Wienken et al., 2010). The MST method enables quantification of the biomolecule interactions by the thermophoretic detection of even minute changes in conformation, charge and size of a molecule as they are induced by binding events and can be performed in every buffer of choice (Jerabek-Willemsen et al., 2011). No limitation on buffer allowed us to adjust the conditions

that prevented aggregation of UBL5 and gave reliable affinity values. We were successful in determining the dissociation constants of UBL5-HIND and Hub1-yHINDI interactions. The MST experiments consist of direct measurements of the protein binding to the fluorescently labeled peptide. Concentration of the peptide is kept constant, whereas the protein is titrated within the micromolar to nanomolar range. The binding of the molecules is observed as a change in the thermophoretic property of the fluorescently labeled peptide upon complex formation. The complexes show a stronger increase of the normalized fluorescence than the unbound peptide (Figure 3.3.15). The dissociation constants are determined to be around 11 µM for the UBL5-HIND complex and around 1.3 µM for the Hub1-yHINDI complex. The value obtained for the yeast molecules is in good agreement with the literature data (Mishra et al., 2011) and confirms the reliability of the method.
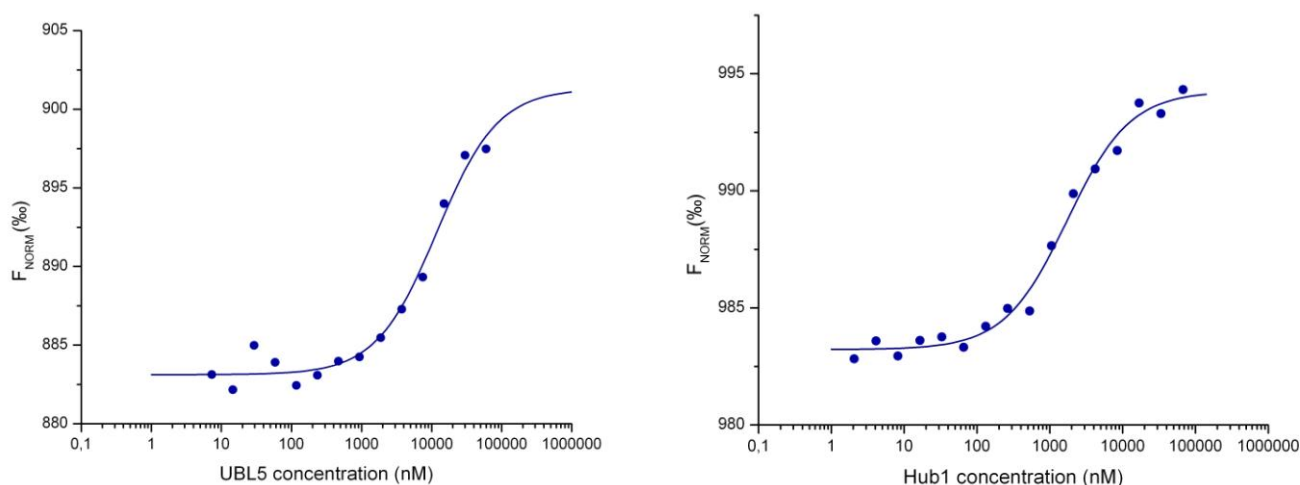
Figure 3.3.15 Protein-peptide interactions. Direct interaction study between UBL5 and the fluorescently labeled HIND peptide (A) and between Hub1 and the labeled yHINDI (B). The concentration of both peptides was kept constant at 50 nM. The difference in the normalized fluorescence (‰) is plotted for analysis of thermophoresis. The K_D is fitted to 11 µM (A) and 1.3 µM (B).

3.3.6 Discussion

We present the structure of the UBL5-HIND complex. The binding mode of UBL5-HIND as well as the Hub1-HIND complex from *S. cerevisiae* described by Mishra et al. (2011) reveals a very unusual binding paradigm, unseen in previously described interactions of ubiquitin and UBLs with their binding partners. Protein binds the HIND peptide on the site almost exactly opposite to the usual ligand-binding interface of the ubiquitin fold (Figure 3.3.16).
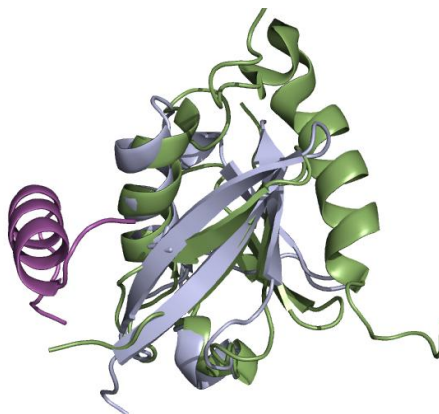
Figure 3.3.16 Comparison of the structure of the UBL5-HIND complex with that of the ubiquitin-UIM (PDB: 1P9D) shows an unusual binding paradigm.

Furthermore, the binding affinities of ubiquitin and the ubiquitin receptors are usually weak ($K_D \geq 100$ µM) (Song et al., 2005), as opposed to the strong binding of the yeast HINDI by Hub1 (~ 1.5 µM K_D), and slightly weaker binding of human HIND by UBL5 (~ 11 µM K_D).

The sequence and structural similarities between Hub1 and UBL5 are not reflected in their similar binding affinities, as assayed by NMR and microscale thermophoresis

117

(MST). NMR demonstrated that UBL5 interacts with human HIND and yeast HINDI and HINDII in the "fast exchange" ($K_D \geq 1$) in contrast to Hub1, which binds all peptides in the "slow exchange" ($K_D \leq 1$). This indicates that differences in binding affinities depend on the nature of the protein, not peptides, since cross-interactions are of similar affinities as interactions among the species. The weaker binding of HINDs by UBL5 might be caused by aggregation of the protein observed during purification procedures. However, we can also speculate that during spliceosome modification UBL5 can be bound by some additional splicing factors, whereas binding of Snu66 in *S. cerevisiae* might be less complex and more exclusive. The most significant difference between both structures refers to the loop α2-β3 containing an aromatic patch. These solvent exposed hydrophobic residues are highly unusual and, being thermodynamically unfavoured, are likely to be involved in formation of an interface, suggesting therefore existence of the binding partner.

UBL5   MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDYEIHDGMNLELYYQ 73

Hub1   MIEVVVNDRLGKKVRVKCLAEDSVGDFKKVLSLQIGTQPNKIVLQKGGSVLKDHISLEDYEVHDQTNLELYYL 73

Figure 3.3.17 Alignment of UBL5 and Hub1 shows high sequence similarity between proteins. Identical residues are colored blue; similar, grey. Residues belonging to the characteristic loop are labeled in red.

Despite the fact that the structures of *H. sapiens* and *S. cerevisiae* proteins are highly similar, the arrangement of molecules in the crystal unit turned out to be completely different. The Hub1-HINDI(II) complexes are found to be monomeric, whereas the human equivalent appeared to form a filament. The main residues

contributing in the supra-helix formation are Trp47 and Tyr48, which were previously

mentioned to probably being involved in some unrevealed interactions. Additionally,

the new binding interface is of a significantly larger surface than the binding interface

between UBL5 and the HIND peptide and seems to be necessary for the helix

formation. In order to examine the helix formation in solution and check possible

influence of the crystallization buffer on polymerization, we performed ^{1}H-^{15}N HSQC

experiment in the crystallization buffer (0.1 M Tris pH 9.0, 0.15 M sodium acetate,
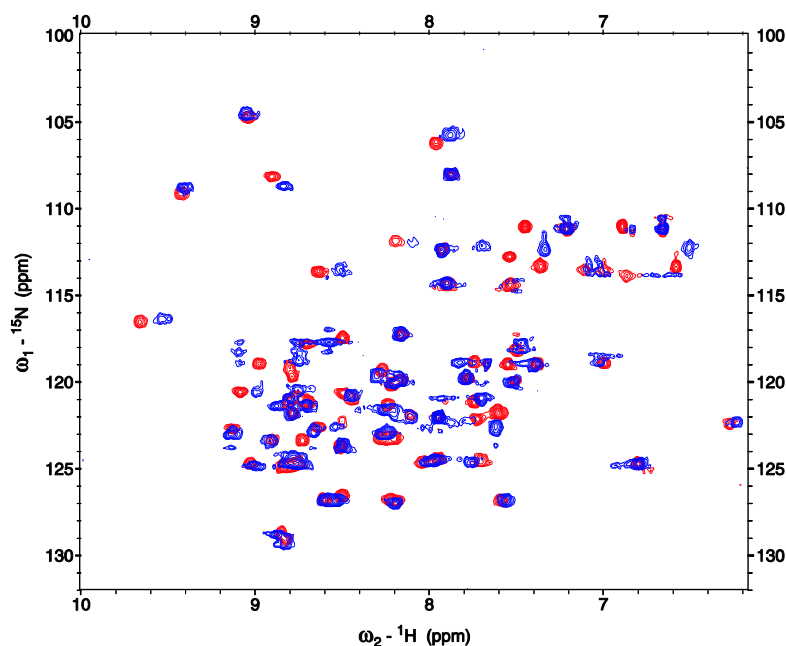
20 % PEG 4000).

Figure 3.3.18 ^{1}H-^{15}N HSQC spectra of UBL5 titrated with the human HIND under crystallization conditions (0.1 M Tris pH 9.0, 0.15 M sodium acetate, 20 % PEG 4000).

Upon the peptide binding we observed chemical shift perturbations which

represented the same pattern as UBL5 with HIND at physiological conditions (Figure

3.3.18). Since possible polymerization would cause peaks disappearance, this

indicates that the helical assembly does not occur in solution of the crystallization

buffer. Furthermore, gel filtration profile of the complex corresponds to a monomer form of the protein and no aggregates or oligomers were observed. However, increased complex concentration induces oligomerization in the experiments in vitro and we can speculate that the filament formation might be concentration dependent. Nevertheless, the helix existence should be investigated in experiments *in vivo* to exclude any bias.

4 Summary

Two projects were undertaken in this thesis. The first project involved functional investigation of CD44, development of small-molecule inhibitors of CD44, and coupling the hyaluronic acid, the CD44 ligand, to potential anti-cancer drugs. The focus of the second project is the structural characterization of the ubiquitin-like protein 5.

The first part of the thesis focuses on the CD44 glycoprotein, the receptor for components of the extracellular matrix such as hyaluronic acid, collagen, and fibronectin. The protein is involved in numerous processes like, for example, tumor cell proliferation, adhesion and invasion and the expression of wt-CD44 is frequently elevated in epithelial and some hematopoietic tumours. Both CD44 on cancer cells and hyaluronan (HA) have been important targets of anticancer therapy. HA was widely used in the arthritis treatment, ophthalmic surgery, drug delivery and tissue engineering. Especially for drug delivery applications HA exhibits a number of properties which make it a successful drug carrier. HA is water soluble, non-toxic, non-immunogenic, non-inflammatory, and a completely biodegradable polysaccharide. The aim of the thesis was to determine structural attributes and to develop small-molecule compounds for CD44. CD44 has recently been associated with the regulation of cancer-initiating cells (CICs). These cells are, in most cases, postulated to have properties of stem cells (cancer stem cells, CSCs). We used NMR spectroscopy to structurally characterize CD44 and its ligand-protein interactions and to develop small-molecule inhibitors of CD44 with the aim to kill

cancer stem cells. Additionally, as hyaluronic acid was reported to be used for conjugation with various chemical drugs, we prepared the low molecular weight HA and tried to couple it with potential anti-cancer substances.

The ubiquitin-like protein 5 (UBL5), known as Hub1 in *S. cerevisiae*, is structurally similar to ubiquitin and equally highly conserved and ancient as ubiquitin, yet the two proteins function in completely different ways. Hub1 binds the spliceosomal protein Snu66, and was reported to be required for the usage of certain non-canonical 5' splice sites. The protein binds non-covalently to a conserved element HIND, revealing a new binding paradigm unseen in interactions of ubiquitin and ubiquitin-like proteins with their binding partners. Hub1 binds the HIND peptide on the site almost exactly opposite to the usual ligand-binding interface of the ubiquitin fold. For this study, the crystal structure of the complex between UBL5 and the human HIND peptide was determined. The structure gives insight into the HIND binding mode and shows differences between the human and *S. cerevisiae* complexes. The spatial arrangement of the UBL5-HIND complexes in the crystal unit indicates that the molecules can evolve into a filament. We investigated this interaction using purified proteins and biophysical methods, like nuclear magnetic resonance (NMR) and microscale thermophoresis (MST).

5 Zusammenfassung

Zwei Projekte wurden im Zuge dieser Dissertation durchgeführt. Das erste Projekt bestand in einer funktionalen Untersuchung von CD44, der Entwicklung von niedermolekularen Inhibitoren des CD44-Proteins und der Verbindung von Hyaluronsäure, dem CD44-Liganden, an potentielle Anti-Krebs-Medikamente. Das Fokus des zweiten Projekts bezieht sich auf Strukturanalyse des ubiquitinähnlichen Proteins 5.

Der erste Teil der Dissertation befasste sich mit dem CD44-Glykoprotein, dem Rezeptor für die Komponenten einer extrazellulären Matrix wie beispielsweise Hyaluronsäure, Kollagen und Fibronektin. Das Protein ist an zahlreichen Prozessen beteiligt wie zum Beispiel dem Wachstum, der Adhäsion und der Invasion von Tumorzellen. Außerdem ist die Expression von wt-CD44 häufig in dem Gewebe von epthelialen und einigen hematopoetischen Tumoren erhöht. CD44 und Hyaluronsäure (HA) sind beides wichtige Ziele der Krebstherapie. HA war weitverbreitet bei der Arthritistherapie, der Augenchirurgie, der Medikamentenverabreichung und der Entwicklung von Gewebe. Insbesondere die Anwendung von HA bei der Medikamentenverabreichung eröffnet etliche Möglichkeiten, wodurch sie zu einer erfolgreichen Trägersubstanz für Medikamente wurde. HA ist ein wasserlösliches, ungiftiges, nichtimmunogenes, nichtinflammatorisches und vollständig biologisch abbaubares Polysaccharid. Das Ziel dieser Dissertation bestand darin, Strukturmerkmale aufzuweisen und niedermolekulare Verbindungen gegen CD44 zu entwickeln, welches kürzlich mit der Funktionsweise krebsauslösender Zellen (CICs) in Verbindung gebracht wurde.

Von diesen Zellen wird angenommen, dass sie in den meisten Fällen Eigenschaften von Stammzellen aufweisen. Wir benutzten NMR-Spektroskopie, um CD44 und seine Ligand-Protein-Interaktionen strukturell zu charakterisieren und um niedermolekulare Inhibitoren gegen CD44 zu entwickeln mit dem Ziel, Krebsstammzellen zu töten. Da berichtet wurde, dass Hyaluronsäure für Verbindungen mit verschiedenen chemischen Präparaten genutzt wurde, stellten wir außerdem HA mit niedrigem Molekulargewicht her und versuchten es mit potentiellen Anti-Krebs-Substanzen zu kombinieren.

Das ubiqitinähnliche Protein 5, bekannt als Hub1 in *S. cerevisiae*, ist dem Ubiquitin strukturell sehr ähnlich und gleichermaßen hochkonserviert und alt, obgleich beide Proteine vollkommen unterschiedliche Funktionen haben. Hub1 bindet das spleißosomale Protein Snu66 und ist für den Gebrauch von bestimmten nicht-kanonischen 5er-Spleißstellen erforderlich. Das Protein bindet nichtkovalent an ein konserviertes Element HIND und legt dabei ein neuartiges Bindungssystem zutage, welches bei der Interaktion von Ubiquitin und ubiquitinähnlichen Proteinen mit ihren Bindungspartnern bislang unbekannt war. Hub1 bindet das HIND-Peptid an einer Stelle, welche der gewöhnlichen Verbindungsstelle der Ubiquitinstruktur für Ligand-Bindungen fast genau gegenüberliegt. Für diese Studie wurde die Kristallstruktur des Komplexes von UBL5 und menschlichem HIND-Peptid bestimmt. Die Struktur gewährt Einblicke in die Bindungsweise von HIND und zeigt Unterschiede zwischen den menschlichen Komplexen und denen bei *S. cerevisiae* auf. Die mehrdimensionale Formation der UBL5-HIND-Komplexe in den Kristalleinheiten weist darauf hin, dass die Moleküle sich zu Filamenten formieren

können. Wir erforschten diese Interaktion, indem wir gereinigte Proteine und biophysikalische Methoden anwandten wie Kernspinresonanz (NMR) und Microscale Thermophoresis Technology (MST).

6 Appendix

6.1 Abbreviations and symbols

1D	one-dimensional
2D	two-dimensional
Å	Ångstrøm (10^{-10} m)
aa	amino acid
ADH	adipic acid hydrazide
AEMA	aminoethyl methacrylate
AP	alkaline phosphatase
APMAm	aminopropyl methacrylamide
APS	ammonium peroxodisulfate
bp	base pair
BSA	bovine serum albumin
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
ChS	chondroitin sulfate
CIC	cancer initiating cells
CIP	calf intestinal alkaline phosphatase
CLK	cyclin-like kinase
Da	Dalton (g mol^{-1})
DMSO	dimethylsulfoxide

DNA	deoxyribonucleic acid
DTT	dithiothreitol
DUB	deubiquitilating enzyme
ECM	extracellular matrix
EDTA	ethylenediamine tetraacetic amid
g	gravity (9.81 m s^{-2})
GAG	glycosaminoglycan
GSH	reduced glutathione
GSSG	oxidized glutathione
HA	hyaluronic acid
HABD	hyaluronan binding domain
HIND	Hub1-interaction domain
HMDA	hexamethylenediamine
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single quantum coherence
Hub1	homologous to ubiquitin
Hz	Hertz
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-thiogalactopyranoside
LB	Luria-Broth medium
LMW	low molecular weight

MM	minimal medium
NiNTA	nickel-nitrilotriacetic acid
NHS	N-hydroxylsuccinimide
NMR	nuclear magnetic resonance
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PDPH	propionyl hydrazide
PEG	polyethylene glycol
ppm	parts per million
SDS	sodium dodecyl sulphate
SUMO	small ubiquitin-like modifier
TAE	Tris-acetate-EDTA buffer
TBA	tetrabytylammonium salt
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylendiamine
Tris	tris(hydrixylmethyl)aminomethane
UBD	ubiquitin-binding domain
UBL	ubiquitin-like protein
UPS	ubiquitin-proteasome system

6.2 Protein sequences

6.2.1 CD44

```
        10         20         30         40         50         60
MDKFWWHAAW GLCLVPLSLA QIDLNITCRF AGVFHVEKNG RYSISRTEAA DLCKAFNSTL
        70         80         90        100        110        120
PTMAQMEKAL SIGFETCRYG FIEGHVVIPR IHPNSICAAN NTGVYILTSN TSQYDTYCFN
       130        140        150        160        170        180
ASAPPEEDCT SVTDLPNAFD GPITITIVNR DGTRYVQKGE YRTNPEDIYP SNPTDDDVSS
       190        200        210        220        230        240
GSSSERSSTS GGYIFYTFST VHPIPDEDSP WITDSTDRIP ATTLMSTSAT ATETATKRQE
       250        260        270        280        290        300
TWDWFSWLFL PSESKNHLHT TTQMAGTSSN TISAGWEPNE ENEDERDRHL SFSGSGIDDD
       310        320        330        340        350        360
EDFISSTIST TPRAFDHTKQ NQDWTQWNPS HSNPEVLLQT TTRMTDVDRN GTTAYEGNWN
       370        380        390        400        410        420
PEAHPPLIHH EHHEEEETPH STSTIQATPS STTEETATQK EQWFGNRWHE GYRQTPKEDS
       430        440        450        460        470        480
HSTTGTAAAS AHTSHPMQGR TTPSPEDSSW TDFFNPISHP MGRGHQAGRR MDMDSSHSIT
       490        500        510        520        530        540
LQPTANPNTG LVEDLDRTGP LSMTTQQSNS QSFSTSHEGL EEDKDHPTTS TLTSSNRNDV
       550        560        570        580        590        600
TGGRRDPNHS EGSTTLLEGY TSHYPHTKES RTFIPVTSAK TGSFGVTAVT VGDSNSNVNR
       610        620        630        640        650        660
SLSGDQDTFH PSGGSHTTHG SESDGHSHGS QEGGANTTSG PIRTPQIPEW LIILASLLAL
       670        680        690        700        710        720
ALILAVCIAV NSRRRCGQKK KLVINSGNGA VEDRKPSGLN GEASKSQEMV HLVNKESSET
       730        740
PDQFMTADET RNLQNVDMKI GV
```

6.2.2 UBL5

```
          10          20          30          40          50          60
MIEVVCNDRL GKKVRVKCNT DDTIGDLKKL IAAQTGTRWN KIVLKKWYTI FKDHVSLGDY
          70
EIHDGMNLEL YYQ
```

7 Bibliography

Allen, C. F. H., J. R. Jr. Byers, W. J. Humphlett. 1963. Oleoyl chloride, *Organic Syntheses*. 4:739

Archer, C.T., L. Burdine, B. Liu, A. Ferdous, S.A. Johnston, and T. Kodadek. 2008. Physical and functional interactions of monoubiquitylated transactivators with the proteasome. *The Journal of biological chemistry*. 283:21789-21798.

Auzenne, E., S.C. Ghosh, M. Khodadadian, B. Rivera, D. Farquhar, R.E. Price, M. Ravoori, V. Kundra, R.S. Freedman, and J. Klostergaard. 2007. Hyaluronic acid-paclitaxel: antitumor efficacy against CD44(+) human ovarian carcinoma xenografts. *Neoplasia*. 9:479-486.

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

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

Baumann, E. 1886. "Ueber eine einfache Methode der Darstellung von Benzoësäureäthern". *Berichte der deutschen chemischen Gesellschaft* 19: 3218.

Bayer, P., A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke, and J. Becker. 1998. Structure determination of the small ubiquitin-related modifier SUMO-1. *Journal of molecular biology*. 280:275-286.

Bedford, L., J. Lowe, L.R. Dick, R.J. Mayer, and J.E. Brownell. 2011. Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets. *Nature reviews. Drug discovery*. 10:29-46.

Beeson, J.G., S.J. Rogerson, B.M. Cooke, J.C. Reeder, W. Chai, A.M. Lawson, M.E. Molyneux, and G.V. Brown. 2000. Adhesion of Plasmodium falciparum-infected erythrocytes to hyaluronic acid in placental malaria. *Nature medicine*. 6:86-90.

Bennett, J.W., J.L. Robertson, D.R. Hospenthal, S.E. Wolf, K.K. Chung, K. Mende, and C.K. Murray. 2010. Impact of extended spectrum beta-lactamase producing Klebsiella pneumoniae infections in severely burned patients. *Journal of the American College of Surgeons*. 211:391-399.

Blundell, C.D., D.J. Mahoney, A. Almond, P.L. DeAngelis, J.D. Kahmann, P. Teriete, A.R. Pickford, I.D. Campbell, and A.J. Day. 2003. The link module from ovulation- and inflammation-associated protein TSG-6 changes conformation on hyaluronan binding. *The Journal of biological chemistry*. 278:49261-49270.

Brown, T.A., T. Bouchard, T. St John, E. Wayner, and W.G. Carter. 1991. Human keratinocytes express a new CD44 core protein (CD44E) as a heparan-sulfate intrinsic membrane proteoglycan with additional exons. *The Journal of cell biology*. 113:207-221.

Camenisch, T.D., and J.A. McDonald. 2000. Hyaluronan: is bigger better? *American journal of respiratory cell and molecular biology*. 23:431-433.

Cavanagh, J., W. J. Fairbrother, A. G. Palmer, M. Rance and N. J. Skleton. 2007. Protein NMR spectroscopy: Principles and Practice. *Elsevier Academic Press.*

CCP4 (Collaborative Computational Project, Number 4). 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallographica Section D, Biological Crystallography* 50:760-763.

Chung, C.T., S.L. Niemela, and R.H. Miller. 1989. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America*. 86:2172-2175.

Cohen, S.X., R.J. Morris, F.J. Fernandez, M. B. Jelloul, M. Kakaris, V. Parthasarathy, V.S. Lamazin, G. J. Kleywegt, A. Perrakis. 2004. Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr*aphy 60:2222-2229.

Collier, G.R., J.S. McMillan, K. Windmill, K. Walder, J. Tenne-Brown, A. de Silva, J. Trevaskis, S. Jones, G.J. Morton, S. Lee, G. Augert, A. Civitarese, and P.Z. Zimmet. 2000. Beacon: a novel gene involved in the regulation of energy balance. *Diabetes*. 49:1766-1771.

Day, A.J., and G.D. Prestwich. 2002. Hyaluronan-binding proteins: tying up the giant. *The Journal of biological chemistry*. 277:4585-4588.

De Maeyer, E., and J. De Maeyer-Guignard. 1992. The growth rate of two transplantable murine tumors, 3LL lung carcinoma and B16F10 melanoma, is influenced by Hyal-1, a locus determining hyaluronidase levels and polymorphism. *International journal of cancer. Journal international du cancer*. 51:657-660.

Dikic, I., S. Wakatsuki, and K.J. Walters. 2009. Ubiquitin-binding domains - from structures to functions. *Nature reviews. Molecular cell biology*. 10:659-671.

Dittmar, G.A., C.R. Wilkinson, P.T. Jedrzejewski, and D. Finley. 2002. Role of a ubiquitin-like modification in polarized morphogenesis. *Science*. 295:2442-2446.

Filion, M.C., and N.C. Phillips. 2001. Pro-inflammatory activity of contaminating DNA in hyaluronic acid preparations. *The Journal of pharmacy and pharmacology*. 53:555-561.

Friedman, J.S., B.F. Koop, V. Raymond, and M.A. Walter. 2001. Isolation of a ubiquitin-like (UBL5) gene from a screen identifying highly expressed and conserved iris genes. *Genomics*. 71:252-255.

Gallatin, W.M., I.L. Weissman, and E.C. Butcher. 1983. A cell-surface molecule involved in organ-specific homing of lymphocytes. *Nature*. 304:30-34.

Gill SC, and P. H. von Hippel. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* 182:319-326.

Goddard T. D. and D. G. Kneller, SPARKY 3, University of California, San Francisco

Goetinck, P.F., N.S. Stirpe, P.A. Tsonis, and D. Carlone. 1987. The tandemly repeated sequences of cartilage link protein contain the sites for interaction with hyaluronic acid. *The Journal of cell biology*. 105:2403-2408.

Goldstein, L.A., D.F. Zhou, L.J. Picker, C.N. Minty, R.F. Bargatze, J.F. Ding, and E.C. Butcher. 1989. A human lymphocyte homing receptor, the hermes antigen, is related to cartilage proteoglycan core and link proteins. *Cell*. 56:1063-1072.

Hochstrasser, M. 2009. Origin and function of ubiquitin-like proteins. *Nature*. 458:422-429.

Hoeller, D., and I. Dikic. 2009. Targeting the ubiquitin system in cancer therapy. *Nature*. 458:438-444.

Hurley, J.H., S. Lee, and G. Prag. 2006. Ubiquitin-binding domains. *The Biochemical journal*. 399:361-372.

Iwai, K., and F. Tokunaga. 2009. Linear polyubiquitination: a new regulator of NF-kappaB activation. *EMBO reports*. 10:706-713.

Jackson, D.G. 2003. The lymphatics revisited: new perspectives from the hyaluronan receptor LYVE-1. *Trends in cardiovascular medicine*. 13:1-7.

Jackson, D.G., R. Prevo, S. Clasper, and S. Banerji. 2001. LYVE-1, the lymphatic system and tumor lymphangiogenesis. *Trends in immunology*. 22:317-321.

Jaracz, S., J. Chen, L.V. Kuznetsova, and I. Ojima. 2005. Recent advances in tumor-targeting anticancer drug conjugates. *Bioorganic & medicinal chemistry*. 13:5043-5054.

Jentsch, S., and G. Pyrowolakis. 2000. Ubiquitin and its kin: how close are the family ties? *Trends in cell biology*. 10:335-342.

Jerabek-Willemsen, M., C.J. Wienken, D. Braun, P. Baaske, and S. Duhr. 2011. Molecular interaction studies using microscale thermophoresis. *Assay and Drug Development Technologies*. 9:342-353.

Kabsch, W. 2010a. Integration, scaling, space-group assignment and post-refinement. *Acta crystallographica. Section D, Biological crystallography*. 66:133-144.

Kabsch, W. 2010b. Xds. Acta crystallographica. Section D, Biological crystallography. 66:125-132.

Kantham, L., L. Kerr-Bayles, N. Godde, M. Quick, R. Webb, T. Sunderland, J. Bond, K. Walder, G. Augert, and G. Collier. 2003. Beacon interacts with cdc2/cdc28-like kinases. *Biochemical and biophysical research communications*. 304:125-129.

Kawashima, H., M. Hirose, J. Hirose, D. Nagakubo, A.H. Plaas, and M. Miyasaka. 2000. Binding of a large chondroitin sulfate/dermatan sulfate proteoglycan, versican, to L-selectin, P-selectin, and CD44. *The Journal of biological chemistry*. 275:35448-35456.

Keren, H., G. Lev-Maor, and G. Ast. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nature reviews. Genetics*. 11:345-355.

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

Kreil, G. 1995. Hyaluronidases--a group of neglected enzymes. *Protein science : a publication of the Protein Society*. 4:1666-1669.

Kurland, C., and J. Gallant. 1996. Errors of heterologous protein expression. *Current opinion in biotechnology*. 7:489-493.

Leonelli, F., A. La Bella, L.M. Migneco, and R.M. Bettolo. 2008. Design, synthesis and applications of hyaluronic acid-paclitaxel bioconjugates. *Molecules*. 13:360-378.

Lesley, J., N. English, A. Perschl, J. Gregoroff, and R. Hyman. 1995. Variant cell lines selected for alterations in the function of the hyaluronan receptor CD44 show differences in glycosylation. *The Journal of experimental medicine*. 182:431-437.

Liu, D., E. Pearlman, E. Diaconu, K. Guo, H. Mori, T. Haqqi, S. Markowitz, J. Willson, and M.S. Sy. 1996. Expression of hyaluronidase by tumor cells induces angiogenesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America*. 93:7832-7837.

Luders, J., G. Pyrowolakis, and S. Jentsch. 2003. The ubiquitin-like protein HUB1 forms SDS-resistant complexes with cellular proteins in the absence of ATP. *EMBO reports*. 4:1169-1174.

Luo, Y., and G.D. Prestwich. 1999. Synthesis and selective cytotoxicity of a hyaluronic acid-antitumor bioconjugate. *Bioconjugate chemistry*. 10:755-763.

Luo, Y., M.R. Ziebell, and G.D. Prestwich. 2000. A hyaluronic acid-taxol antitumor bioconjugate targeted to cancer cells. *Biomacromolecules*. 1:208-218.

Maeda, H., L.W. Seymour, and Y. Miyamoto. 1992. Conjugates of anticancer agents and polymers: advantages of macromolecular therapeutics in vivo. *Bioconjugate chemistry*. 3:351-362.

Mahoney, D.J., R.T. Aplin, A. Calabro, V.C. Hascall, and A.J. Day. 2001. Novel methods for the preparation and characterization of hyaluronan oligosaccharides of defined length. *Glycobiology*. 11:1025-1033.

Makrides, S.C. 1996. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological reviews*. 60:512-538.

Markovic-Housley, Z., G. Miglierini, L. Soldatova, P.J. Rizkallah, U. Muller, and T. Schirmer. 2000. Crystal structure of hyaluronidase, a major allergen of bee venom. *Structure*. 8:1025-1035.

McNally, T., Q. Huang, R.S. Janis, Z. Liu, E.T. Olejniczak, and R.M. Reilly. 2003. Structural analysis of UBL5, a novel ubiquitin-like modifier. *Protein science : a publication of the Protein Society*. 12:1562-1566.

Meyer, B., and T. Peters. 2003. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition English*. 42:864-890.

Mishra, S.K., T. Ammon, G.M. Popowicz, M. Krajewski, R.J. Nagel, M. Ares, Jr., T.A. Holak, and S. Jentsch. 2011. Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature*. 474:173-178.

Moeslein, F.M., M.P. Myers, and G.E. Landreth. 1999. The CLK family kinases, CLK1 and CLK2, phosphorylate and activate the tyrosine phosphatase, PTP-1B. *The Journal of biological chemistry*. 274:26697-26704.

Mori, S., C. Abeygunawardana, M.O. Johnson, and P.C. van Zijl. 1995. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *Journal of magnetic resonance. Series B*. 108:94-98.

Naor, D., S. Nedvetzki, I. Golan, L. Melnik, and Y. Faitelson. 2002. CD44 in cancer. *Critical reviews in clinical laboratory sciences*. 39:527-579.

Naor, D., R.V. Sionov, and D. Ish-Shalom. 1997. CD44: structure, function, and association with the malignant process. *Advances in cancer research*. 71:241-319.

Neises, B., and W. Steglich. 1978. 4-Dialkylaminopyridines as Acylation Catalysts .5. Simple Method for Esterification of Carboxylic-Acids. *Angewandte Chemie International Edition English* 17:522-524.

Ogino, S., N. Nishida, R. Umemoto, M. Suzuki, M. Takeda, H. Terasawa, J. Kitayama, M. Matsumoto, H. Hayasaka, M. Miyasaka, and I. Shimada. 2010. Two-state conformations in the hyaluronan-binding domain regulate CD44 adhesiveness under flow condition. *Structure*. 18:649-656.

Oh, E.J., K. Park, K.S. Kim, J. Kim, J.A. Yang, J.H. Kong, M.Y. Lee, A.S. Hoffman, and S.K. Hahn. 2010. Target specific and long-acting delivery of protein, peptide, and nucleotide therapeutics using hyaluronic acid derivatives. *Journal of controlled release : official journal of the Controlled Release Society*. 141:2-12.

Pellecchia, M., D.S. Sem, and K. Wuthrich. 2002. NMR in drug discovery. *Nature reviews. Drug discovery*. 1:211-219.

Pham, H.T., N.L. Block, and V.B. Lokeshwar. 1997. Tumor-derived hyaluronidase: a diagnostic urine marker for high-grade bladder cancer. *Cancer research*. 57:778-783.

Piotto, M., V. Saudek, and V. Sklenar. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*. 2:661-665.

Platt, V.M., and F.C. Szoka, Jr. 2008. Anticancer therapeutics: targeting macromolecules and nanocarriers to hyaluronan or CD44, a hyaluronan receptor. *Molecular pharmaceutics*. 5:474-486.

Popowicz, G.M., A. Czarna, S. Wolf, K. Wang, W. Wang, A. Domling, and T.A. Holak. 2010. Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery. *Cell Cycle*. 9:1104-1111.

Pouyani, T., and G.D. Prestwich. 1994. Functionalized derivatives of hyaluronic acid oligosaccharides: drug carriers and novel biomaterials. *Bioconjugate chemistry*. 5:339-347.

Ramage, R., J. Green, T.W. Muir, O.M. Ogunjobi, S. Love, and K. Shaw. 1994. Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *The Biochemical journal*. 299 (Pt 1):151-158.

Ramelot, T.A., J.R. Cort, A.A. Yee, A. Semesi, A.M. Edwards, C.H. Arrowsmith, and M.A. Kennedy. 2003. Solution structure of the yeast ubiquitin-like modifier protein Hub1. *Journal of structural and functional genomics*. 4:25-30.

Rosenberg, A. H., E. Goldmann, J. J. Dunn, F. W. Studier, G. Zubay. 1993. Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *Journal of Bacteriology*. 175:716-722.

Schagger, H., and G. von Jagow. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical biochemistry*. 166:368-379.

Schotten, C. 1884. Ueber die Oxydation des Piperidins. *Berichte der deutschen chemischen Gesellschaft* 17: 2544

Shuker, S.B., P.J. Hajduk, R.P. Meadows, and S.W. Fesik. 1996. Discovering high-affinity ligands for proteins: SAR by NMR. *Science*. 274:1531-1534.

Sironen, R.K., M. Tammi, R. Tammi, P.K. Auvinen, M. Anttila, and V.M. Kosma. 2011. Hyaluronan in human malignancies. *Experimental cell research*. 317:383-391.

Song, J., Z. Zhang, W. Hu, and Y. Chen. 2005. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *The Journal of biological chemistry*. 280:40122-40129.

Stamenkovic, I., M. Amiot, J.M. Pesando, and B. Seed. 1989. A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. *Cell*. 56:1057-1062.

Stetler-Stevenson, W.G., S. Aznavoorian, and L.A. Liotta. 1993. Tumor cell interactions with the extracellular matrix during invasion and metastasis. *Annual review of cell biology*. 9:541-573.

Stoll, R., C. Renner, S. Hansen, S. Palme, C. Klein, A. Belling, W. Zeslawski, M. Kamionka, T. Rehm, P. Muhlhahn, R. Schumacher, F. Hesse, B. Kaluza, W. Voelter,

R.A. Engh, and T.A. Holak. 2001. Chalcone derivatives antagonize interactions between the human oncoprotein MDM2 and p53. *Biochemistry*. 40:336-344.

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

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

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

Toole, B.P., A. Zoltan-Jones, S. Misra, and S. Ghatak. 2005. Hyaluronan: a critical component of epithelial-mesenchymal and epithelial-carcinoma transitions. *Cells, tissues, organs*. 179:66-72.

Wienken, C.J., P. Baaske, U. Rothbauer, D. Braun, and S. Duhr. 2010. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*. 1:100.

Wuthrich, K. 1986. NMR of Proteins and Nucleic Acids. *Wiley-Interscience, New York*

Yashiroda, H., and K. Tanaka. 2004. Hub1 is an essential ubiquitin-like protein without functioning as a typical modifier in fission yeast. *Genes to cells : devoted to molecular & cellular mechanisms*. 9:1189-1197.

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

Zsoldos, Z., D. Reid, A. Simon, S.B. Sadjad, and A.P. Johnson. 2007. eHiTS: a new fast, exhaustive flexible ligand docking system. *Journal of molecular graphics & modelling.* 26:198-212.

www.glycoforum.gr.jp/science/hyaluronan/HA28/HA28E.htmlS (2004)