Charles University in Prague

Faculty of Science

Department of Biochemistry

α-*N*-Acetylgalactosaminidase as a tools in the synthesis of complex oligosaccharide immune stimulators

Summary of Ph. D. Thesis

Mgr. Hynek Mrázek

Supervisor: Prof. RNDr. Karel Bezouška, Dsc.

Prague 2011

Introduction

Glycoproteins

Glycoproteins consist of proteins to which carbohydrate are covalently linked. The distinction between proteoglycans and glycoproteins residues is in the level and types of carbohydrate modification. Carbohydrates are linked to the protein component through either *O*-glycosidic or *N*-glycosidic bonds. The *N*-glycosidic linkage is through the amide group of asparagine. The *O*-glycosidic linkage is to the hydroxyl of serine, threonine or hydroxylysine.

The predominant carbohydrate attachment in glycoproteins of mammalian cells is via *N*-glycosidic linkage. The site of carbohydrate attachment to *N*-linked glycoproteins is found within a consensus sequence of amino acids, N-X-S(T), where X is any amino acid except proline. While the *N*-glycosylation is governed by the above rules, no exact rules were found for glycosylation of *O*-type. This process is the step-wise addition of the sugar residues directly onto the polypeptide chain. Biosynthesis of glycoproteins occurs via protein glycosylation (the addition of chains of sugar units, or oligosaccharides, to proteins). Protein glycosylation is a group of complex posttranslational modifications that occur through the function of many enzymes working together in the endoplasmic reticulum (ER) and Golgi apparatus.

Glycoside hydrolases

Glycoside hydrolases are classified into EC 3.2.1 as enzymes catalyzing the hydrolysis of *O*- or *S*-glycosides yielding smaller sugar moieties. They can be classified as either *retaining* or *inverting* enzymes. Glycoside hydrolases can also be classified as *exo* or *endo* acting, dependent upon whether they act at the (usually non-reducing) end or in the middle, respectively, of an oligo/polysaccharide chain.

α-N-Acetylgalactosaminidase

α-N-Acetylgalactosaminidase (α-NAGA; EC.3.2.1.49) is an exoglycosidase specific for the hydrolysis of terminal GalNAc α-linked to amino acids serin or threonine, or to various sugar chains. According to enzyme nomenclature of IUB-MB (International Union of Biochemistry and Molecular Biology) these enzymes belong to Hydrolase-Glycosidase-Glycosidase hydrolyzing *O*- and *S*- glycosidic linkage. According to CAZY system prokaryotic α-NAGA belong to enzyme family 36 (Clan GH-D) and eukaryotic α-NAGA to enzyme family 27 the same clan like prokaryotic enzyme.

Synthesis of glycopeptides and glycoproteins

Two major approaches can be distinguished: the linear and the convergent assembly. In the linear assembly, carbohydrates are coupled to amino acids to give modified amino acids carrying either mono- or oligosaccharides. These are used as building blocks in solution or solid-phase peptide synthesis (SPPS) to provide glycoconjugates and, respectively.

In convergent approaches, carbohydrates are coupled to presynthesized peptides or to proteins. Depending on the type of chemistry used for carbohydrate attachment, there might be a need for protecting groups at the peptide and carbohydrate, an approach which is confined to synthetic peptides. Conjugation of carbohydrates to full-length proteins is possible via chemoselective ligation to amino acids with a unique reactivity, for example cysteine residues. Compared to linear approaches, the convergent synthesis of glycopeptide and glycoprotein mimetics offers greater flexibility with respect to the sugars attached to a peptide. Thus, the preparation of several well-defined glycoforms of the same peptide/protein becomes possible. If chemoselective ligation reactions are employed for the sugar attachment, it is possible to modify unprotected peptides and even whole proteins.

Convergent approaches using chemoselective ligation reactions provide access to homogeneous glycoprotein mimetics that are likely to impact our understanding of how specific glycoforms mediate physiological processes. Despite the progress made in the field, many challenges remain, e.g., the development of methods for the controlled introduction of multiple (different) glycans into proteins.

NK cells (natural killer cells)

NK cells have been described only in 1975 as lymphocytes, which are able to kill without prior stimulation, proliferation and differentiation of some tumor or virus-infected cells. They are large granular lymphocytes that develop in the bone marrow with other cells from pluripotent stem cells. Developmentally they are closer to T lymphocytes - this relationship became apparent by the recent discovery of T lymphocytes, which bear on their surface markers characteristic of NK cells (NKR-P1). This group was called NKT and cells are similarly as the Tc lymphocytes involved in cytotoxic reactions, which are directed against tumor or virus-infected altered cells. Production of cytokines is also important in the regulation of specific immune responses, cell differentiation, and cell adhesion. The most important cytokine produced by NK cells, is INF-γ (Horejsi and Bartunkova, 2001), which stimulates macrophages.

C-type lectins

The C-type lectins were first properly defined in the 80th of the 20th century (Drickamer, 1988). A common feature of this family is called the CRD (Carbohydrate Recognication Doman). CRD requires calcium ions for its binding activity. The lenght of CRD of C-type lectins is around 125 amino acids. CRD is composed of a highly conserved combination of two α-helixes and two anti-parallel β-sheets connected by random coils with 14 invariant and another 18 highly conserved amino acids (Bezouska et. al. 1991). This carbohydrate

domain is stabilized by two or more disulphide bonds. Different binding specificity is provided by the participation of variable loops in binding site formation. During evolution some members of C-type lectins have lost the ability to bind carbohydrates or calcium, instead they can gained another substrate specifity (Drickamer, 1999).

Aim of the study

The aim of this study was to find suitable expression system for recombinat production of enzymatic active α-NAGA with dual activity. The second aim was characterize enzymatic and biochemical properties of this enzyme. In order to achieve this goal, the following specific aims heve been adopted for this study.

- To determined amino acid sequence of native of α-NAGA isolated from filamentous fungi *Aspergillus niger.*
- To develop expression system suitable for recombinant expression of α-NAGA in enzymatic active form.
- To develop the optimal conditions providing sufficient amount of α-NAGA.
- To develop purification protocol providing sufficient amount of α-NAGA for functional and structural studies.
- To investigate biochemical and enzymatic properties of α-NAGA.
- To determined *N*-glycosylation sites of α-NAGA.

In order to achieve these goals, we have adopted the suitable methodology including the bacterial and yeast expression system, different biochemical and enzymatic techniques (chromatography, analytic ultracentrifuge, mass spectrometry, electron microscopy etc.).

Results

Determination of amino acids sequence of α-NAGA and computer modeling

This issue has been addressed experimentally by screening a large library of filamentous fungi (42 strains) and a series of inducers and cultivation conditions for the presence of α-NAGA activity. Only a single enzyme from *A niger* CCIM K2 demonstrated this activity together with the one of α-GA. The enzyme was isolated to very high purity using chromatofocusing with narrow range polybuffers, but the highly purified enzyme still had the dual activity. As the final separation technique applied at small scale, two dimensional electrophoresis using narrow pH strips was applied that resolved the preparation into two spots (Weignerova et. al. 2008). Both spots were analyzed by N-terminal sequencing and by mass spectrometry sequencing.

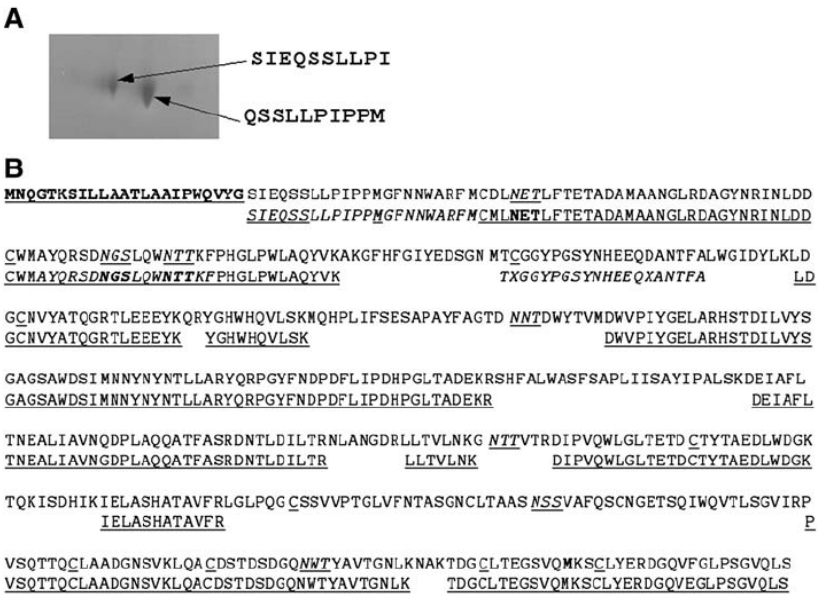

Figure 1. Summary of sequencing data for enzyme acting as α-NAGA. (A) Separation of prepared enzymes differing in their indicated N-terminal sequences by 2D electrophoresis, indicating minor heterogeneity in enzyme preparation confirmed by N-terminal sequencing. (B) Upper lane indicates amino acid sequence of *aglA* gene from *Aspergillus niger* with signal peptide sequence bold. Lower lane shows summary of sequence data obtained by N-terminal Edman degradation of entire enzyme or isolated peptides after CNBr cleavage (italics) or by mass spectrometric analysis of peptide fragments.

Alignment of the found *AglA* sequence against structurally solved α-*N*-acetylgalactosaminidases revealed high sequence identity allowing to construct a molecular model of the enzyme (Figure 2). The molecular architecture of α-NAGA is composed of catalytic (mellibiase) domain containing TIM barrel, followed by a beta sandwith, and the C-terminal ricin domain. Related anzymes such as *AglB* contain similar catalytic domain buy lack the C-terminal ricin-like domain (Kulik et. al. 2010).

\

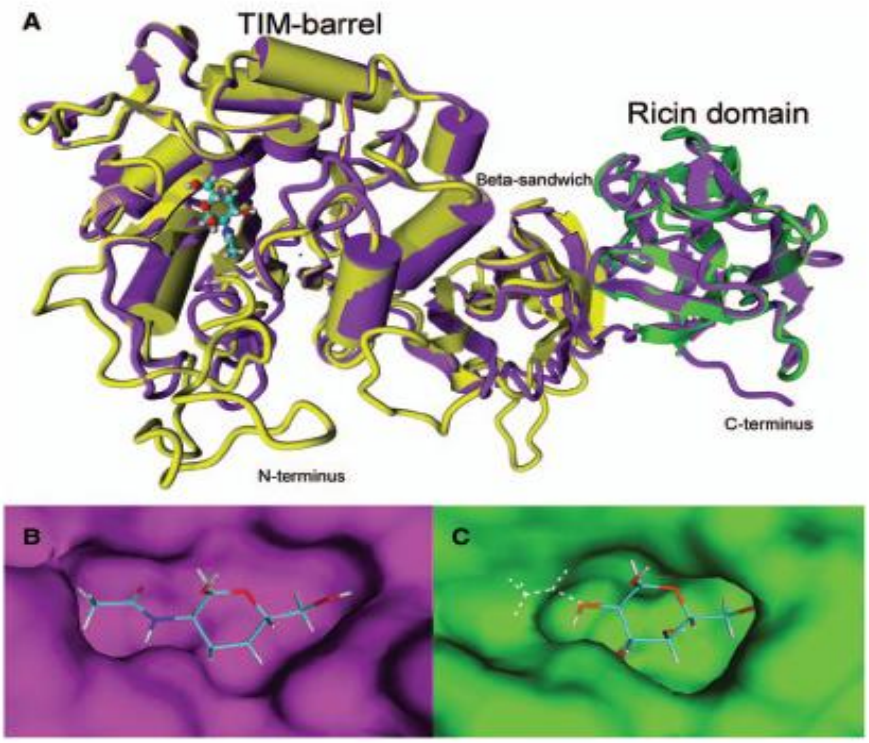

Figure 2. Structure of the α-NAGA from *Aspergillus niger*. (A) Overall fold shows a TIM-barrel with the active site at the N-terminus, a small domain of eight antiparallel β-strands packed in β-sandwich in the middle, and a ricin-like domain on the right. The generated model (magenta) is overlaid with the crystal structure of the homologs α-*N*-acetylgalactosaminidase from chicken (yellow), with α-*N*-acetylgalactosamine and the ricin-like domain from the xylanase from *Streptomyces olivaceoviridis* E-86 (green). (B) and (C) Molecular surface of the active site of *aglA* enzyme (magenta) and *aglB* enzyme (yellow) with *o*-NP-α-GalNAc. The active site of *aglA* enzyme has extra space for accommodating the *N*-acetyl-group of the substrate, while in *aglB* enzyme this space is occupied by Trp205.

***Recombinant expression of Aspergillus niger α*-NAGA**

α-NAGA isolated from *Aspergillus niger* displayed unique properties. The production of α-NAGA carried out in cultivation medium 6 days at 26˚C as the inductor was used soy flour.

The enzyme was optimally active at 55˚C and at of pH 1.8. The enzyme deglycosylation shifted the pH optimum to 1.5. Both enzymes forms were stable at of pH 1.5 - 4 and 4˚C. The enzyme was maintained without loss activity several months. The purified enzyme exhibited a Km value of 0.73 mmol/l for *o*-NP-α-GalNAc in 50 mM citrate-phosphate buffer (pH 3.5) at 35°C (Weinerova et. al. 2008). The dual α-GA activity was also found in the native form of α-NAGA. The dual activity of α-NAGA was explained by evolutionary mapping of this enzyme class and computer model. On the basic of this mapping was observed that the α-NAGA is evolved from α-GA described above. The large differ in active site between α-GA and α-NAGA is so-called binding pocket for *N*-acetylamine. This binding pocket was revealed by substrate docking experiments.

On the basic of these properties were decided to prepare the recombinat form of this enzyme. Since the deglycosylated α-NAGA was active several months without significantly loss of enzyme activity, prokaryotic expression system *Escherichia coli* was chose. The gene of interest was gained from total RNA isolated from *Aspergillus niger* and reverze transcripted into DNA sequence. The gene of interest was cloned into prokaryotic expression vector. Two prokaryotic expression strains were used. Firstly, *E. coli* BL-21 (DE3) Gold. In the case of this strain we observed that the α-NAGA is produced in form of inclusion bodies without enzymatic activity. We have a lot of experiences with renaturation of recombinant proteins in our laboratory. For these reasons were tried many renaturation protocols and techniques. Unfortunately, not one was successful. The *E. coli* ArticExpressCells were used

as a second possibility of expression in prokaryotic expression system. The advantage of this system is low temperature of production around 12˚C. The low temperature should help to correct protein folding and preclude the aggregation of proteins to inclusion bodies. Unfortunately, in this case was no observed the expression of α-NAGA.

This failure can be explained a large glycosylation of the α-NAGA, this posttranslation modification is probably very important in refolding process.

Of these reasons we decided for using of yeast expression system *Pichia pastoris*. Unfortunately, the expression was not success.

As a last expression system was used *Saccharomyces cerevisiae* (Ashida et. al. 2000). The production of α-NAGA in this expression system was successful. The recombinant enzyme was purified and biochemically and enzymatically characterised. Biochemical and enzymatic properties of recombinant protein was identical except the subunit structure and activity in neutral pH.

The recombinant enzyme occurs in two active forms monomeric and dimeric. The explanation of these two forms can be in overexpression and the placement of this enzyme inside the cell. Part of enzyme can occur in the membrane organels, where are other conditions for refolding of the proteins. The second part of enzyme can be in cytoplasm, where is reduction environment. This hypothesis confirmed the experiments with addition of detergent to lysate mixture. The enzymatic activity increases two times after addition of detergent. The second differ between wild-type and recombinant α-NAGA is activity in neutral pH. The recombinant enzyme displays around 15% activity in this pH unlike the wild-type which has no activity.

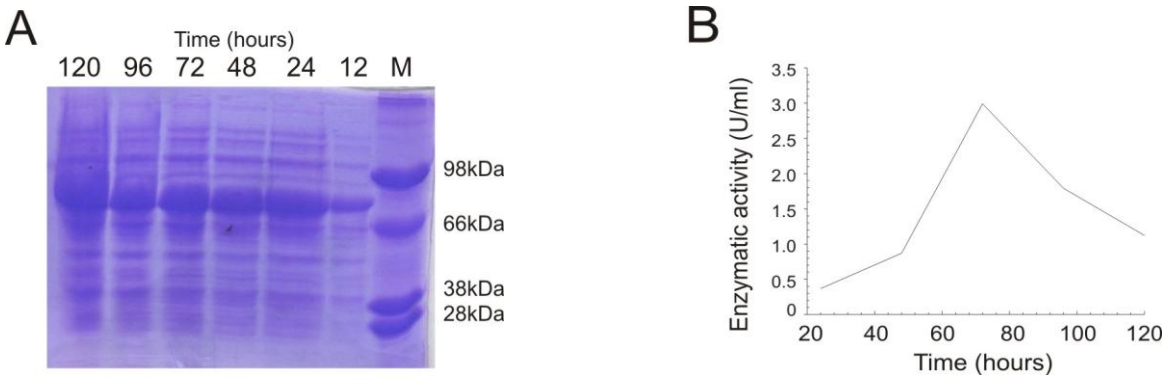

Figure 3. Time profile of the recombinant intracellular α-NAGA production by *S. cerevisiae* W303: (A) SDS-PAGE electrophoresis (B**)** enzyme activity. Time optimization was carried out at 30˚C. Aliquots of the cell culture were harvested 12, 24, 48, 72, 96 and 120 hours after transfer to the SC medium with galactose as an inducer. The recombinant α-NAGA was identified as a band with an apparent molecular mass of approx. 76 kDa.

Interestingly, the addition of the detergent into the lysis buffer caused two fold increase in activity.

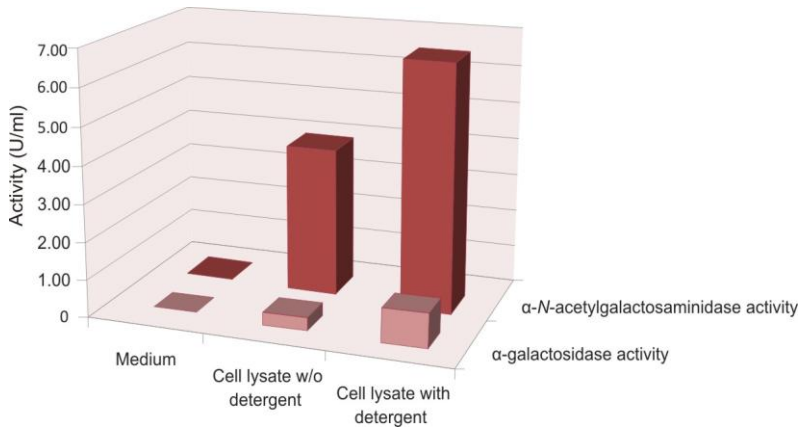

Figure 4. α-NAGA and α-GA activity measured in cell lysates with and without dodecylmaltoside (detergent) in lysis buffer.

Step	Protein (mg)	Activity (U)	Spec.activity (U mg^{-1})	Purity (fold)	Yield (%)
Cell lysate	378.0	150.0	0.4	1.0	100
Phenyl-sepharose HR	31.4	102.3	3.3	8.3	68.2
S-Sepharose FF	12.8	81.0	6.3	15.8	54.0
Superdex 200	1.5	41.9	27.9	69.9	27.9
Mono P 5/200	0.4	18.1	42.3	105.9	12.1

Table 1. Purification of the recombinant α-NAGA from *Saccharomyces cerevisiae* W303-1A. Purity is related to the starting material.

The pH profil of recombinant α-NAGA was changed, the recombinant enzyme more active in neutral pH compare to wild-type α-NAGA. This fact predetermines this enzyme for the red blood cell gropu A transformation in to the group of H(0), beeing a universal donor.

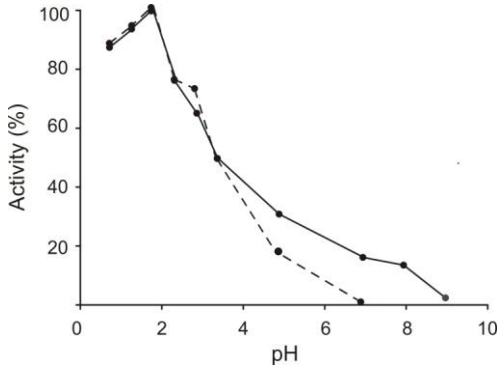

Figure 5. Effect of pH on the activity of wild α-NAGA (- -●- -) and recombinant α-NAGA (─●─).

The molecular weight of the native and recombinant α-NAGA was investigated using gel filtration and analytical ultracentrifuge (Schutz P. 2003).

The gel filtration of the recombinant α-NAGA showed two active forms with estimated molecular mass of approximately 70 kDa nad 130 kDa. The wild α-NAGA occured only as 70 kDa monomer. The analytical centrifugation confirmed the molecular mass of both enzymes (Schutz P. 2000).

α-NAGA was found to be a glycoprotein. There are eight of potential *N*-glycosylation sites located at Asn 14, 52, 58, 88, 168, 320, 401 and 456.

The recombinant α-NAGA was treated with Endo Hf and PNGase F. After Endo Hf treatment we identified peaks corresponding to the peptides containing *N*-glycosylation sites based on increased *m/z* of peptides containing *N*-acetylglucosamine residues (203.079 Da). When α-NAGA was treated with PNGase F and digested with AspN, new peaks appeared. These peaks corresponded to the peptides that arose from the conversion of asparagine to asparatic acid as shown in Figure 6b. Six asparagines of the eight potential *N*-glycosylation sites in the α-NAGA located at residues 14, 52, 58, 88, 320 and 456 were found to be glycosylated.

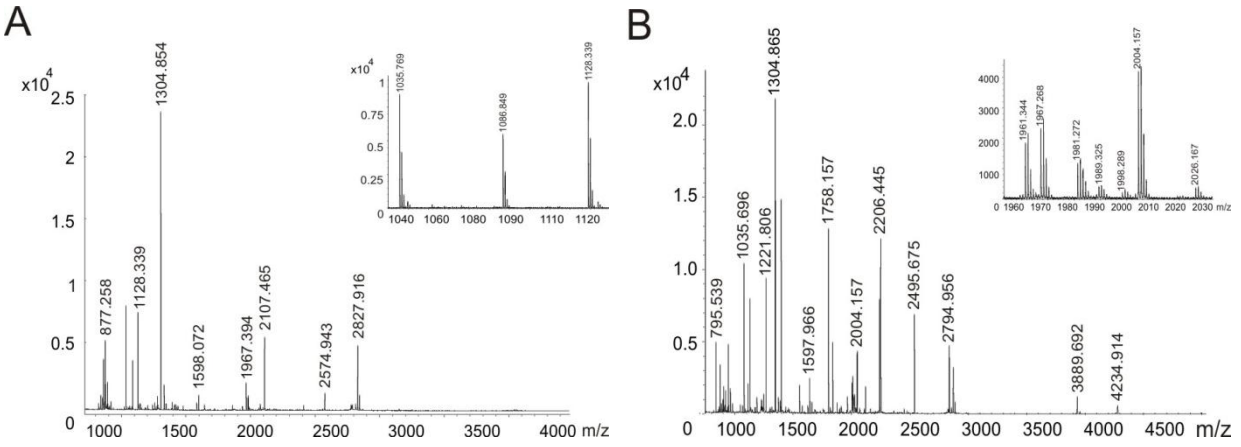

Figure 6. Determination of *N*-glycosylation sites. (A) α-NAGA was digested with Asp-N and treated with PNGase F. MALDI-TOF mass spectrometry analysis showed that the high mass peaks had disappeared, while a new peptide signal at *m/z* 1128.337 appeared. This peak corresponded to the peptide residue 310-319 (calculated *m/z* for $[M+H]^+$ 1127.337) with a 1 Da mass increase from the conversion of asparagines to aspartic acid by PNGase F. (B) α-NAGA was digested with Asp-N and deglycosylated with Endo Hf, which cleaves the bond between two GlcNAc units attached to asparagines. The final peptide has a mass increase of 203,079 Da. The peptide signal at *m/z* 2004.157 corresponded to peptide residue 310-325 (calculated *m/z* for $[M+H]^+$ 1801.076) with a 203.079 Da mass increase for GlcNAc.

In conclusion, we prepared stable, active recombinant α-NAGA in large quantities in a simple eukaryotic system of *Saccharomyces cerevisiae*. The notable advantage of our expression system is in shorter production times, and, up to fourfold increase of the enzyme yields compared to the native production system. Unique properties of this enzyme can find a use for the enzymatic synthesis of various carbohydrate structures and for transformation of the red blood cell group A to the group of H (0), the universal donor.

Cooperation between Subunits of CD69 receptor

We reported that the binding of HexNAc to soluble CD69 is highly cooperative at molecular level, and this cooperativity is not seen for Q93A and R134A mutants with disturbed formation of noncovalent dimers. Similarly at the cellular level, efficient signaling after CD69 crosslinking by antibody or bivalent ligand is diminished for the above mutants with a damaged subunit cross-talk more dramatically than for CD69 bearing C68A mutation, and thus lacking the disulfide bridge forming the covalent dimer identified previously as the critical signaling element.

Several constructs were generated for preparation of highly stable soluble recombinant CD69 proteins, which is suitable for ligands identification experiments. Preliminary ligand binding experiments were performed to evaulate the ability of these constructs to bind calcium nad monosaccharide units. In the case of the ability to bind calcium there has not been difference between the covalent dimeric protein and noncovalent dimeric protein when compared to the monomeric form. Each of these proteins bound 1 mol of calcium/mol of CD69 subunit with Kd of approximately 58 μM. On the other hand, significant differences between these protein constructs were observed with regard to the binding of HexNAc. While the IC50 values for the soluble monomeric CD69 with regard to binding of the two active HexNAc, D-GlcNAc

and D-GalNAc, were each approximately 10^{-5} M, these values were about 10 times lower for the covalent dimeric protein and about 100 times lower for the other two highly stable noncovalent dimeric proteins.

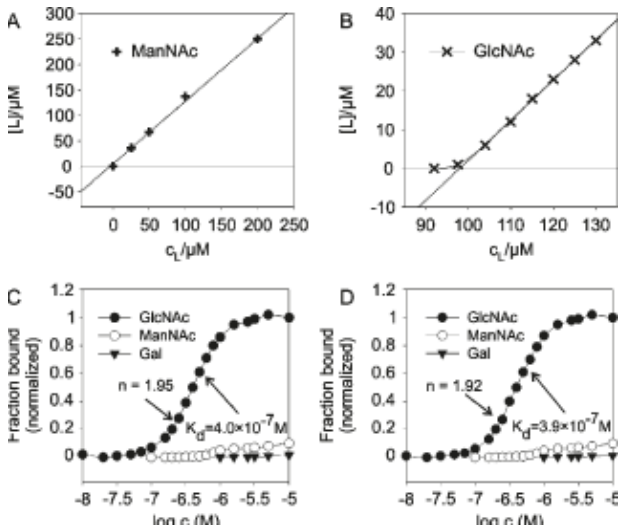

Figure 7. Measurements of direct interaction of soluble noncovalent dimer CD69 with ManNAc and GlcNAc. (A, B) NMR titration of soluble CD69 with ManNAc and GlcNAc, respectively. (C, D) Concentration dependence of receptor saturation measured by equilibrium dialysis and tryptophan fluorescence quenching, respectively, using GlcNAc, ManNAc, and Gal as indicated.

Synthetic GlcNAc Based Fully Branched Tetrasaccharide, a Mimetic of the Endogenous Ligands fo CD69

Three series of carbohydrate ligands for NK cells receptors NKR-P1 and CD69 were prepared (Veprek et. al. 2006). Individual compounds were tested as inhibitors of binding of the soluble radiolabeled receptor. In the case of NKR-P1, the results indicate that synthesized compounds were average or poor ligands compared to the GlcNAc control and the *β1-4* linkage is preferred to other linkages. Branching of the oligosaccharide resulted in significant decrease in the inhibitory potencies independently of the series used. On the other hand more interesting results were obtained in the case of CD69 receptor. Only minor differences have been found in the linear GlcNAc/GalNAc series compared to

the GlcNAc monosaccharide control. However, hierarchical increase in inhibitory potencies has been found in the branched GlcNAc/GalNAc series. On the basic of these results the detailed structure-activity studies were performed with branched GlcNAc.

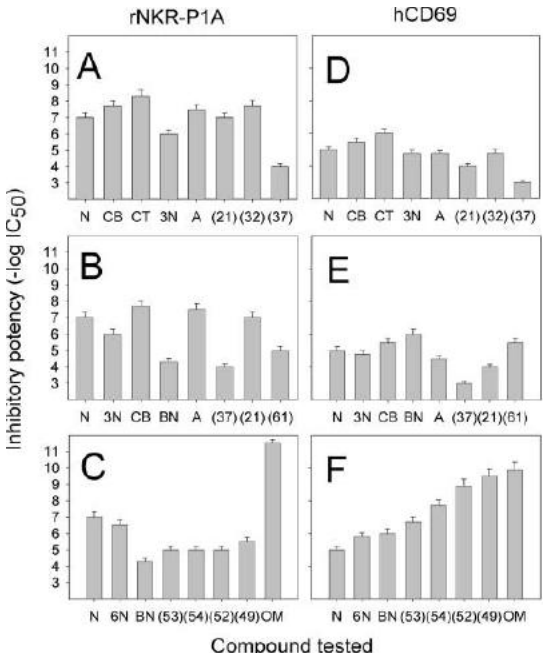

Figure 8. Biological testing of the synthesized HexNAc based oligosaccharides using inhibition assay. Indicated compounds were tested as the inhibitors of binding of the radiolabeled rNKR-P1A (left) or hCD69 (right) to the high affinity GlcNAc23BSA ligand. From the complete inhibition curves, IC% values were calculated.

References

Ashida H., Tamaki H., Fujimoto T., Yamamoto K., Kumagai H. (2000). Molecular cloning of cDNA encoding α-N-acetylgalactosaminidase from Acremonium sp. and its expression in yeast. *Arch. Biochem. Biophys*. **384**: 305–310.

Bezouska, K.; Crichlow, G. V.; Rose, J. M.; Taylor, M. E.; Drickamer, K. (1991). Evolutionary conservation of intron position in a subfamily of genes encoding carbohydrate-recognition domains. *J. Biol. Chem*. **266**: 11604-11609.

Drickamer K. (1988). Two distinct classes of carbohydrate-recognition domains in animals lectins. *J Biol. Chem*. **263**: 9557-9560.

Drickamer K. (1999). C-type lectin-like domains. *Curr. Opin. Struct. Biol*. **9**: 585-590.

Hořejší V., Bartůňková J. (2001). *Základy imunologie*. Triton. Praha.

Kulik N., Weignerova L., Filipi T., Pompach P., Novak P., Mrazek H., Slamova K., Bezouska K., Kren V., Ettrich R. (2010). The α-galactosidase type A gene aglA from Aspergillus niger encodes a fully functional α-N-acetylgalactosaminidase. *Glycobiology* **20**: 1410-1419.

Schutz P. (2000). Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modelling. *Biophys. J* . **78**: 1606-1619.

Schutz P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem*. **320**: 104-124.

Veprek, P.; Hajduch, M.; Dzubal, P.; Kuklik, R.; Polakova, J.; Bezouska, K. (2006). Comblike dendrimers containing Tn antigen modulate natural killing and induce the production of Tn specific antibodies. *J Med Chem* **49** (21): 6400-6407.

Weignerová L., Filipi T., Manglová D., Kren V. (2008). Induction, purification and characterization of alpha-N-acetylgalactosaminidase from Aspergillus Niger. *Appl. Microbiol. Biotechnol*. **79**: 769-774.

List of publication

Kulik N., Weignerova L., Filipi T., Pompach P., Novak P., Mrazek H., Slamova K., Bezouska K., Kren V., Ettrich R. (2010). The α-galactosidase type A gene aglA from Aspergillus niger encodes a fully functional α-N-acetylgalactosaminidase. *Glycobiology* **20**: 1410-1419.

Rey M., Mrazek H., Pompach P., Novak P., Pelosi L., Brandolin G., Forest E., Havlicek V., Man P. (2010). Effective removal of nonionic detergents in protein mass spectrometry, hydrogen/deuterium exchange, and proteomics. *Anal. Chem.* **82**: 5107-5116.

Kovalova A., Ledvina M., Saman D., Zyka D., Kubickova M., Zidek L., Sklenar V., Pompach P., Kavan D., Bily J., Vanek O., Kubinkova Z., Libigerova M., Ivanova L., Antolikova M., Mrazek H., Rozbesky D., Hofbauerova K., Kren V., Bezouska K. (2010). Synthetic N-acetyl-D-glucosamine based fully branched tetrasaccharide, a mimetic of the endogenous ligand for CD69, activates CD69+ killer lymphocytes upon dimerization via a hydrophilic flexible linker. *J. Med. Chem.* **53**: 4050-4065.

Kavan D., Kubickova M., Bily J., Vanek O., Hofbauerova K., Mrazek H., Rozbesky D., Bojarova P., Kren V., Zidek L., Sklenar V., Bezouska K. (2010). Cooperation between subunits is essential for high-affinity binding of N-acetyl-D-hexosamines to dimeric soluble and dimeric cellular forms of human CD69. *Biochemistry* **49**: 4060-4067.

Mrazek H., Benada O., Man P., Vanek O., Kren V., Bezouska K., Weignerova L. (2011). Facile production of A*spergillus niger* α-*N*-acetylgalactosaminidase in yeast. *Submitted for Protein expression and purification (under review).*

Mrazek H., Weignerova L., Bezouska K. (2011). Aktivní forma α-*N*-acetylgalaktosaminidázy z vláknité houby *Aspergillus niger* a její rekombinantní exprese. *Czech Patent Application*.