Charles University in Prague

Faculty of Science

Department of Biochemistry

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

Mgr. Hynek Mrázek

Ph. D. Thesis

Supervisor: Prof. RNDr. Karel Bezouška, DSc.

Prague 2011

I declare that I have worked on this thesis under the guidance of my supervisor and that all sources of the previous knowledge are properly cited. No part of this work was used and will not be used for obtaining any other academic degree than Ph. D. from Charles University in Prague.

Prague…… ……

Acknowledgements

First I would like to thank to my supervisor, prof. RNDr. Karel Bezouska DSc for his valuable advices and for his patient support during the work on this thesis.

On the second place I would like to thank to every one of my colleagues from the Laboratory of Protein Architecture for their kind help and support. Many thanks belong also to Lenka Weignerova from Laboratory of Biotransformation at the Institute of Microbiology for her help with expression and characterization of α-*N*-acetylgalactosaminidase.

Finally, I would like to thank my wife for her support and tolerance.

The work was supported by grants from: Grant Agency of Charles University (grant no. 19309/2009), from Ministry of Education of Czech Republic (grants MSM_21620808 and 1M0505), and from Grant Agency of Czech Republic (Grant 303/09/0477).

Contents

Abbreviations

Ab	antibody
ADCC	Antibody-Dependent Cellulat Cytotoxicity
α-GA	α -galactosidase
α-NAGA	α-*N*-acetylgalactosaminidase
ATP	Adenosine-5'-TriPhosphate
BLAST	Basic Local Alignment Search Tool
bp	base pair
CD	Cluster of Differentiation
CMP	Cytidine-5'-MonoPhosphate
CTP	Cytidine-5'-TriphosPhate
DCC	DiCyklohexylCarbodiimide
DMF	DiMethylFormamide
DMSO	DiMetylSulfOxid
DNA	DeoxyriboNucleic Acid
DTT	DiThioThreitol
EDTA	EthyleneDiamineTetraacetic Acid
EPO	Erythropoietin
ER	Endoplasmic Reticulum
ESI	ElectroSpray Ionisation
Fc	Fragment crystalizable
FPLC	Fast Protein Liquid Chromatography
FT-ICR	Fourier Transform - Ion Cyclotron Resonance
Fuc	Fucose
G	Glucosidase
GDP	GuanosineDiPhosphate
GH	Glycosidase Hydrolase
GMP	GuanosineMonoPhosphate
GPI	GlykosylphosphatidylInositol
GTP	GuanosineTriPhosphate
Hal	Halogenide
HATU	2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate

HexNAc	*N*-Acetylhexosamine
H/D	Hydrogen/Deuterium (exchange)
HGF	Hepatocyte Growth Factor
HMPA	HexaMethylPhosphorAmide
HPLC	High Performance Liquide Chromatography
HSQC	Heteronuclear Single Quantum Coherence
Ig	Immunoglobulin
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
ITAM	Immunoreceptor Tyrosin-based Activation Motif
ITIM	Immunoreceptor Tyrosin-based Inhibition Motif
KIR	Killer-cell Immunoglobulin-like Receptor
LC	Liquid Chromatography
LRC	Leukocyte Receptor Complex
MHC	Major Histocompatibility Complex
MALDI	Matrix-Assisted Laser Desorption Ionisation
MS	Mass Spectrometry
Ni-NTA	Nickel-NitriloTriacetic Acid
NK	Natural Killer
NKC	Natural Killer gene Complex
NKR-P1	Natural Killer Receptor - Protein 1
NMR	Nuclear Magnetic Resonance
o-NP-α-GalNAc	*ortho*-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside
Ost	Oligosaccharyltransferase
PAMP	Pathogen-Associated Molecular Pattern
PCR	Polymerase Chain Reaction
PRR	Pattern Recognition Receptor
PTK	Protein Tyrosine Kinase
RNA	RiboNucleic Acid
SDS	Sodium Dodecyle Sulphate
SHP	Src Homology Phosphatase
TFA	TriFlouracetic Acid
THF	TetraHydroFuran
TLR	Toll-like Receptor

TMSOTf	TriMethylSilylTriflourmethanesulphonate
Tris	Tris(hydroxymethyl)aminomethane
UDP	UridineDiPhosphate
UMP	UridineMonoPhosphate
UTP	UridineTriPhosphate

Carbohydrates are abbreviated using the IUPAC nomenclature. All monosaccharides are in D-configuration except L-fucose.

1. Introduction

1.1 Glycoproteins

Glycoproteins consist of proteins to which carbohydrate are covalently linked. The predominant sugars found in glycoproteins are Glc, Gal, Man, Fuc, GalNAc, GlcNAc and NeuAc. The distinction between proteoglycans and glycoproteins depends on the level and types of carbohydrate modifications. 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 group of serine, threonine or hydroxylysine. The linkage of carbohydrate to hydroxylysine is generally found only in the collagens. The linkage of carbohydrate to 5-hydroxylysine is either through the single sugar galactose, or the disaccharide glucosylgalactose. In Ser- and Thr-type *O*-linked glycoproteins, the carbohydrate directly attached to the protein is GalNAc. In *N*-linked glycoproteins, it is GlcNAc. 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.

1.1.1 Biosynthesis of glycoproteins

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. Oligosaccharides can be attached to proteins through a variety of linkages, with the two most common linkages being *N*-linked and *O*-linked glycosylation. Monosaccharides are joined together and to protein by the glycosidic bonds. Formation of these bonds requires free energy which is acquired through the conversion of monosaccharide units to nucleotide diphosphate or monophosphate sugars (enzymes using nucleotide sugars as glycosyl donors during biosynthesis of oligosaccharides belong to *Leloir pathway* enzymes; the *non-Leloir pathway enzymes* use glycosyl phosphates) (Gerardy-Schahn et. al. 2001, Goto et. al. 2001) (Figure 1).

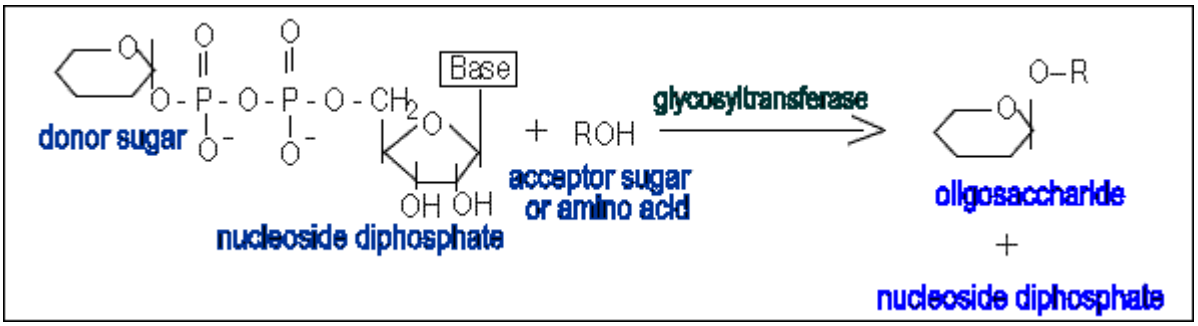


Figure 1. Monosaccharides are joined together with protein by the glycosidic bonds. Formation of these bonds requires free energy input which is acquired through the conversion of monosaccharide units to nucleotide diphosphate or monophosphate sugars. Nucleotide sugars are used by the glycosyltransferases during glycosylation.

Nucleotide sugars are synthesized in the cytosol via following reaction:

sugar-1-phosphate + XTP ---------- sugar-XDP + PPi

This process is catalyzed by nucleoside transferases and pyrophosphatases.

An exception is CMP-NeuAc which is synthesized from the sugar (not sugar phosphate) and CTP:

NeuAc + CTP ---------- CMP-NeuAc + PPi

Nucleotide diphosphate	Monosaccharide units
UDP	Glc, GlcNAc, GalNAc, Gal, NeuAc, Xyl
GDP	Fuc, Man
CMP	NeuAc

Table 1. Nucleotide diphosphate used for individually monosaccharides units.

The ER and Golgi membranes contain transporters for nucleotide sugars. They are antiporters : when sugar nucleotides are imported, nucleotides are exported from the ER or Golgi vesicles (Hirschberg et. al. 1998).

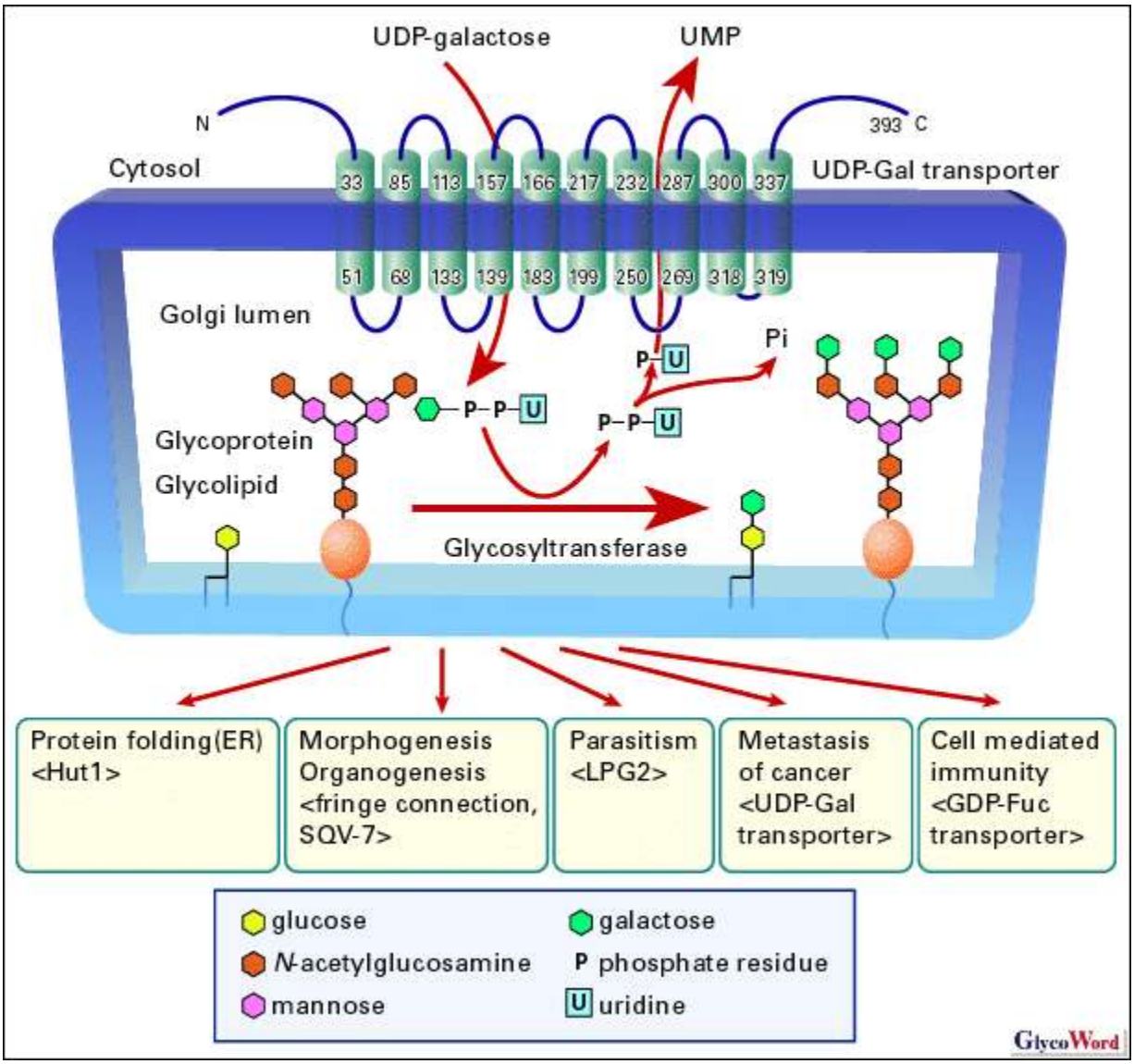


Figure 2. Nucleotide sugar transporters are very hydrophobic proteins from 340 amino acids to 400 amino acids long. They reside in Golgi apparatus and/or ER membrane. They have ten transmembrane helices, and their C- and N-terminal regions are placed in cytosol. They antiport nucleotide sugars pooled in cytosol into lumen of Golgi apparatus or ER for the corresponding nucleoside monophosphates such as UMP for UDP-sugars, GMP for GDP-sugars or CMP for CMP-sugars. The transported nucleotide sugars are utilized as sugar donors by glycosyltransferases for synthesis of sugar chains of glycoproteins, glycolipids and polysaccharides. UDP-glucuronic acid also serves as a substrate for glucuronidation in ER lumen. There are many molecular species of nucleotide sugars, and it is assumed that there is a specific nucleotide sugar transporter for each of them. Hut1 protein is molecular chaperon, and SQV-7 complex participate in morphogenesis. LPG2 protein inhibits the release of midgut proteases and thereby may protect the parasite surface from attack.

1.1.1.1 Glycoproteins with *N*-glycosidically linked oligosaccharides

The core of *N*-glycans is transferred onto proteins *en bloc* and has a clearly defined structure. It is a branched oligosaccharide unit which consists of three Glc, nine Man and two GlcNAc residues (Glc3Man9GlcNAc2). The oligosaccharide is transferred from carrier lipid dolichol pyrophosphate to amide nitrogen of selected asparagine residues of polypetide chains. The acceptor polypeptide is characterized by the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except proline.

The synthesis of the core oligosaccharide starts with the transfer of α-linked *N*-acetylglucosaminphosphate from uridinephosphate to dolicholphosphate giving rise to GlcNAc-PP-Dol. There follows the transfer of the second GlcNAc residue and five Man groups at the cytosolic side of ER, where the sugar nucleotides UDP-GlcNAc and GDP-Man serve as sugar donors. Subsequently, the Man5GlcNAc2-PP-Dol heptasaccharide is transferred to the luminal side of ER (Kornfeld and Kornfeld, 1985). This reaction is catalysed by ATP-independent flippase, which is anchored in the cytoplasmic membrane of ER and it is specifically inhibited by Verapamil. Flippases were identified as transmembrane proteins from the RTF1 family. This group of enzymes was originally discovered in yeast. Family of RTF1 proteins has not ATP-binding part, so they do not belong to the family of the ABC translocators which use ATP for enzymatic activity. Genes for homologous protein of the RTF1 families were also found in the genomes of other eukaryotic organisms (Helenius et. al. 2002). After translocation into ER the residual four Man and three Glc were added according to a defined reaction sequence determined by the specifity of the individual glycosyltransferases, and the chain was elongated to the full-lenght lipid linked oligosaccharide Glc3Man9GlcNAc2-PP-Dol (Kukuruzinska and Robbinson, 1987). This glycan that contains three terminal Glc residues, but not Man9GlcNAc2 is efficiently transferred to protein by oligosaccharyltransferase (Haselbeck and Tanner, 1982). In mammalian cells the importance of the terminal Glc residues is evident from the fact that transfer of Man9GlcNAc2-PP-dolichol is approximatelly 25 times less efficient than of the complete structure. The terminal Glc units are required for recognition of the oligosaccharide structures with enzymes oligosaccharyltransferase. Oligosaccharyltransferase (Ost) is an enzyme associated with translocation complex (Kelleher et. al. 1992). Translocation complex is composed of protein channel, a signal peptidase, BiP, calnexin and probably also from other factors. Oligosaccharyltransferase is composed of several transmembrane subunits (Knauer and Lehle, 1994). For example, Ost from *Saccharomyces cerevisiae* contains eight of them, Ost1p, Ost2p, Wbp1, Swp1, Stt3p, Ost3p/Ost6p, and Ost4p Ost5p, the first five subunits is essential (Silberstein and Gilmore, 1996; Yan and Lennarz, 2005). Ost interacts with nascent

polypeptides and is looking for a specific glycosylation sequence (Asn-X-Ser/Thr). Then it catalyzes the binding of the oligosaccharides with the amide nitrogen of asparagine to form *N*-glycosidic bond (Ecker et. al. 2003). Unlike yeast, where only one type of Ost was found, in mammalian cells two Ost complexes occur, differing in their specificity to lipid carriers and oligosaccharide structure. Complexes containing STT3-A domain prefer complete oligosaccharide structure, while the complex containing STT3-B domain has a higher affinity for oligosaccharide incomplete structures.

Initially, the terminal Glc is removed through the action of glucosidase I (GI), a membrane bound enzyme recognizing α-1,2-linked Glc (Shailubhai et. al. 1987; Shailubhai et. al. 1991). It is a type II membrane glycoprotein of approximately 82 kDa with a short N-terminal cytosolic peptide, transmembrane region, and a large catalytic domain. The remaining two Glc residues are then removed by glucosidase II (GII). GII is a soluble protein recognizing α-1,3-linked Glc, composed of two closely linked glycoprotein chains, α and β, with molecular weight of 107 and 54 kDa. Part of α chain is a catalytic domain belonging to the family of hydrolases. β chain is highly conserved glycoprotein, which contains in its C-terminal part of sequence homologous to mannose-6-P receptor. This enzyme plays an important role in calnexin/calreticulin cycle (Helenius, 1994). The existence of this cycle was demonstrated in 1993 for calnexin and later was discovered that this cycle also includes the soluble ER protein calreticulin. After removal of the Glc residues, the action of α-mannosidases removes several Man residues as the protein progresses to the Golgi (Measaeli et. al. 1999). The action of the various glucosidases and mannosidases leaves *N*-linked glycoproteins containing a common glycan core consisting of three Man and two GlcNAc residues (Molinari and Helenius, 1999). Through the action of a wide range of glycosyltransferases and glycosidases a variety of other sugars are attached to this core as the protein progresses through the Golgi. These latter reactions generate the three major types of *N*-linked glycoproteins. High-mannose type contains all Man outside the core in varying amounts, hybrid type contains various sugars and amino sugars and complex type is similar to the hybrid type, but in addition, contains Gal and NeuAc (Ballou, 1990). In terms of biosynthesis *N*-glycosylation is cotranslation phenomenon, because the connection of oligosaccharide structures to the protein chain in the ER takes place before the end of its biosynthesis.

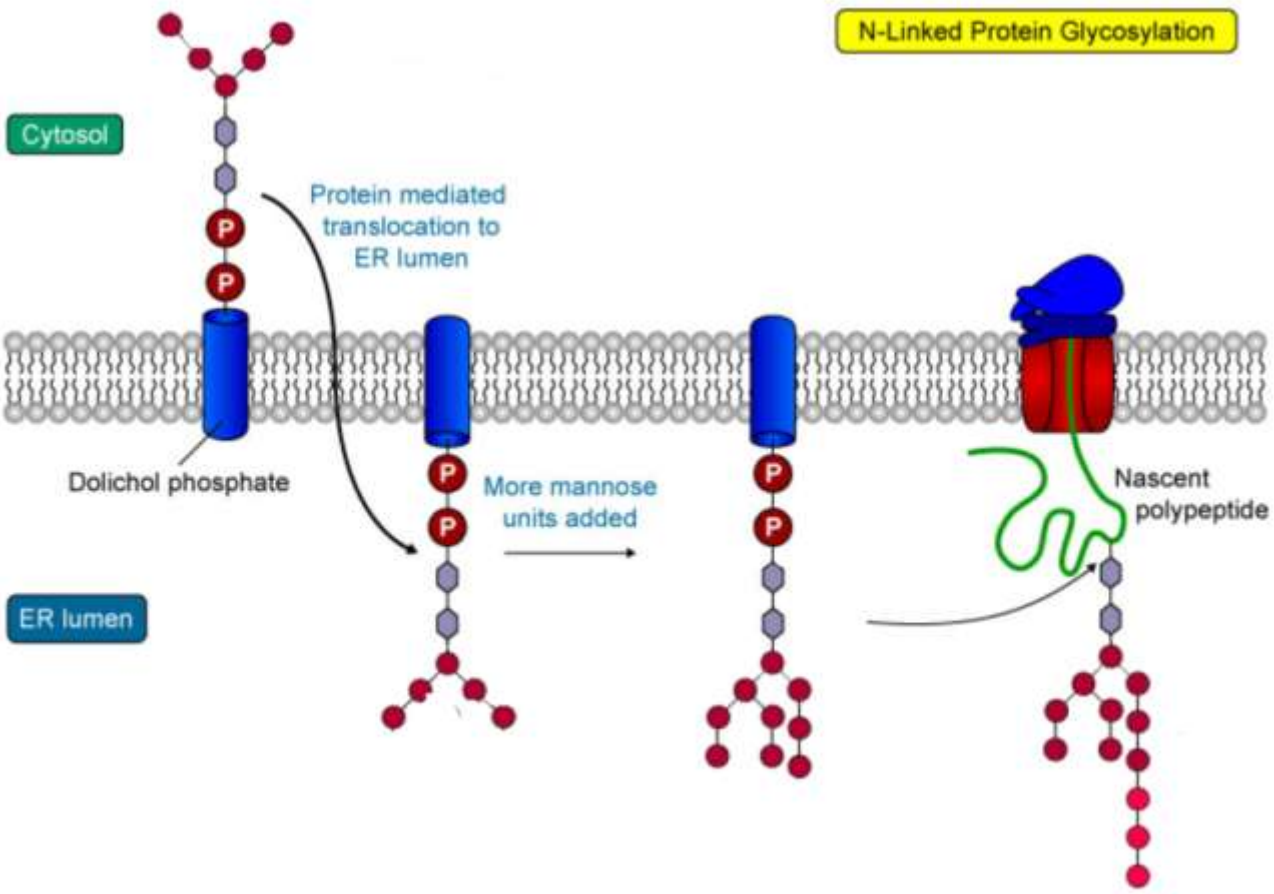


Figure 3. The mechanism for *N*-linked protein glycosylation. The core oligosaccharide is assembled and attached to dolichol phosphate (a lipid carrier) on the outer ER membrane. This lipid-oligosaccharide complex is flipped across the lipid bilayer (so it is now facing the ER lumen) by proteins referred to as flippases. Then, the oligosaccharide is elongated by specific enzymes. Finally, it is transfered to the nascent polypeptide while it is being translated. In *N*-linked glycosylation, the addition of an oligosaccharide to the polypeptide occurs by a covalent attachment of the sugars to an asparagine side chain.

1.1.1.2 *O*-glycosylation

While the *N*-glycosylation is governed by the above rules, no exact rules were found for glycosylation of *O*-type. This process is the stepwise addition of the sugar residues directly onto the polypeptide chain. Carbohydrate structures are bound via the hydroxyl group of serine or threonine. However, glycosylation on tyrosine, hydroxyproline or hydroxylysine has also been described. Eight basic structures representing the core consist of GalNAc and next sugars (Gal, GlcNAc) bound to the initial GalNAc by different types of bonds. From these eight structures a large variety

of *O*-linked oligosaccharides may be synthesized. *O*-glycosylation proceeds late, mostly in the Golgi apparatus. At first, the GalNAc residue is transferred from UDP-GalNAc to the hydroxyl group of Ser or Thr residue. It is catalyzed by *N*-acetylgalactosaminyltransferase. Then the protein moves to the trans-Golgi vesicles where the carbohydrate chain is elongated. The specific glycosyltransferase adds the Gal residue to the GalNAc. The last steps in biosynthesis of typical *O*-glycans are the additions of two NeuAc residues in the trans-Golgi. *O*-glycosylation is most prevalent in regions rich in serine or threonine. Functionaly, the *O*-glycosylation applies especially in protective function of proteins, as well as the formation of blood groups and participates in protecting the epithelium.

1.1.2 Variations in the carbohydrate moieties between mammals and yeasts

Yeasts are capable of performing many eukaryotic posttranslational modifications, including *N*-glycosylation. However, the *N*-linked glycans from yeast differ significantly from those of mammalian and human cells. Mammalian cells and yeast share the initial biosynthetic pathway for the synthesis of *N*-glycans. The process is initiated by the transfer of GlcNAc from UDP-GlcNAc onto dolichol phosphate on the cytoplasmic face of the ER. Subsequent glycosyltransferase reactions that catalyze the addition of GlcNAc and Man mature the structure to Man5GlcNAc2-P-dolichol, which is then translocated to the luminal face of the ER membrane by a flipase enzyme. Once inside the ER the structure is further extended to Glc3Man9GlcNAc2-P dolichol, at which point the glycan is transferred to the N-X-S/T motif of the target peptide by the oligosaccharyltransferase complex. Following transfer to the nascent polypeptide, three Glc residues and one Man residue are removed to produce Man8GlcNAc2, at this stage the glycopeptide is transported to the Golgi apparatus. Although this glycan structure arriving into the Golgi is identical in yeast and mammals, the pathways diverge significantly as the protein proceeds through the rest of the respective secretory pathways (Ballou, 1990). In mammalian cells the Man8GlcNAc2 glycan is trimmed to Man5GlcNAc2 by the action of several α-1,2-mannosidases, at which stage GlcNAc is added by *N*-acetylglucosaminyltransferase I (GnTI). Subsequent trimming and extension reactions lead to the production of complex type structures. In contrast to mammals, yeast do not further trim the Man8GlcNAc2 glycan in the Golgi, but rather extend the existing high mannose structure with additional Man residues, to produce hypermannosylated glycan chains (Gemmil and Trimble, 1999).

1.1.3 Biological significance of glycosylation

Protein folding

The single most important role of glycosylation is the promotion of proper folding of newly sythesized polypeptides in the ER (Helenius, 1994). It is well known that glycans can have a direct effect on the protein folding process (Imperiali and O'Connor, 1999). The presence of a large polar carbohydrate unit is also likely to affect the folding process locally by orienting the polypeptide toward the surface of protein domains. *In vitro* refolding experiments showed that the glycosylation can have positive effect on the folding process (Kern et. al. 1993; Imperiali and O'Connor, 1999). The most important indirect effect of glycans on folding involves a unique chaperone system called calnexin/calreticulin cycle, found in ER of eukaryotes. The existence of this cycle was demonstrated in 1993 for calnexin, and later discovered that this cycle includes also soluble ER protein calreticulin. Every single step has been extensively studied, and is described in detail. The process begins with connecting the core glycan to the emerging polypeptide chain. The first terminal Glc is rapidly removed by glucosidase I. After removing of the first Glc, there follows removing of the second Glc catalyzed by glucosidase II. Monoglucosylated core binds specifically to calnexin or calreticullin. In the formation of such a complex calnexin and calreticulin serve as molecular chaperones, protecting protein against aggregation and export of incorrectly folded peptides from the ER. Further, these chaperones enable interaction of ERp57 complex glycoproteins with thiol-disulfide oxidoreductase. This enzyme catalyzes the formation of correct disulfide bonds during protein biosynthesis. The release of polypeptide chains from the calnexin / calreticulin complex occurs after removing of the last Glc residue by glucosidase II. After interaction of a new polypeptide with UDP-Glc: glycoprotein glucosyltransferase, which serves as a sensor of correct protein conformation, correctly folded proteins can leave the ER. In the opposite case incorrectly folded proteins are degraded. Degradation process is called ER associated degradation (ERAD), and it is very important, since it prevents accumulation of incorrectly folded proteins in the ER.

Calnexin belongs to a group of type I of transmembrane proteins. Calreticulin is soluble ER protein. Both belong to the family of lectins isolated from legumes. Both are monomeric calcium-binding proteins and have ER localization sequence. NMR structure of the calreticulin and X-ray crystallography revealed the unusual architecture of this lectin. It consists of two separate domains, globular β-sandwich domain that is homologous with lectins isolated from the legumes, followed by proline rich domain. Calnexin differs by trasmembrane part and cytoplasmic domain of a size of 96 amino acids. This domain is phosphorylated and responsible for interaction with the ribosome.

Calnexin and calreticulin also have an identical carbohydrate-binding specificity. For proper interaction of these lectins one Glc and three Man residues present in the oligosaccharide nucleus are important. Glycoprotein glucosyltransferase is the sensor capture of the correct protein conformation. It is one of the most interesting enzymes of the calnexin/calreticuline cycle. It is a large soluble protein with a C-terminal sequence which ensures retention in the ER. Catalytic glucosyltransferase segment contains 300 amino acids with high homology to members of the glycosyltransferase family 8. This segment is located in the C-terminal part. The N-terminal part consists of approximately 1200 amino acids. This part of the GT is probably involved in the recognition of a suitable substrate. C-and N-terminal domains are structurally and functionally related. How accurately has been the composiiton of glucosyltransferase protein determined? To this question there is not currently a clear answer. Two ways how could the glucosyltransferase check the correct conformation of the protein have been suggested. The first way would be through hydrophobic peptides, which are exposed on the surface of the protein. The second way would be through the dynamics of the protein surface. Glucosyltransferase uses a substrate of UDP-glucose, transported to the ER from cytosol. Furthermore, the correct operation of glucosyltransferase requires the presence of Ca^{2+}, Mg^{2+} or Mn^{2+}.

Effect on protein stability and solubility

An example of a glycoprotein that is medically important is erythropoetin also known as EPO. This glycoprotein has improved the treatment for anemia particularly induced by cancer chemotherapy. It is secreted by kidney and stimulates the production of red blood cells. EPO is made of 165 amino acid (Dordal et. al. 1985). It is *N*-glycosylated at the asparagine residue and *O*-glycosylated at a serine residue. It is composed of 40% of carbohydrates by weight. The glycosylation enhances stability of the protein in the blood as compared to the unglycosylated protein, which only carries about 10% of the bioactivity (Kubota et. al. 1990). This low bioactivity is due to the rapid removal from the blood by the kidney. Although recombinant human EPO has aided the treatment of anemia it has also been misused by athletes to increase their red blood cell count and their oxygen carrying capacity. However modern drug testing can usually distinguish between this and natural EPO.

Protection of protein

Carbohydrates play an important role in the protection of glycoproteins from proteolytic attack. Experiments showed that heavily glycosylated porcine pancreatic ribonuclease is very

insensitive to the proteolytic attacks. Subtilisin undergoes autoproteolysis when its carbohydrates moieties are removed while the native glycosylated enzyme is stable (Wang and Hirs, 1977). EPO described above can serve as the second example of such a protection. It was reported that by removal of terminal NeuAc from EPO leads to an increase in its susceptibility to proteolysis by trypsin (Goldwasser et. al. 1974).

Influence on biological activity

Generally, only completely glycosylated proteins have full biological activity as shown for HGF (hepatocyte growth factor). The studies demonstrated that nonglycosylated protein retains only mitogenic activity, and loses morphogenetic activities. The second example is the blood clotting Factor IX which completely loses its activity after removal of terminal sialic acids (Yan et. al. 1990). The glycosylated tissue plasminogen activator and glycosylated EPO have much higher *in vivo* activities than their unglycosylated forms (Berg et. al. 1993, Higuchi et. al. 1992, Delorme et. al. 1992).

Influence on in vivo clearence rate

Oligosaccharides structures on proteins play also dominant role in *in vivo* clearence rate of glycoproteins. Two clearence pathways were found. In the first clearence pathway the glycoproteins without NeuAc with exposed terminal Gal and GlcNAC (desialylated complex *N*-glycans) bind to the sialoglycoprotein receptor which occurs on the surface of hepatocytes and consequently are eliminated from circulatory system (Weis and Ashwell, 1989). The second clearence pathway is represented by mannose receptors on the surface of endothelial cells in liver and kidney and on resident macrophages which bind glycoproteins with high mannose structures (Ezekowitz and Stahl, 1988).

1.1.4 Glycosidases

The intracelular degradation of glycoproteins and glycosaminoglycans occurs predominantly in lysosomes. This degradation involves the concerted action of about 20 hydrolytic enzymes (glycosidases). The amino acids and cabohydrates result from the degradative process are released from lysosome, and used again in biosynthesis.

Glycoside hydrolases are classified by international enzyme nomenclature into group 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 nonreducing) end or in the middle, respectively, of an oligo/polysaccharide chain.

Glycoside hydrolases are found in essentially all domains of life. In prokaryotes they are found both as intracellular and extracellular enzymes that are largely involved in nutrient acquisition. One of the important occurrences of glycoside hydrolases in bacteria is the enzyme β-galactosidase (LacZ), which is involved in regulation of expression of the *lac* operon in *Escherichia coli.* In higher organisms glycoside hydrolases are found within the ER and Golgi apparatus where they are involved in processing of *N*-linked glycoproteins, and in the lysozyme as enzymes involved in the degradation of carbohydrate structures. Deficiency in specific lysozomal glycoside hydrolases can lead to a range of lysosomal storage disorders that result in developmental problems or even death. Glycoside hydrolases are found in the intestinal tracts and saliva where they degrade complex carbohydrates such as lactose, starch and sucrose. In the gut they are found as glycosylphosphatidyl anchored enzymes on endothelial cells. The enzyme lactase is required for degradation of the milk sugar lactose and is present at high levels in infants, but in most populations will decrease after weaning or during infancy, potentially leading to lactose intolerance in adulthood. The enzyme *O*-GlcNAcase is involved in removal of GalNAc groups from serine and threonine residues in the cytoplasm and nucleus of the cell.

They are very common enzymes with roles in nature including degradation of biomass such as cellulose and hemicellulose, in anti-bacterial defense strategies (e.g., lysozyme), in pathogenesis mechanisms (e.g., viral neuraminidase), and in normal cellular function (e.g., trimming manosidase involved in *N*-linked glycoprotein biosynthesis). Together with glycosyltransferases described above, glycosidases form the major catalytic machinery for the synthesis and breakage of glycosidic bonds. A classification system for glycosyl hydrolases, based on sequence similarity, has led to the definition of more than 100 different families (Henrissat and Davies, 1995). This classification is available on the CAZy (CArbohydrate-Active EnZymes) web site. The database provides a series of regularly updated sequence, and allows a reliable prediction of the catalytic mechanism (retaining/inverting), active site residues and possible substrates. The online database is supported by CAZy system, an online encyclopedia of carbohydrate active enzymes (Henrissat and Coutinho, 1999). Based on three dimensional structural similarities, the sequence-based families have been classified into 'clans' of related structure. Recent progress in glycosidase sequence analysis and 3D structure comparison has allowed the proposal of an extended hierarchical classification of the glycoside hydrolases (Naumoff, 2006).

Reaction mechanism can be of two types - inverting and retaining. Inverting enzymes utilize two enzymic residues, typically carboxylate residues, acting as acid and base, respectively, as shown below for a β-glucosidase.

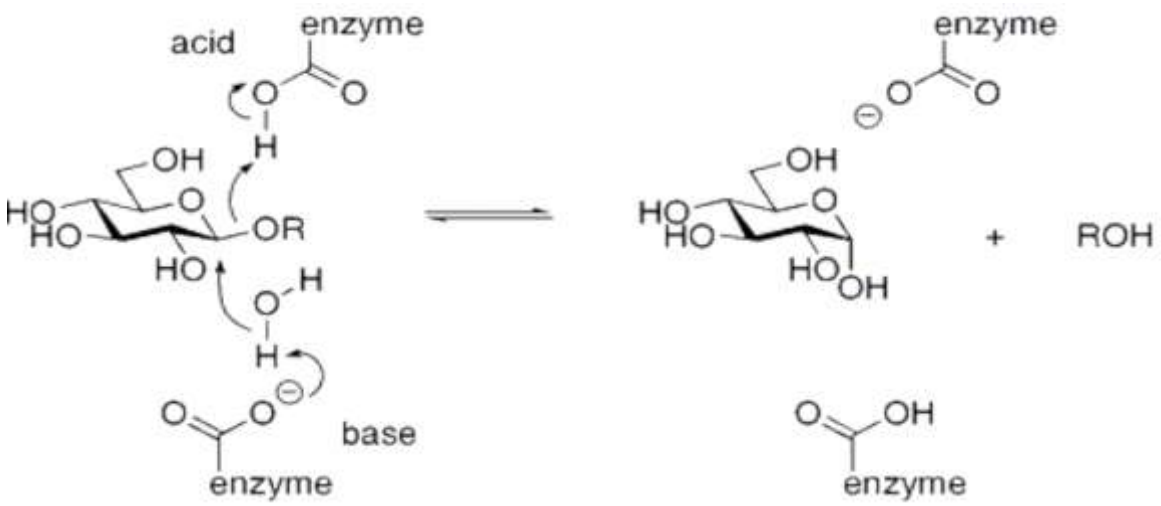


Figure 4. Inverting mechanism of the glycosidase reaction. Inverting enzymes utilize two enzymatic residues, typically carboxylate residues, that act as acid and base respectively.

Retaining glycosidases operate through a two-step mechanism, with each step resulting in inversion for a net retention of stereochemistry. Again, two residues are involved, which are usually enzyme-borne carboxylates. One acts as a nucleophile and the other as an acid/base. In the first step the nucleophile attacks the anomeric centre, resulting in the formation of a glycosyl enzyme intermediate, with acidic assistance provided by the acidic carboxylate. In the second step the now deprotoned acidic carboxylate acts as a base and assists a nucleophilic water to hydrolyze the glycosyl enzyme intermediate, giving the hydrolyzed product. The mechanism is illustrated below for hen egg white lysozyme (Vocadlo et. al. 2001).

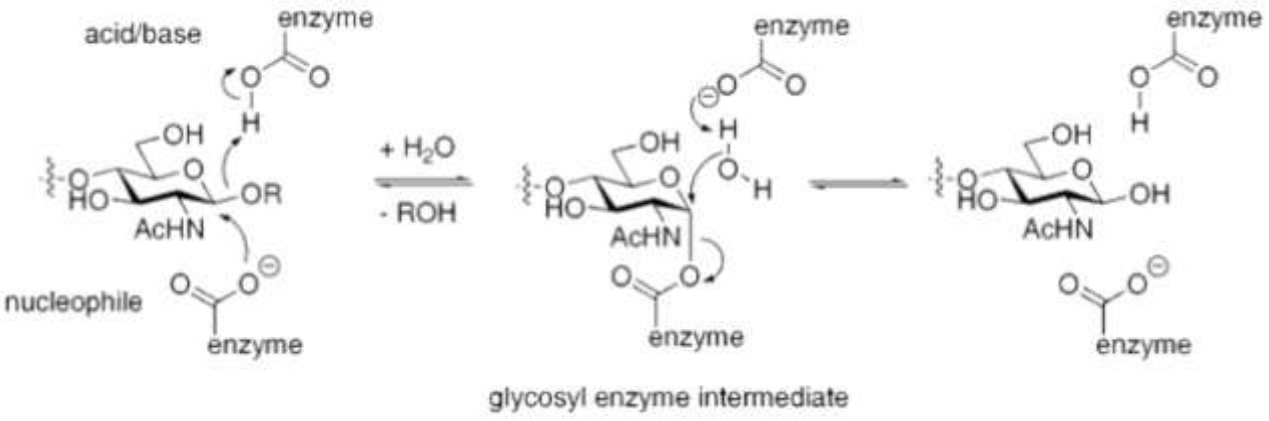


Figure 5. Retaining glycosidases operate through a two-step mechanism, with each step resulting in inversion, for a net retention of stereochemistry.

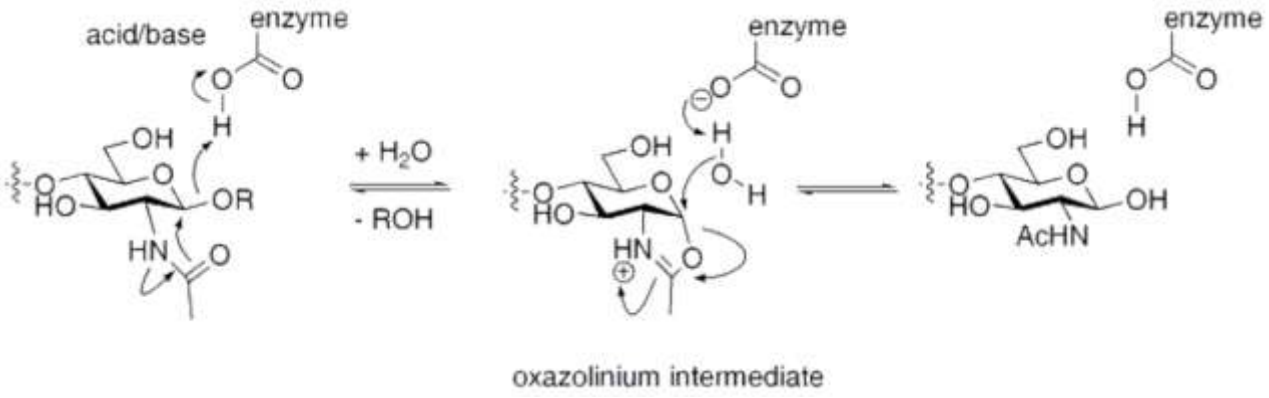


Figure 6. An alternative mechanism for hydrolysis with retention of stereochemistry.

An alternative mechanism for hydrolysis with retention of stereochemistry can occur that proceeds through a nucleophilic residue that is bound to the substrate, rather than being attached to the enzyme. Such mechanisms are common for certain *N*-acetylhexosaminidases, which have an acetamido group capable of neighboring group participation to form an intermediate oxazoline or oxazolinium ion. Again, the mechanism proceeds in two steps through individual inversions to lead to a net retention of configuration, described above.

Many compounds are known that can act to inhibit the action of a glycoside hydrolase.

Nitrogen-containing, 'sugar-shaped' heterocycles have been found in nature including deoxynojirimycin, swainstonine and castospermine From these natural templates many other inhibitors have been developed, including isofagomine and deoxygalactonomin and various unsaturated compounds such as 1,5-hydroximolactone. Inhibitors that are used clinically include antidiabetic drugs or antiviral drugs like oseltamivir and zanamivir. Moreover, some proteins have been found to act as glycoside hydrolase inhibitors.

1.1.5 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 IUBMB (International Union of Biochemistry and Molecular Biology) these enzymes belong to Hydrolase-Glycosidase-Glycosidase hydrolyzing *O*- and *S*- glycosidic linkage group. According to CAZY system prokaryotic α-NAGA belong to enzyme family 36 (Clan GH-D) and eukaryotic α-NAGA to enzyme family 27 in the same clan like prokaryotic enzyme. In phylogenetic tree of enzyme family 27 and 36, α-NAGA are divided to three groups. The first group includes α-NAGA and α-galactosidases (α-GA) of vertebrates, second group includes these enzymes from yeast, and in the last group there are α-NAGA and α-GA from plants and fungi. The phylogenetic mapping showed that the evolution of genes for α-NAGA from vertebrates is distinct from evolution of genes for α-NAGA from fungi (evolutionary paralelism). α-NAGA were evolved from fungal α-GA occurring in fungi growing on solid substrates as a way for metabolic utilization of compounds containing terminal α-linked GalNAc.

α-NAGA can occur in organisms in various isoforms, which can be different in their catalytic activities (Weignerova et. al. 2008). Unlike monomeric forms, dimeric and tetrameric forms of these enzymes are enzymatically activite. The mutual equilibrium of individual forms is influnced by pH value. If the pH value is very low, enzyme occur in dimeric or tetrameric active forms. Alkaline environment preferably creates monomeric (inactive) form of the enzyme.

Enzymatic hydrolysis of glycosidic bond uses the retaining mechanism, when the anomeric of bond for substrate and product is unchanged. The carbon C1 of GalNAc involved in glycosidic linkage is subjected to two nucleophilic attacks. First, atom of oxygen Asp140 attacks electrophilic carbon C1 of the carbohydrate skeleton. This interaction leads to creating of covalent enzyme - substrate complex. Then follows the attack of deprotonated molecule of water, this reaction causes a release of the product and regeneration of an enzyme.

1.1.6 Schindler disease

One of the most studies disorders related to metabolism of glycoproteins and malfunction of glycosidases is Schindler disease. This disease is also known as Kanzaki disease. This disease was first described in 1988 by Detlev Schindler. Schindler/Kanzaki disease is an inherited metabolic disorder belonging to the lysosomal storage disorders. A deficient enzyme, α-NAGA, leads to an abnormal accumulation of certain compounds (glycosphingolipids) in the cells of the body. This abnormal accumulation causes damage to the cells that may get worse over time (Wang et. al. 1990; Cantz and Ulrich-Bott, 1990). There are two forms of Schindler disease. The classical form of the disorder, known as Type I Schindler disease, has an infantile onset (Sakuraba et. al. 2004). Affected individuals appear to develop normally until approximately 1 year of age, when they begin to lose previously acquired skills that require the coordination of physical and mental activities (developmental regression). Additional neurological and neuromuscular symptoms may become apparent, including diminished muscle tone (hypotonia) and weakness; involuntary, rapid eye movements (nystagmus); visual impairment; and episodes of uncontrolled electrical activity in the brain (seizures). With continuing disease progression, affected children typically develop restricted movements of certain muscles due to progressively increased muscle rigidity, severe mental retardation, hearing and visual impairment, and a lack of response to stimuli in the environment. Type II Schindler disease, which is also known as Kanzaki disease, is the adult-onset form of the disorder (Umehara et. al. 2004). Associated symptoms may not become apparent until the second or third decade of life. In this milder form of the disease, symptoms may include the development of clusters of wart-like discolorations on the skin (angiokeratomas); permanent widening of groups of blood vessels (telangiectasia), causing redness of the skin in affected areas; relative coarsening of facial features; and mild intellectual impairment. The progressive neurological degeneration characteristically seen in the infantile form of the disease has not occurred in association with type II Schindler disease. Both forms of Schindler disease are inherited as autosomal recessive traits. According to investigators, different changes (mutations) of the same gene are responsible for the infantile- and adult-onset forms of the disease. The gene has been mapped to the long arm (q) of chromosome 22 (22q11). Diagnosis of Schindler disease is based on the symptoms the individual has, as well as the age the symptoms began. A urine test, blood test, or skin sample (biopsy) may help confirm the diagnosis. In Schindler disease, the blood or skin sample will show decreased activity of α-NAGA. Treatment for Schindler disease focuses on its symptoms, since there is as yet no cure for the disease. Specialists such as a neurologists (for seizures or nervous system

complications), eye doctor (ophtalmologist), and geneticist will be involved in the individual's care. Physical and occupational therapy can help the individual with Type I disease maintain muscle movement and relieve discomfort.

1.2 Synthesis of glycopeptides and glycoproteins

The need for carbohydrates in biological events may be explained through the diverse range of chemical properties that they can transfer onto other compounds. Glycosylation is the most complex protein modification event and nature does not provide a coding template of the type that is available for nucleic acids and polypeptides. Carbohydrates are unparalleled in the number of structures they can adopt, and as a consequence, nature appears to exploit this structural diversity to convey information at a molecular level. In terms of oligomerization, proteins and nucleic acids are effectively linear in structure (amide bonds in proteins and 3′-5′ linkage of phosphodiesters within DNA), where limited basis sets (4 for DNA, 20 for amino acids) give rise to limited variations. In addition to inherent configurational variation (gluco, manno, etc.), additional variety caused by ring size, branching, anomeric configuration, and modification (e.g., acylation, sulfation, and phosphorylation) gives carbohydrates strong potential for diversity. This inherent structural diversity parallels a wide range of functions within nature (Dwek, 1996; Varki, 1993), ranging from a source of energy and metabolic intermediates to the structural components of plants (cellulose), animals (chitin), and nucleic acids (DNA, RNA). However, greater variation in oligosaccharide structure is exploited by nature through the combination of carbohydrates with proteins (glycopeptide/glycoprotein), the products of which have caused an explosion of interest within the scientific community. This structural diversity generated by carbohydrates is assumed to be ruled by a *glycocode*, a term representing the potential level of complex information that carbohydrate structures are able to convey (Davis, 2002; Gabius et. al. 2004) This vast number of potential variations represent a technological barrier and means that oligosaccharide portions of glycoproteins can not be made simply on an iterative basis since there are far too many possible synthetic targets. It is therefore crucial that the design of new glycoproteins is guided by identification of the associated functions and activities of existing natural and resulting synthetic structures.

1.2.1 Potential aplications

Recent developments in glycopeptide/glycoprotein synthesis have been further driven by the

practical applications of the synthetic products (Wong, 1995). For example, Verez-Bencomo developed a glycoconjugate vaccine composed of a fully synthetic capsular polysaccharide antigen of *Haemophilus influenzae* type b (Hib) (Verez-Bencomo et. al. 2004) highlighting the fact that access to synthetic complex carbohydrate-based vaccines is feasible. Initial results from clinical trials demonstrated long-term protective antibody titers as compared to the already licensed commercial products. This Cuban work has set the international benchmark for synthetic vaccines, and highlighted the need for further development of similar approaches to other human pathogens and, perhaps, greater innovation and courage in other countries.

The second example can be modulation of the serum half-life with glycans, and an attempt to modulate the pharmacokinetics of synthetic insulin in the way best suited for diabetes patients. Nishimura and colleagues combined mutagenesis with enzymatic synthesis in the production of a glycosylated insulin (Sato et. al. 2004). Standard insulin is rapidly degraded by the liver (within a few hours of administration), thus requiring frequent booster shots. Methods to increase in vivo activity of insulin have been investigated (Markussen et. al. 1987), but these are limited by intricate administration regimes resulting in uncontrolled fluctuation in glucose blood levels caused by decreased water solubility (Markussen et. al. 1988). Nishimura tackled the water solubility and degradation problems by introducing sialic acid, a sugar that is known to increase glycoprotein half-life *in vivo* (Egrie et. al. 2003). A protocol used required the installation of accessible Gln residues through mutagenesis into the B chain of insulin and subsequent transglutaminase-catalyzed transamidiation with a lactosyl amine. The resulting glycoprotein was further modified through enzymatic sialylation with a Sia-α-2,6-transferase to make Siaα2,6-Lac. The modified insulin showed a long-lasting *in vivo* activity compared to the unmodified insulin. Interestingly, the same synthetic methods were also used in the creation of a dendrimeric display of sialic acid, leading to an insulin “glycodendriprotein”. Although the binding affinity of these modified insulins to receptor decreased as dendrimer size increased, an overall *in vivo* activity increase was observed due to the enhanced half-life caused by the higher presence of sialic acid.

Fatalities resulting from malaria are caused by an inflammatory surge initiated by malarial toxin released from the parasite *Plasmodium falciparum*; glycosylphosphatidylinositols (GPI) are thought to be the primary toxins underlying this pathology. In an attempt to produce a vaccine against malaria, Seeberger and co-workers chemically synthesized the GPI oligosaccharide and conjugated it to carrier proteins (Schofield et. al. 2002). Anti-GPI antibodies were obtained from immunized mice and shown to neutralize the pro-inflammatory activity of *P. falciparum in vitro*. More impressively, it was demonstrated that deaths from malarial parasites in animal models were

greatly decreased, thereby establishing the idea that GPI conjugates could be used for antimalarial vaccine design. The rapid and high-yielding assembly of the GPI motif was possible through automated solid-phase glycosylation chemistry (Hewitt et. al. 2002). Such chemistry is useful since it could be modified to include alternative building blocks for structure – activity studies.

1.2.2 Chemical synthesis of glycopeptides

Chemical synthesis of *N*- and *O*-glycopeptides can be done in three different ways: synthesis of preformed glycosyl amino acids building block, synthesis in solution and synthesis on solid supports.

1.2.2.1 *N*-Glycosides

The most useful techniques for synthesis of glycopeptides is synthesis by preformed glycosyl amino acids bulding blocks. It is commonly used for the synthesis of *N*-glycosides exploiting the formation of a peptidic bond between a glycosylamine **1** and a protected aspartic acid derivative **2**. Glycosylazides are usually employed as glycosylamine precursors, which can be obtained by treatment of glycosyl halides or 2-iodoglycosylsulfonamides with azide salts (Szilagyi and Györgydeak, 1985).

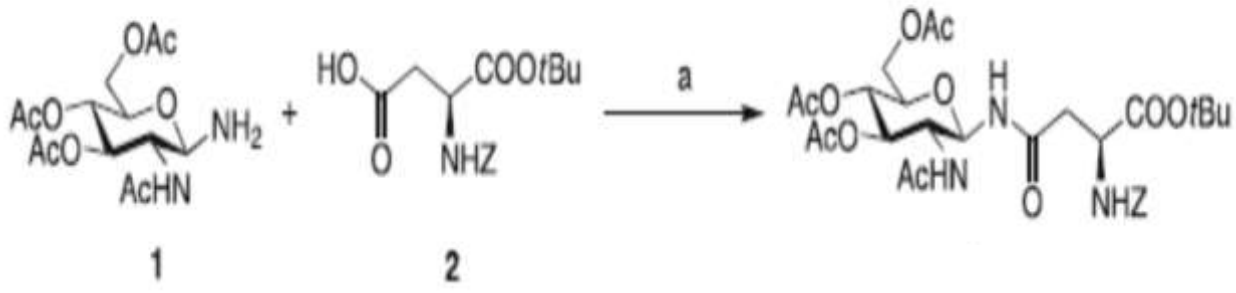


Figure 7. Synthesis of *N*-glycosides. Chemical conditions of reaction (a) include DCC, HOBt, THF, 79.5%.

In a strategy reported by Kiessling et al., glycosylazides are used directly for a stereoselective *N*-glycosylation by Staudinger ligation. Herein the glycosylazide **3** was reacted with phosphinothioester **4** furnishing glycosylasparagine **5** in a moderate yield (Figure 8) (He et. al. 2004). Takahashi and coworker described a synthesis that avoids the need to prepare glycosylazides. The *β*-glycosyl trifluoroacetoimidate **6** was shown to react with asparagine **7** to deliver *β*-

configurated *N*-glycosyl asparagine **8** (Figure 8) (Tanaka et. al. 2005). Glycosylamines can be prepared directly from unprotected sugars by Kochetkov's procedure, which involves treatment with ammonium bicarbonate. An improved procedure based upon ammonium carbamate was introduced by Likhosherstov. The carbamic acid salts **10** are stable upon storage and are converted into the free amines **11** by base-treatment or by applying vacuum (Figure 8). Flitsch et al. have shown the acceleration of the Kochetkov procedure by irradiation with microwaves. The conversions of the sugars into the corresponding amines were finished within 90 min with yields ranging from 35 to 87%. Complex *N*-linked glycosyl amino acids can also be obtained by releasing *N*-linked carbohydrates from natural.

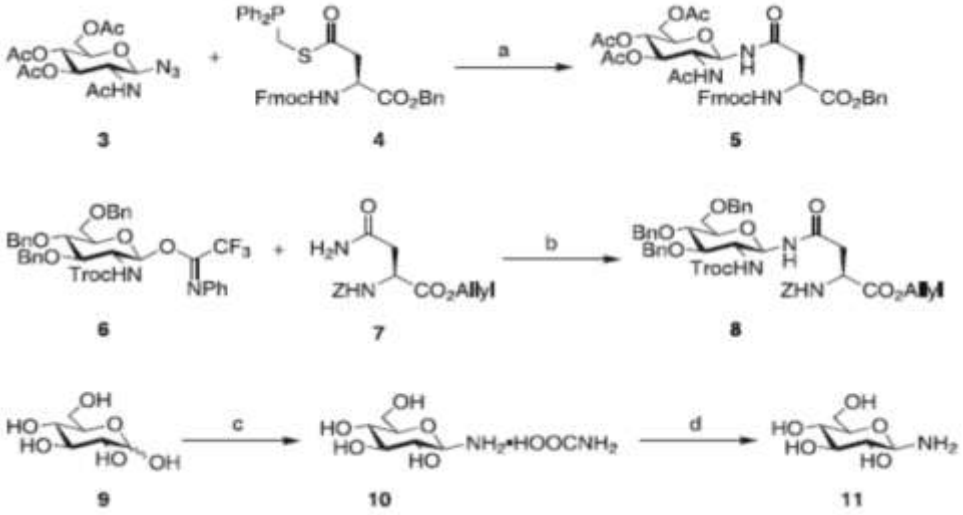


Figure 8. Alternative methods for the synthesis of *N*-glycosidic bonds. Conditions for individual reactions are as follows: (a) DMF, 45%; (b) TMSOTf, MeNO2, 0 ◦C to r.t., 68%; (c) NH4COONH2 in sat. NH3 solution (d) base or high vacuum.

The second technique is synthesis of glycopeptides in solutions. This techniques is analogy to biosynthetic pathway, glycosyltransferases can be used to step by step assemble carbohydrate structures by using sugar nucleotides as a donor. This approach was predominantly utilized in the synthesis of *O*-linked glycopeptides like fragments of the P-selectin glycoprotein ligand-1 PSGL-1 (Holm et. al. 2005). There are glycosidases that catalyze the release of *N*-glycans by cleaving the *β*-glycosidic linkage between two GlcNAc residues adjacent to asparagines in *N*-glycans. In a trans-glycosylation approach, these enzymes can transfer saccharide moieties from glycosyl asparagine

donors to *N*-linked GlcNAc-peptides. Wang applied the endo-*β*-*N*-acetylglucosaminidases from *Arthrobacter protophormiae* (Endo-A) and from *Mucor hiemalis* (Endo-M) for the synthesis of peptides displaying high-mannose type *N*-glycans (Singh et. al. 2003). Glycopeptide products are obtained in relatively modest yields and, though elegant, access to the complex *N*-glycopeptide donors is required. Significant improvements of transglycosylation yields were brought about by Shoda's proposal of using 1,2-oxazolines as substrates (Fujita et. al. 2001). During the solid-phase synthesis of *N*-glycopeptides it is practical to employ a block glycosylation approach in which the full-length carbohydrate is coupled to the aspartic acid side chain (Wang et. al. 2001). For example, in Danishefsky's synthesis of the HIV envelope glycoprotein gp120-fragment the 3-OH of the Man residue of precursor was subjected to α-mannosylation with building block followed by debenzoylation and extension with thiolactoside donor. Regioselective opening of the benzylidene exposed the primary hydroxyl group of hexasaccharide. The glycosylation with the trimannosyl building block completed the assembly of the nonasaccharide. In preparing the stage for the connection of the sugar unit with the peptide, the protecting groups were removed followed by conversion of to the glycosylamine by using Kochetkov's procedure. The aspartic acid side chain of *N, S*-protected peptide was activated by HATU in DMSO to accomplish the coupling with glycosylamine. Removal of the remaining protecting groups afforded glycopeptide (Mandal et. al. 2004). A similar strategy was employed in the synthesis of the high-mannose-type analogue of glycoconjugate, which also spanned a fragment of HIV virus envelope-protein gp120. Both highly complex glycopeptides are putative HIV vaccines (Dutkin et. al. 2004).

The last techniques using in synthesis of glycans is synthesis on solid support. This synthesis allows the automation of the highly repetitive process of building block coupling - if desired in parallel and combinatorial formats. Further advantages of solid-phase synthesis are the speed and ease of synthesis and the possibility to drive reactions to completion by using a large excess of building blocks and reagents. To cope with the acid sensitivity of many carbohydrates, the solid-phase synthesis of glycopeptides is most commonly performed by coupling *N*-Fmoc-protected amino acid building blocks, which allow the application of relatively mild conditions for the cleavage of temporary and permanent protecting groups. With few alterations concerning the removal of carbohydrate protecting groups, the standard protocols of Fmoc-based solid-phases peptide synthesis can be applied.

In solid-phase synthesis, the choice of the linker that connects the growing glycopeptide with the solid support is of critical importance. Acid-labile linkers such as the one used by Wang (Wang, 1973), the HMPA (Shepard and Williams, 1982), the linker used by Rink (Rink, 1987), or

the PAL-linkers (Albericio et. al. 1990) have frequently been used to release unprotected glycopeptide acids and glycopeptide esters. The synthesis of protected glycopeptide fragments has been accomplished by using very acid-labile linkers such as SASRIN (Mergler et. al. 1988), HMPB (Floersheimer and Riniker, 1991), Sieber's linker of quite complex structure (Sieber, 1987) or trityl linkers (Barlos et. al. 1989) or the Pd(0)- labile allylic HYCRON anchor (Seitz and Kunz, 1995).

1.2.2.2 *O*-glycosides

The preparation of the basic *O*-linked glycosyl amino acids (*α*GalNAc)Ser/Thr **3** is typically accomplished by treatment of glycosyl halides such as **1** with the free hydroxyl groups of otherwise protected amino acids **2** in the presence of silver salts (Paulsen et. al. 1995). High *α*-selectivities can be achieved when benzyl-protected glycosyl fluorides are employed as glycosyl acceptors (Shao and Guo, 2005). The alternative reaction of acetimidates **1** with the protected amino acids **2** in the presence of Lewis acids also assures access to the desired compounds. For the synthesis of complex *O*-glycosides, a stepwise glycosylation strategy has emerged as reliable. Thus, the carbohydrate part is assembled by using a glycosyl amino acid as an acceptor, to which further glycan residues are appended. Versatile syntheses for the preparation of the more complex 2,6-sialyl-T, 2,3-sialyl-T and sialyl-T*N*-antigens were developed by Kunz.

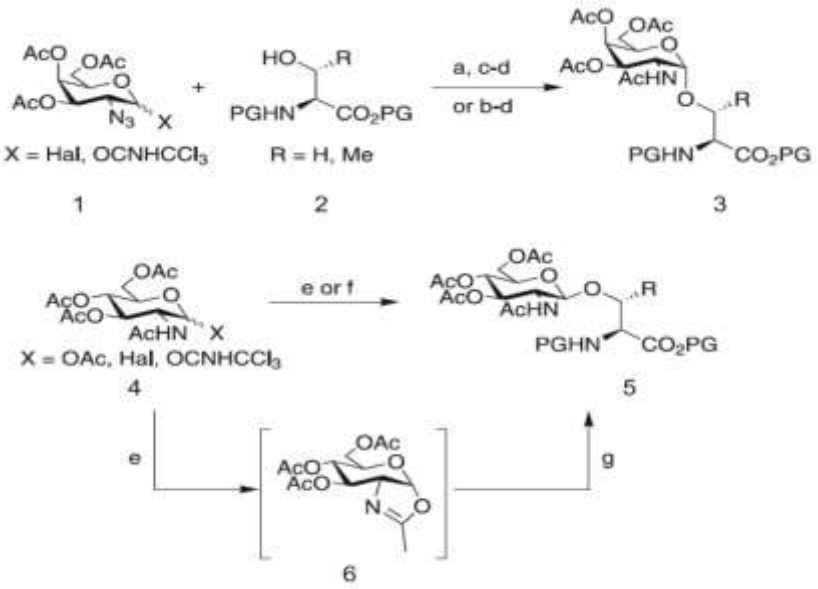


Figure 9. Synthesis of *O*-linked glycosylamino acids. Chemical conditions for individual reactions are as follows: (a) Ag2CO3/AgClO4 (X = Hal); (b) TMSOTf (X = OCNHCCl3); (c) NaBH4, NiCl2; (d)Ac2O, pyridine; (e) BF3OEt2, **2** (X = OAc); (f) AgOTf, **2** (X = Hal, OCNHCCl3); (g) CuCl2.

In principle, the synthesis of *O*-linked glycopeptides in solutions can be performed by attaching the glycan as block to the amino acid/peptide aglycon or by stepwise assembly of the peptide part by using preformed glycosyl amino acids. The modest yields of block glycosylations and their relatively poor *α/β*-selectivities call for the application of the latter approach. The step-by-step condensation of presynthesized building blocks in solution has found widespread application. Danishefsky reported the synthesisof various glycopeptides bearing complex *O*-linked carbohydrate structures. The synthesis of a short trivalent T antigen glycopeptide was reported by Guo. The approach benefited from straightforward work-up steps due to the *O*-unprotected glycan moieties. The convergent synthesis of the peptide part of *O*-glycopeptides hasbeen performed by fragment ligations of protected and unprotected peptide segments in organic and aqueous solvents, respectively. Ligation methods that allow the use of unprotected (glyco)peptides provide the opportunity to employ segments accessed by both chemical and recombinant means. In an approach known as expressed protein ligation, Macmillan and Bertozzi coupled chemically synthesized multiply glycosylated peptide thioesters with recombinant peptide fragments (Macmillan and Bertozzi, 2004). The peptides were expressed in *E. coli* as conjugates with a *C*-terminal chitin-binding domain. The removal of the chitinbinding domain proceeded through the intermediary formation of a thioester that was hydrolyzed. Subsequently peptide was treated with factor Xa protease which cleaved behind *C*-terminal to the recognition sequence Ile – Glu – Gly – Arg. The exposed cysteinylpeptide was submitted without purification to native chemical ligation with glycopeptidethioester (Macmillan and Bertozzi, 2004).

Glycan *O*-acetylation offers a convenient means to stabilize *O*-glycosidic bonds against electrophilic cleavage. Using glycosyl serine and threonine building blocks with *O*-acetyl protected glycans the Kunz group prepared a host of *O*-glycopeptides (Dziadek et. al. 2004). For example, a tandem repeat sequence of the epithelial mucin MUC4 was synthesized on Wang–Tentagel resin. The synthesis started from Fmoc-Asp(O*t*Bu) resin. The chain assembly was performed by employing standard protocols of Fmoc solid-phase peptide synthesis. However, the excess of acylating agents was reduced from ten to two equivalents in the coupling of the precious glycosyl amino acid building blocks. After the TFA treatment, performed to simultaneously cleave the Wang-linker and the amino acid side chain protection groups, the glycan moiety was deblocked. The sialic acid benzyl ester groups were removed by hydrogenolysis followed by mild acetate hydrolysis at pH 9 (Brocke and Kunz, 2004).

1.2.3 Enzymatic *in vitro* synthesis of glycopeptides or glycoproteins

Various types of sugar-modifying enzymes have been used in the synthesis of glycopeptides. Glycosyltransferases, exoglycosidases, and endoglycosidases have proven to be valuable catalysts for the formation and elaboration of glycans on glycopeptides.

The biosynthesis of oligosaccharides can be divided into *Leloir* and non-*Leloir* pathway enzymes. In mammalian systems, *Leloir* pathway enzymes account for the synthesis of most cell-surface glycoforms. These enzymes catalyze the transfer of a given carbohydrate from the corresponding sugar nucleotide donor substrate to a specific hydroxyl group of the acceptor molecule. Non-*Leloir* glycosyltransferases, which consume sugar phosphates as glycosyl donors, have not been applied widely in the synthesis of glycopeptides, and will not be discussed further. *Leloir* glycosyltransferases are ideal for the synthesis of glycopeptides due to their high regio- and stereospecific glycosidic linkage formation and high yield. Nevertheless, glycosyltransferases suffer from two shortcomings. First, the nucleoside diphosphates (NDPs) produced are often inhibitors of the glycosyltransferase. In addition, the expense of sugar nucleotide (NDP-sugar) can become a concern for large-scale synthesis. Fortunately, both of these problems can be solved by the use of multi-enzyme sugar nucleotide regeneration, wherein the product inhibition is avoided and expense reduced by recycling catalytic amounts of the NDPs to NDP-sugars (Wong et. al. 1982). Another obstacle is the availability of a wide diversity of glycosyltransferases. Though enzymes for all desired glycosidic linkages have not been isolated, several glycosyltransferases are available commercially and many others can be expressed recombinantly or isolated from tissue sources.

1.2.3.1 Use of glycosyltransferases for the synthesis of *O*-glycoproteins

Leloir glycosyltransferases are a favorable tool for the synthesis of *O*-linked glycopeptides, as the enzymatic transfer of individual monosaccharides to a preformed simple *O*-linked glycan of a synthetic glycopeptide is more convenient than difficult chemical glycosylations involving issues of protecting group manipulation and stereoselectivity. Such elaboration of glycans was used in the synthesis of glycopeptides containing an *O*-linked sialyl-Lewis-X (SLe*X*) tetrasaccharide, a glycopeptide fragment of P-selectin glycoprotein ligand-1 (PSGL-1), and tumor associated Tn (GalNAcαThr/Ser)-antigens and T (Galα1-3GalNAcαThr/Ser)-antigens and their sialylated forms. Various glycoforms of SLe*x*, SLe*a*, and their sulfated derivatives are moieties on adhesion molecules that are recognized by selectins. Interactions between selectins and their ligands leads to a “rolling” of the leukocyte on the endothelial cell surface as an inflammatory response to an injured site. In an effort to construct specific glycoforms of SLe*x* of interest to the study of cell-cell

adhesion, an *O*-linked glycopeptide containing β-linked GlcNAc was elaborated sequentially with β1,4-galactosyl-, α2,3-sialyl-, and α1,3-fucosyltransferases to yield a SL*e*x tetrasaccharide-containing glycopeptide (Seitz and Wong, 1997). The enzymatic glycosylation of the *O*-GlcNAc-octapeptide substrates was performed both in solution and on solid support, highlighting the flexibility of glycosyltransferases in chemoenzymatic methodologies for glycopeptide synthesis. More recently, a similar approach has been reported for the chemoenzymatic synthesis of glycopeptides with a spacer between the SL*e*x or SL*e*x mimetic glycan and peptide (Matsuda et. al. 2001).

Another example can be synthesis of glycoproteins containing mucins. The most common *O*-linked carbohydrates in eukaryotes are the mucins, which contain a core structure represented by the Tn- and T-antigens. Sialylated Tn- and T-antigens are expressed in low levels in many normal tissues but can become expressed abundantly in several types of human malignancies. In addition, mucins are involved in inflammation and cellular recognition. Due to such important biological connections, efficient methods for the preparation of Tn- and T-antigens and their sialylated forms have been realized. Chemoenzymatic syntheses of the α2-3sialylated T-antigen, including the glycosyl amino acid Neu5Acα2-3Galβ1-3GalNAcα- threonine and a neoglycopeptide, using the sialyltransferase ST3Gal, it has been reported (Gamblin et. al. 2009).

1.2.3.2 Use of glycosyltransferases in the synthesis of *N*-glycoproteins

The synthesis of *N*-linked glycopeptides is also enhanced by chemoenzymatic approaches incorporating the use of glycosyltransferases, with advantages analogous to those realized with *O*-linked glycopeptides. Similar to *O*-glycopeptides, the use of sialyltransferases is favorable as chemical introduction of sialic acid and deprotection of sialylated glycans remains challenging. Earlier work highlighted an efficient method for the chemoenzymatic synthesis of *N*-linked sialyllactosamine glycopentapeptides using galactosyl- and sialyltransferases in a two-step, one-pot reaction (Unverzagt et. al. 1990). Slight modification of this procedure was used for the enzymatic transfer of galactose and sialic acid onto glycopeptides with multiple acceptor sites, and these multivalent sialoglycopeptides were tested as inhibitors of influenza virus (Unverzagt et. al. 1994). Additional employment of this method by Unverzagt demonstrated a chemoenzymatic synthesis of a diantennary complex-type *N*-linked glycosyl asparagine by means of sequential, one-pot glycan elaboration with β1,4-galactosyl- and α2,6-sialyltransferases (Figure 10 on the next page). In this case the choice of protecting groups was critical for obtaining unprotected glycosyl asparagine for

enzymatic glycosylation. This chemoenzymatic approach provides a complex glycosyl amino acids for chemical synthetic strategies.

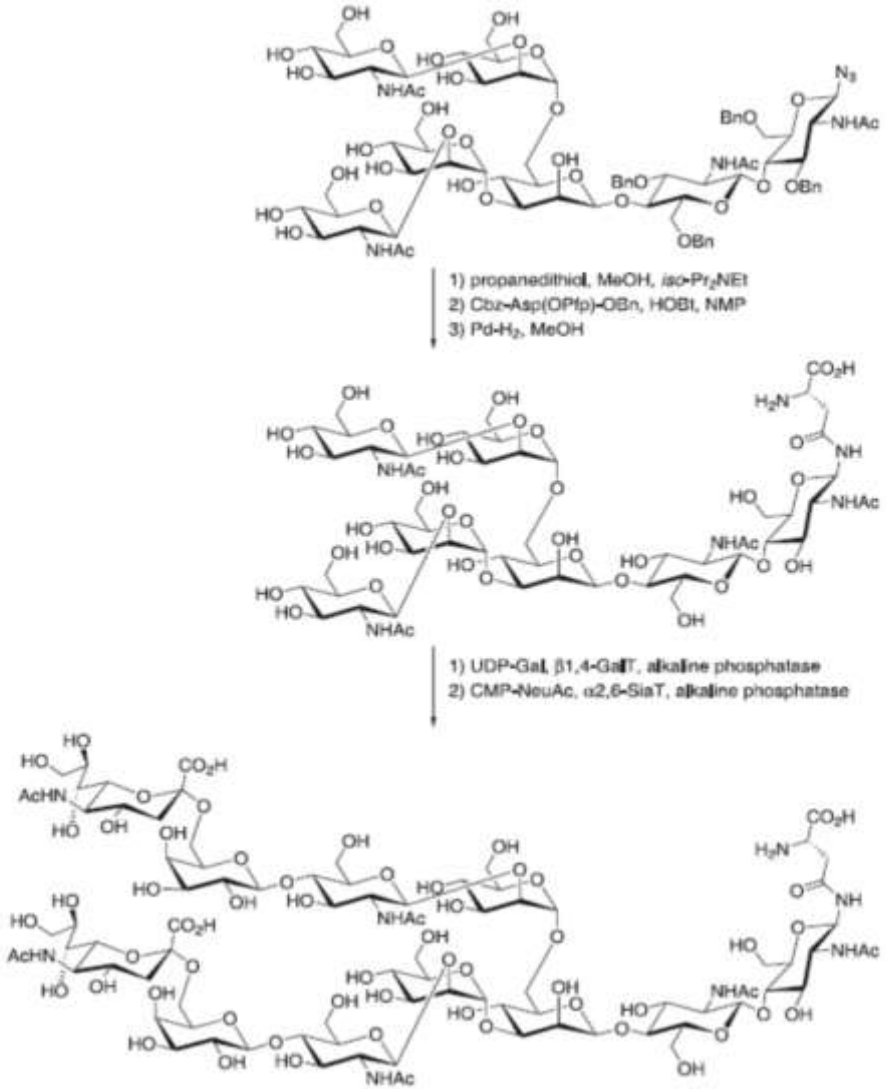


Figure 10. Chemoenzymatic synthesis of an *N*-linked complex-type diantennary glycosyl asparagine. The heptasaccharide core was elaborated to a complex-type undecasaccharide by sequential one pot enzymatic glycosylation.

1.2.3.3 *In vivo* synthesis of glycoproteins using specialized production cells

In vivo methods for the synthesis of glycoproteins have been established. Such methods, consequently, are the equivalent of in vivo enzymatic methods, whereby enzymes within the cell are used for synthetic purposes. Two types of *in vivo* glycoprotein synthesis have been explored,

including cotranslational incorporation of an unnatural amino acid by *in vivo* suppressor tRNA technology and post-translational modification of glycoproteins via pathway re-engineering in yeast. *In vivo* suppressor tRNA technology allows the site-specific incorporation of non-natural amino acids into proteins (Liu et. al. 2003). This process involves the evolution of an orthogonal tRNA synthetase and tRNA pair that selectively charges and inserts a selected unnatural amino acid into the protein in the place of an amber codon TAG. Using this method, engineered *E. coli* were used to produce neoglycoproteins and glycoproteins. Neoglycoproteins were generated by *in vivo* insertion of *p*-acetylphenylalanine into recombinant protein followed by chemical modification with aminooxy saccharides. In addition, homogeneous *O*-linked glycoproteins were produced by direct incorporation of the glycosyl amino acids GlcNAc-*β*-serine (Zhang et. al. 2004), and GalNAc-*α*-threonine (Xu et. al. 2004). These glycoproteins containing core monosac-charides glycans were then elaborated by glycosyltransferases *in vitro*. Though this method is distinguished for its ability to introduce glycans site-specifically into proteins, further development of this method is needed to achieve favorable production levels. Another method for *in vivo* synthesis of glycopeptides is pathway reengineering in yeast. *N*-linked glycosylation occurs in yeast, though conservation with human glycosylation only exists in the early steps of *N*-glycan assembly and processing. Differences in later steps result in high-mannose type glycans in yeast, which are not suitable for therapeutic human glycoproteins. Thus, efforts have been made to humanize *N*-glycosylation pathways in yeast for the production of therapeutic glycoproteins (Wildt and Gerngross, 2005). Such re-engineering requires altering endogenous *N*-glycosylation reactions and genetic incorporation of human *N*-glycosylation pathways into yeast and other fungi. Although the ability to humanize glycosylation pathways in yeast has progressed recently, future research is necessary for efficient yeast strains to be developed for homogeneous human glycoprotein synthesis.

1.2.4 Synthesis and application of glycopeptides and glycoprotein mimetics

Glycan chains of glycoproteins modulate physical and chemical properties of proteins, such as solubility, viscosity, charge, conformation, and dynamics and thereby can bestow stability and resistance to proteolytic degradation and can lead to improved protein properties. On the other hand, the glycans provide unique epitopes for molecular recognition that are involved in cell–cell communication, cell growth and differentiation, cancer metastasis, bacterial and viral infection, and they direct protein folding. Oligosaccharides are attached to proteins mainly via an *N*-glycosidic bond to asparagine or an *O*-glycosidic bond to hydroxylated amino acids, such as serine and

threonine, as is described abow. Whereas all *N*-glycoproteins share the same core structure, which is derived from a common biosynthetic oligosaccharide precursor, a variety of carbohydrate - peptide linkages are found in *O*-glycoproteins. In the mucins, the GalNAc(α1–*O*)Ser/Thr linkage (also called Tn antigen) is found. Another widely occurring *O*-glycosidic bond is the GlcNAc(β1-*O*)Ser/Thr linkage. Less common *O*-glycosidic linkages include those to hydroxylysine or tyrosine. The synthesis of glycopeptide and glycoprotein mimetics, represents an alternative strategy to produce single glycoforms for biological studies. The replacement of a natural structure with a non-natural one allows study of the influence of distinct structural elements on biological activity, but has many practical applications as well. Use of chemoselective ligation reactions makes glycoconjugate synthesis accessible to a broader community. Furthermore, *S*- and *C*-glycosidic bonds for example are more stable than the corresponding *O*-glycosides (both chemically and with respect to enzymatic degradation), which leads to an increased half life of a glycoconjugate within a biological system. Thus, beneficial properties of glycans may be permanently linked to a protein, which is an important aspect for pharmaceutical applications.

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. Combined with site directed mutagenesis, this approach allows for control of both site of attachment and type of saccharide. Alternatively, genetically engineered proteins containing non-proteinogenic amino acids with a unique chemical reactivity can be employed. Mixed linear and convergent assembly strategies have also been described. Glycopeptide mimetics or synthesized by the linear route have been extended in the glycan part by convergent attachment of an oligosaccharide through chemoselective ligation.

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. The approach is also of interest for combinatorial glycopeptide synthesis. If chemoselective ligation

reactions are employed for the sugar attachment, it is possible to modify unprotected peptides and even whole proteins.

The biological importance of glycans and the need for pure glycoforms of glycoproteins for biological studies and pharmaceutical applications have been the major driving forces toward the development of strategies for the preparation of these structures. The synthesis of glycopeptide and glycoprotein mimetics allows for structure-activity relationship studies but has many practical applications as well. Replacement of labile *N*- and *O*-glycosidic bonds by more stable *C*- and *S*-glycosides generates compounds with longer half-lives and simplifies their preparation by linear assembly strategies due to the increased stability of the corresponding building blocks. 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.

1.3 Immune system and its regulation

Immunology is a science that examines the structure and function of the immune system. It originates from medicine and early studies on the causes of immunity to disease. The earliest known mention of immunity was during the plague of Athens in 430 BC. Thucydides noted that people who had recovered from a previous bout of the disease could nurse the sick without contracting the illness a second time. In the 18th century, Pierre-Louis Moreau de Maupertuis made experiments with scorpion venom and observed that certain dogs and mice were immune to this venom. This and other observations of acquired immunity were later exploited by Louise Pasteur in his development of vaccinations and his proposed germ theory of disease. Robert Koch confirmed that the microorganisms are the cause of infectious disease. Viruses were confirmed as human pathogens in 1901, with the discovery of the yellow fever virus by Walter Reed.

Immunology made a great advance towards the end of the 19th century, through rapid developments, in the study of humoral immunity and cellular immunity. Particularly important was the work of Paul Ehrlich who proposed the side-chain theory to explain the specificity of the antigen-antibody reaction; his contributions to the understanding of humoral immunity were recognized by the award of a Nobel Prize in 1908, which was jointly awarded to the founder of cellular immunology, Elie Metchnikoff.

The immune system is composed of many interdependent cell types that collectively protect

the body from bacterial, parasitic, fungal, viral infections and from the growth of tumor cells. Many of these cell types have specialized functions. The cells of the immune system can recognize and destroy bacteria, kill parasites or tumor cells, or kill viral-infected cells. Often, these cells are depended on the other cells type for activation signals in the form of secretions known as cytokines, lymphokines, and interleukins.

An immune response to foreign antigen requires the presence of an antigen-presenting cell (APC), (usually either a macrophage or dendritic cell). When an APC presents an antigen on its cell surface to a B cell, the B cell is signalled to proliferate and produce antibodies that specifically bind to that antigen. If the antibodies bind to antigens on bacteria or parasites it acts as a signal for macrophages to engulf and kill them. Another important function of antibodies is to initiate the "complement destruction cascade." When antibodies bind to cells or bacteria, serum proteins called complement bind to the immobilized antibodies and destroy the bacteria by creating holes in them. Antibodies can also signal natural killer cells and macrophages to kill viral or bacterial-infected cells.

If the APC presents the antigen to T cells, the T cells become activated. Activated T cells proliferate and become secretory in the case of CD4+ T cells, or, if they are CD8+ T cells, they become activated to kill target cells that specifically express the antigen presented by the APC. The production of antibodies and the activity of CD8+ killer T cells are highly regulated by the CD4+ helper T cell. The CD4+ T cells provide growth factors or signals to these cells that signal them to proliferate and function more efficiently. This multitude of interleukins or cytokines that are produced and secreted by CD4+ T cells are often crucial to ensure the activation of natural killer cells, macrophages, CD8+ T cells.

1.3.1 Lymphocytes

Lymphocytes are type of leukocytes (white blood cells) that have major responsibility for the activities of the immune system. In humans lymphocytes make up 25 to 33 percent of the total number of leukocytes. They are found in circulation and they concentrated in central lymphoid organs such as spleen and lymph nodes where the specific immune response is initiated. Two major types of lymphocytes were found. The first type is B lymphocytes or B cells, which grow to maturity independend of the thymus. The second type is called T lymphocytes or T cells, this type migrate to thymus, where develop into adults cells. B and T cells are able to recognize specific antigens through receptors on their surface. The ability to respond to virtually any antigen comes

from the enormous variety of lymphocyte populations that the body contains, each of them with a receptor capable of recognizing a unique antigen. The cells are stimulated by binding to a foreign antigen, such as a component of a bacterium or virus, a lymphocyte multiplies into a clone of identical cells. Some of the cloned B cells differentiate into plasma cells that produce antibody molecules. These antibodies are closely modeled after the receptors of the precursor B cell, and, once released into the blood and lymph, they bind to the target antigen and initiate its neutralization or destruction. Antibody production continues for several days or months, until the antigen has been eliminated. Other B cells, the memory B cells are stimulated to multiply but do not differentiate into plasma cells; they provide the immune system with long-lasting memory. T-lymphocytes multiply and differentiate in the thymus into helper, regulatory, or cytotoxic T cells or become memory T cells. They are seeded to peripheal tissues or lymphatic system. Once stimulated by the appropriate antigen, helper T cells secrete chemical compounds called cytokines which stimulate the differentiation of B cells into plasma cells, thereby promoting antibody production. Cytotoxic T cells, which are activated by various cytokines, bind to and kill infected cells and cancer cells. There are not only disparaties among T and B cells, but also between subsets of these cells. Every mature T cells containg marker known as CD3. Most helper T cells carry a CD4 marker, a molecul that recognizes MHC class II types. Cytotoxic/supressor T cell containing molecule known as CD8 on their surface. This molecule is able to recognize MHC antigens class I. Different T cells have different class of antigen receptors either alpha/beta or gamma/delta.

NK cells (natural killer 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 to be converted into an activated form, an increased expression of MHC class II and secretion of IL-12, which is a significant differentiating factor for Th1 lymphocytes and inhibiting factor for the development of Th2 lymphocytes. Additionally, NK cells are producers of TNF-α and β which are necessary for stimulation (TNF-α) or inhibition (TNF-β) of mitosis, and participate also in inflammatory

processes. NK cells are also producers of so-called colony stimulating factors (CSF). These factors stimulate the differentiation of monocytes (M-CSF), granulocytes (G-CSF) and myeloid cells (GM-CSF). Finally, NK cells are also important in regulating autoimmune responses.

1.3.2 Receptors of lymphocytes

1.3.2.1 B cells receptors (BCRs)

B lymphocytes recognize antigens via their antigen-specific receptors (BCRs). The BCR is a membrane-bound immunoglobuline and it is this molecule that allows the distinction of B cells from other types of lymphocytes as well as being the main protein involved in B cell activation. The receptor is composed of two heavy and two light domains with variable regions on the N-termini of their polypeptide chains. Formation of a complex between BCR and other transmembrane molecules CD79α and CD79β is responsible for full activation of B lymphocytes.

1.3.2.2 T cells receptors (TCRs)

The T cells receptors are composed of two different chains that form the TCR heterodimer responsible for ligand recognition. CD3 molecules (CD3-gamma, delta, epsilon and zeta), which are assembled together with the TCR heterodimer, possess a characteristic sequence motif for tyrosine phosphorylation, known as ITAMs (Immunoreceptor Tyrosine-based Activation Motifs). The TCR polypeptides themselves have very short cytoplasmic tails,and all proximal signaling events are mediated through the CD3 molecules (Anderson et. al. 1989). Cytotoxic T lymphocytes activation is initiated by the interaction of the TCR with antigenic peptides complexed to MHC-I molecules on the surface of almost every cell in body, so it can be distinguish between self antigens and foreign antigens, for example viral-infected cells. Cytotoxic T lymhocytes are very sensitive to amount of self antigens presented, because increasing of antigen presentation can indicate to malignant cells. Helper T cell and inflammtory T cell TCR recognize antigen presented on MHC class II molecules on surface of antigen-presenting cells such as macrophage or dendritic cells that present antigen to T or B cells.

1.3.2.3 NK cells receptors

NK cells have two basic types of receptors on their surface: activating and inhibitory. According their structure we can divide the receptors of NK cells into immunoglobulin and C-type lectin families.

Activating receptors

Interaction of membrane receptors with their ligands leads to signal transduction into the cell. Since most NK cell receptors do not express ITAM motifs within their polypeptide chains, they associate with signal transmitting adaptor proteins via charged amino acids that are found in the transmembrane of the receptor. These adaptor molecules are e.g. DAP10 and DAP12, which contain tyrosine-binding activation immunoreceptors ITAM (Lanier et. al. 1998) motif ITAM phosphorylation is provided by protein-tyrosine kinases (PTK) of Src group. Signalling is also using the phosphatidylinositol-3-kinase and PTK ZAP10 and SYK. Another possibility of signal transmition can be through the G-proteins which are asocciated with the receptors. After ligand binding the G-proteins consisting of three subunits, α, β, and γ, disconnect from the receptor. GTP binds to α subunit instead of the previously bound GDP, which it leads to dissociation into βγ and α-complex. Subunits α and βγ can bind to other enzymes, and influence their activities. Among the NK cell activating receptors are human NKR-P1, CD69, NKG-2D, mouse NKR-P1A, C, and F, and CD69, and a group of receptors responsible for spontaneous (natural) cytotoxicity of NK cells, NKp30, NKp44, NKp46, and CD16, low affinity Fcγ receptor responsible for antibody-dependent cellular cytotoxicity (Smyth et. al. 2002).

Inhibitory receptors

These receptors contain in its intracellular part imunoreceptor tyrosine-based inhibitory motif (ITIM). To this motif, a protein-tyrosine phosphatase SHP2 and SHP1 is bound that dephosphorylate protein tyrosine kinases and thereby inhibit transmission of signals. Specific ligands for these receptors are MHC class I glycoproteins. Human MHC molecules are called HLA (human leukocyte antigen), which is divided into HLA-A, B, C called classical and HLA-E, F, G series called nonclassic (Horejsi and Bartunkova, 2001). Among the classical inhibitory NK cell receptor molecules belong mouse Ly-49 and NKR-P1B/D, and human KIR2DL, KIR3DL and NKG-2A.

Immunoglobulin receptors

These group of receptors called KIR (Killer Inhibition Receptors) (Biassoni et. al. 2000), so far have been found only on the surface of human NK cells. They are divided into two subfamilies, depending on the number of immunoglobulin domains in the extracellular part of molecule. The first subfamily contains two (KIR2D) and the other subfamily of three (KIR3D) immunoglobulin domain type. Ligands for these receptors are classical HLA molecules.

1.3.2.4 C-type lectin family receptors

Lectins were definied as proteins able to specifically recognize and reversibly (non-covalently) bind saccharide structures (Goldstein et. al. 1980). The lectins contain at least two or more sugar binding sites in their carbohydrate-recognition domains, as well as other protein modules which may be soluble or membrane embedded. The lectins have been found in a variety of species from viruses to plants and humans. The lectins can have variety of functions in organisms as adhesion molecules (selectins), immune-recognition molecules (mannosa binding receptor) and so on. Lectins can be classified by different criteria, but the most useful is classification based on structural and functional properties and similarities. Thus, we can distinguish eight main groups of lectins: calnexins, P-type lectins, L-type lectins, C-type lectins, I-type lectins, M-type lectins, R-type lectins and galectins (Dood and Drickamer, 2001). The term C-type lectins is generally used for the animals to distinguish them from plant lectins (Brewer, 2001).

The C-type lectins were first properly defined in the 80^{th} 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).

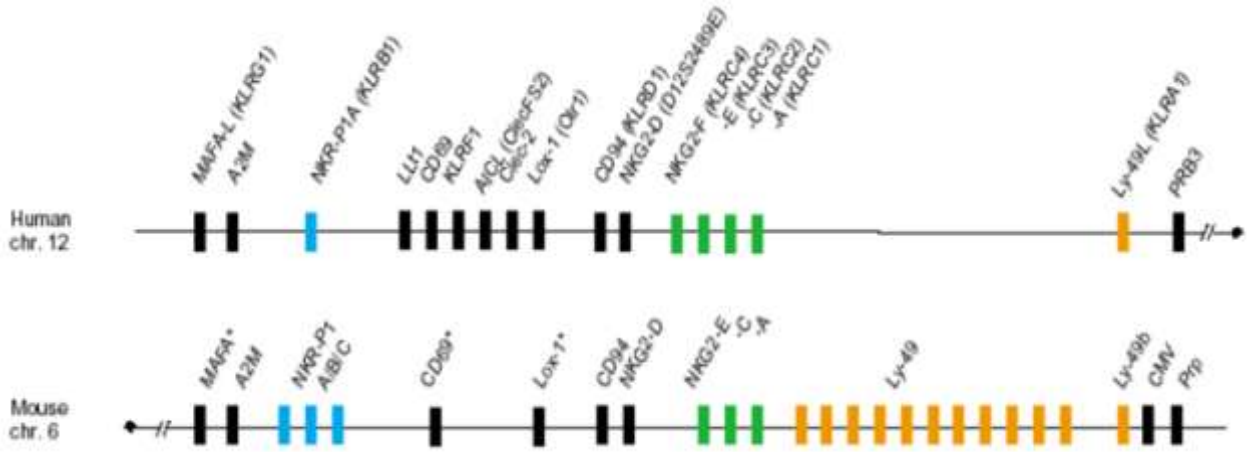


Figure 11. Comparison of NKC gene complex, encoding genes for C-type lectin receptors.

According to amino acids sequence homology we can divide the C-type lectins into seven different groups: proteoglycans, type II transmembrane receptors, selectins, collectins, type II lymphocyte receptors, macrophage mannose receptors and free CRD (Drickamer, 1993). From these groups mainly lymphocyte receptors are expressed on the surface of NK cells. These proteins contain one CRD on the extracellular C-terminus. Their tertiary structure was initially resolved in case of CD94 molecule although this domain is somewhat atypical due to the lack of one α-helix otherwise structurally conserved in CRD (Boyington et. al. 1999). CRD is usually glycosylated and receptors occur in the form of homodimers (except for heterodimer CD94-NKG2) (Lazetic et. al. 1996).

Genes encoding the lectin receptor of the C-type lectins are located in the NKC gene complex. In humans, NKC occurs on the 12th chromosome, in mice at the 6th and rats on the 4th chromosome. The genes for the polymorphic receptor family of Ly-49 and NKR-P1 are located away from the monomorphic genes for CD69 and CD94 (Figure 11). The group of Ly-49 family includes nine members of the Ly-49A - Ly-49I. Family Ly-49 receptors have inhibitory function for cell cytotoxicity. These receptors are present on the surface of mouse and rat NK cells. On the surface of human NK cells have not been discovered yet (Ryan and Seaman, 1997).

It seems that their function in human NK cells has been taken over by KIR receptors. As more was discovered family of five genes for C-type lectins marked as NKG 2A-E. NKG 2 receptors differ in the extracellular and cytoplasmic parts. This diversity allows great variability of receptors for their ligands. NKG 2 receptors are usually associated with glycoproteins referred to as CD94 (Yokoyama, 1998). NKG 2 can not be expressed on the cell surface without the presence of CD94. CD94 function is not yet entirely clear. Since the intracellular part of CD94 is too short and contains no signal sequence, it seems that facilitates transport of NKG 2 on the surface of cells.

1.3.2.5 Ligands of C-type lectin NK cell receptors

Ligands of the C-type can be divided into three major groups. The first group consists of receptors recognizing MHC class I or similar molecules. The best example is murine Ly-49 (KLRA1) receptor. Ly-49 receptors are inhibitory and serve as a self-recognition system. Murine Ly-49 is comprised approximatelly from 23 highly polymorphic genes (ly-49A – W) (Dimasi and Biassoni, 2005). Ly-49A was the first NK cell receptor whose structure in complex with its MHC class I was solved (Tormo et. al. 1999) and the missing self hypothesis was demonstrated right on this receptor. Ly-49 was found only in rodents, the same function in humans executes the KIR immunoglobulin-like receptor. The second member of first group is homodimeric activation receptor NKG-2D (KLRK1), which recognizes stress induced MHC class I ligands MICA and MICB or viral infection associated ligands ULBP (Lanier, 2005). NKG2/CD94 receptors are the last subfamily of the first group. This family is conserved in human as well as in mice. NKG2 and CD94 are covalent heterodimers which are responsible for signaling. Various forms are different functions: isoform A and B are inhibitory and C is activatory receptor. The gene for CD94 occurs in single form and is known in all species lying close NKG2A, -B, -C gene cluster (KLRC1, -2, -3).

The second group is composed of KLRG1 (MAFA), an inhibitory receptor. These receptors are expressed on the surface of mature T and NK cells. The ligands for KLRG1 have been found E-, N and R-cadherin (Li et. al. 2009). Cadherins (named for "calcium-dependent adhesion") are a class of type-1 transmembrane proteins. They play important roles in cell adhesion, ensuring that cells within tissues are bound together. They are dependent on calcium (Ca^{2+}) ions to function, hence their name. The cadherin superfamily includes cadherins, protocadherins, desmogleins, and desmocollins, and more. In structure, they share *cadherin repeats*, which are the extracelular Ca^{2+} binding domains. There are multiple classes of cadherin molecule, each designated with a prefix (in general, noting the type of tissue with which it is associated). It has been observed that cells containing a specific cadherin subtype tend to cluster together to the exclusion of other types, both in cell culture and during development. For example, cells containing N-cadherin tend to cluster with other N-cadherin expressing cells. However, it has been noted that the mixing speed in the cell culture experiments can have an effect on the extent of homotypic specificity. In addition, several groups have observed heterotypic binding affinity (i.e., binding of different types of cadherin together) in various assays (Volk et. al. 1987). One current model proposes that cells distinguish cadherin subtypes based on kinetic specificity rather than thermodynamic specificity, as different types of cadherin homotypic bonds have different lifetimes (Bayas et. al. 2005).

The last group of C-type lectin NK cell receptor-ligand pairs are able to recognize another

C-type receptors. These receptors are more likely to be conserved during evolution because receptor-ligand pair are mixed with each other in the NKC gene complex. The first receptor ligand pair identified in this group were mNKR-P1F and mNKR-P1B/D (Carlyle et. al. 2004).

1.3.2.6 Cytotoxic activity of NK cells

The NK cells are two mechanisms known cytotoxic effect and it is the natural cytotoxicity, which takes place either via the apoptotic Fas ligand or lytic path in the presence of calcium, and cellular cytotoxicity dependent on antibody against the target structure (ADCC). After meeting the target cells with NK cell realize interaction between the surface receptors of both cells, strengthening mutual ties and run the lytic mechanisms. Target cell after receiving the "kiss of death" is dying, NK cells detach from it and goes to other potential target.

Antibody-dependent cellular cytotoxicity (ADCC)

ADCC as one of the mechanisms used by NK cells is part of a specific adaptive immunity. When the NK cell meets with a cell that is opsonized by IgG antibodies the NK cell trigger the reaction. On the surface of NK cells occurs stimulatory low-affinity receptor for IgG CD16 (FcgRIII). This receptor binds to the Fc portion of antibodies. CD16 was found in most human and murine NK cells, activated monocytes and on the surface of T cells. After binding ligand on CD16 leads to activation of PTKase of Src family bound on tyrosine residues in the cytoplasmic ITAM domen (Leibson, 1997). Further leads to activation of phospholipases, stimulation of fosfatidilinositol-3-kinase and induction of MAP kinase, leading to increased levels of intracellular calcium (Perrusia, 2000). Increasing of Ca^{2+} leads to triggering of cytotoxic mechanisms leading to apoptosis. ADCC has the dominant activity of antibodies against tumor-transformed cells (O'Shea et. al. 1991).

Lytic way dependent on calcium ions

In the cytoplasm of NK cells is a large amount of cytotoxic granules (specialized lysosomes), which contain a protein called perforin and proteases called granzyms. After NK cell recognizing of damaged or otherwise altered cell using of specialized surface receptors, the cytotoxic granules migrate to the plasma membrane, fuse with it and its contents are poured into the narrow gap between the two cells (degranulation). The pores created by perforin in the cytoplasmic membrane (perforin is structurally similar complement protein C9) enter into the cell interior

granzymy. These in turn cleaves proteasis precursors from the group of caspas in the cytoplasm, which are thus activated. Caspases affect to other proteins in the cytoplasm, leading to run the whole series of reactions leading to apoptotic death of target cells (Trapani et. al. 2000). Why granzymy acts only on the target cell and not damage the cells that secret him it is not yet entirely clear. One possible explanation is that granzymy are secreted as a whole and then fuse with the membrane of the target cells without damaging of effector cell.

Fas-Fas ligand path (Fas-FasL)

On the surface of NK cells is a protein called Fas ligand (FasL). This protein belongs to the TNF (Tumor Necrosis Factor) of the surface molecules. FasL binds to the apoptotic receptor Fas (CD95) (Ashkenazi, 2002), which is present on the surface of many different cells. By binding to this receptor are activated a cascade of reactions leading to apoptotic cell death. On the surface of NK cells express three apoptotic receptors: TRAIL ligand, TNF and already have mentioned FasL. On the surface of NK cells is also found apoptotic receptor Fas, which provides an important feedback regulation of NK cells (Montel et. al. 1995). However, some tumor cells use this fact for security again the attack by the immune system, they express surface FasL and thereby induce apoptosis of cytotoxic cells.

2. Introduction to methods

2.1 Molecular biology

The history of molecular biology begins in the 1930s with the convergence of various, previously distinct biological disciplines: biochemistry, genetics, microbiology, and virology. In 1940, George Beadle and Edward Tatum demonstrated the existence of a precise relationship between genes and proteins (Beadle and Tatum, 1941). In the course of their experiments connecting genetics with biochemistry, they switched from the genetics mainstay Drosophila to a more appropriate model organism, the fungus Neurospora. In 1944, Oswald Avery demonstrated that genes are made up of DNA (Avery et. al. 1944). In 1952, Alfred Hershey and Martha Chase confirmed that the genetic material of the bacteriophage is made up of DNA. In 1953, James Watson and Francis Crick discovered the double helical structure of the DNA. In 1961, Francois Jacob and Jacques Monod hypothesized the existence of an intermediary between DNA and its protein products, which they called messenger RNA. Between 1961 and 1965, the relationship between the information contained in DNA and the structure of proteins was determined and named as genetic code, which creates a correspondence between the succession of nucleotides in the DNA sequence and amino acids in proteins. The chief discoveries of molecular biology took place in a period of only about twenty-five years. Another fifteen years were required before new and more sophisticated technologies as are polymerase chain reaction (PCR), reverse transcription, restriction digestions, gel electrophoresis, blotting and probing, DNA or RNA micro arrays, quantitative PCR amplification and directed-site mutagenesis.

The new techniques and methods are being developed and improved for achieving faster methods, higher senstitivity or higher yields. For example, we can buy the kit for isolation of plasmid DNA from bacterial strains. Isolation by this kit takes around one hour. The older protocol takes around 12 hours. On the other hand is it still necessary to know principle of older methods, this knowledge can help us in solving problems.

2.2 Recombinant protein expression

Proteins are best expressed in their native cell type under physiological conditions, which ensure a correct protein synthesis, posttranslation modifications, protein folding or subcellular targeting. Only a small amount of proteins occur naturally in such amounts that would suffice for biochemical or structure studies. Of these reasons many various heterologous expression systems

have been developed to produce of recombinant proteins. The expression systems we can be divided into two groups: prokaryotic (bacterial) which includes commonly *Escherichia coli* expression system and eukaryotic expression system including yeast (*Pichia pastoris* and *Saccharomyces cerevisiae*), baculovirus-infected insect cells, mammalian cells and recently developed cell-free system (Yokoyama, 2003).

2.2.1. Prokaryotic expression systems

Prokaryotic expression systems are generally easier to handle and are satisfactory for most purposes. Of these reasons is this system still the most widely used for recombinant protein expression. More than 90% of all solved protein structures are originated from recombinant proteins produced in *Escherichia coli*. This bacteria is very favorite, because we have many knowledge about its genetic, physiology and complet genomic sequence, which greatly facilitates gene cloning and protein expression. The high growth rates combines with the ability to express high levels heterologous proteins and rapid growing to high densities in simple and inexpensive media. These of properties make *E. coli* a very useful tool for recombinant protein expression system. However, expression in *E. coli* does have some disadvantage, of course. The most important is inability of posttranslation modifications. It is significant problems, because many proteins can not be able to fold to native structure or they are not active. Recombinant proteins are very often produced into insoluble protein agregates called: inclusion bodies. Inclusion bodies might to be harvested and solubilized in chaotropic reagents and successfully refolded *in vitro*. Unfortunately, the refolding yields are very often low, making the whole process ineffective.

2.2.2 Eukaryotic expression systems

Eukaryotic expression systems such as yeast, insect cells or mammalian cells are able to perform a number of posttranslational modifications such as disulfide bridge formation, glycosylation or other. Eukaryotic expression systems can be divided into categories based on the nature of cellular system used for recombinant expression. In this work I will describe only the yeast expression system *Pichia pastoris* and *Saccharomyces cerevisiae,* since these expressions systems were used for production of recombinant α-NAGA.

Yeasts are favoured alternative host for produce of recombinant proteins. Yeasts combine some of the advantage of prokaryotic and eukaryotic based expression systems. For example, they can be grown relative rapidly on simple and cheap media to high cell density. They possess the

cellular machinery to carry out posttranslational modifications. Added advantages are the availability of complete genomic sequence, the using nuclear stable high copy plasmids and ability to secrete the target protein.

Pichia pastoris

The methalotrophyc budding yeast *Pichia pastoris* was first described by Koichi Ogata in 1969. It was initially intended for single cell protein production by Phillips Petroleum Company due to low cost of methanol at the time (Ogata et. al. 1969). Phillips Petroleum Company developed media and methods for growing of Pichia pastoris on methanol in continuous culture at high cell density (>130 g/L dry cell weight). In the early 1980s, Phillips Petroleum Company contracted with the Salk Institute Biotechnology Industrial Associates (SIBIA), to develop *P. pastoris* as heterologous gene expression system. Researchers at SIBIA isolated the AOX1 gene and its promoter and develop vectors, strains and methods for molecular genetic manipulation of *P. pastoris*. In 1993, Phillips Petroleum Company licensed Invitrogen to sell components of the system to researchers worlwide.

Pichia pastoris has many of the advantage of higher eukaryotic expression system such as protein processing, protein refolding and posttranslational modification, while being to easy manipulate as *E. coli*. It is easier, faster and cheaper to use than other eukarytioc expression system. *Pichia pastoris* shares the advantages of molecular and genetic manipulation with *Saccharomyces cerevisiae*, but it has 10 to 100 fold higher heterologous proteins expression levels. Furthermore *Pichia* can grow between pH 3-7 without any significant changes in specific growth and low pH can eliminate the proliferation of many contaminating microorganism. All of these features make *Pichia* very useful as expression system.

Saccharomyces cerevisiae

For a variety of reasons, the common yeast *Saccharomyces cerevisiae* has been used extensively as a host cell for expression of cloned eukaryotic gene. First, it is single-celled, it is extremely well known genetically and physiologically, and it is grown readily in both small culture vessels and large-scale reactors. Second, several string promoters were isolated from yeast and characterized. Naturally occuring plasmids call 2μm, these plasmids can be used as part of and endogenous yeast expression vector system. Third, the *S. cerevisiae* is capable of carrying out many posttranslation modifications. This yeast normally secretes so few proteins that, when it is engineered for extracellular release of recombinant protein, the product can be easy purified.

Saccharomyces cerevisiae is used in baking and brewing industries, so this yeast has been listed in by U.S. Food and Drug Administration as a "generally recognized as safe" (GRAS) organism. Therefore, the use of this organism for production for human therapeutic agents (drugs or pharmaceuticals) does not require the same extensive experimentation demanded for unapproved host cells. A number of proteins that have been produced in *S. cerevisiae* are currently being used as commercial as a vaccines, pharmaceuticals or diagnostic agents. For example at present more than 50% of insulin is produced by *S. cerevisiae.* The advantages of recombinant protein expression in *S. cerevisiae* are the same as in *P. pastoris* described above. However, in some cases expression was low and the yields were modest. In other instances, the recombinant protein was hyperglycosylated with more than 100 mannose residues in *N*-linked oligosaccharides chains. The mannose often alters functions and makes the recombinant protein antigenic. Several approaches to humanizing yeast *N*-glycosylation pathway have been attempted over the past decade with limited success. However recently, was prepared *N*-glycoengineering strain of *S. cerevisiae* with complex type of glycosylation.

The last disadvantage is problem with secretion of recombinant proteins. Proteins were designed for secretion very often retained in the periplasmic space, increasing the time and cost of purification. Despite all of the disadvantages is *S. cerevisiae* still much used as expression system and exist proteins which was successful expressed only in this yeast.

2.3 Protein renaturation *in vitro*

Correct protein function (enzymatic activity, binding specifity) is mainly dependent on its structure and the structure is determinated by amino acids sequence. Even modest changes in protein structure can result in disfunctions of protein. Bruce Merrifield showed that simple protein is able to fold in solution spontaneously. On the basic of these experiments we can conclude, that the amino acids sequence not only determines the structure but also influence the folding of protein. Although the process of denatured protein refolding is different from folding of newly synthetized one, the theory of protein folding can be applied to both of these processes. There is a balance between two kinds of forces - the stabilizing (disulfide bonds formation, van der Waals, hydrogen bonds, hydrophobic interactions) and the destabilizing (electrostatic repulsion, chain entropy) and the protein folding is the result of this balance. The stable and correct folding state is characterized lower value of Gibss free energy. Protein can enter to local minimums (folding

traps) on the way from unfolded to native conformation which could be hard or even impossible to continue to fully native conformation. In refolding protein *in vitro* we try to minimize entering of protein to local minimums.

Generally the process of *in vitro* refolding consists of two steps, the denatured protein (inclusion bodies) solubilization in chaotropic agents such as urea or guanidinium hydrochloride. The protein is solubilized and transfer into non-denaturing enviroment that could be supplemented with stabilizing chemicals such as arginine as low molecular chaperone or glycerol and redox system such as reduced/oxidized glutathione or cysteamine/cystamine for support of disulfide bonds formation. We can use three main methods for *in-vitro* refolding: rapid dilution, dialysis and chromatographic methods.

The principle of fast dilution method is that denatured protein is slowly dropped into the stirred refolding buffer, what ensures the lowest possible local concentration of unfolded protein in solution and possibility of correct refolding of protein.

Dialysis method is based on slow exchange of solution through the semipermeable membrane, however sometimes it is difficult to keep the protein concentration low enough. As the concentration of denaturing agent is used in high concentration, the dialysis should be done in several steps. In any of them the concentration of refolded protein should not exceed 0.1 mg/ml, otherwise the protein could aggregate.

Last widely used refolding techniques are based on chromatographic methods. Gel filtration or affinity chromatography are the most suitable for refolding of proteins.

2.4 Physical biochemistry

2.4.1 Chromatography

Chromatography is physical methods widely used in biochemistry. Chromatography allows effective separation of substances necessary for reliable identification and quantification of components. The distribution of substances based on their different mobility in a two-phase stationary (anchored) and mobile (moving). Different substances differ in their adsorption properties, the values of partition coefficients in its dimensions and in their charges, which can all be used in chromatography for their division of compounds using suitable chromatographic equipment. Chromatography may be preparative or analytical. The purpose of preparative chromatography is to separate the components of a mixture for further use (and is thus a form of purification). Analytical chromatography is done normally with smaller amounts of material and is

for measuring the relative proportions of analytes in a mixture. The first chomatography experiments were separation of plant pigments such as chlorophyll at the begining of 20th century. These separations carried out Rusiian botanist Mikhail Semyonovich Tsvet. The new types of chromatographic techniques (gas chromatography, paper chromatography and high performance liquid chromatography) were developed Archer John Porter Martin and Richard Laurence Millington Synge during 1930s and 1940s. The development tends to higher efficiency, resolution, sensitivity and reproducibility and shorter separation times, similarly as in other analytical and separation techniques. The new stationary phases with better hydrodynamic properties and with different size of the stationary phase particles were developed. The particle size of the stationare phases is very important for determination of the maximal applicable liquid pressure during separation.

We can distinguish types of chromatography according the stationare phases and interaction of sample with this phase (ion exchange, hydrophobic interactions, affinity, size exclusion), and further by embodiment of chromatography (planar, think layer, column) and type of mobile phase (gas, liquid).

2.4.2 Analytical ultracentrifuge

Theodor Svedberg invented the analytical ultracentrifuge in 1925 and won the Nobel Prize in Chemistry in 1926 for his research on colloids and proteins using the ultracentrifuge. The vacuum ultracentrifuge was invented by Edward Greydon Pickels. It was his contribution of the vacuum which allowed a reduction in friction generated at high speeds. Vacuum systems also enabled the maintenance of constant temperature. In 1946, Pickels cofounded Spinco (Specialized Instruments Corp.) and marketed an ultracentrifuge based on his design. Pickels, however, considered his design to be complicated and developed a more “foolproof” version. But even with the enhanced design, sales of the technology remained low, and Spinco almost went bankrupt. The company survived and was the first to commercially manufacture ultracentrifuges, in 1947. In 1949, Spinco introduced the Model L, the first preparative ultracentrifuge to reach a maximum speed of 40,000 rpm. In 1954, Beckman Instruments (now Beckman Coulter) purchased the company, forming the basis of its Spinco centrifuge division.

In an analytical ultracentrifuge, a sample being spun can be monitored in real time through an optical detection system, using ultraviolet light absorption and/or interference optical refractive index sensitive system. This allows the operator to determine sample purity, characterize assembly

and disassembly mechanisms of biomolecular complexes, macromolecular conformational changes, thermodynamic parameters, determine the subunit stoichiometries and measure of equilibrium constant. With modern instrumentation, these observations are electronically digitized and stored for further mathematical analysis. Two kinds of experiments are commonly performed on these instruments: sedimentation velocity experiments and sedimentation equilibrium experiments.

Sedimentation velocity is technique which reffer entire time-course of sedimentation, and report on the shape and molar mass of the dissolved macromolecules, as well as their size-distribution. The size resolution of this method scales approximately with the square of the particle radii, and by adjusting the rotor speed of the experiment size-ranges from 100 Da to 10 GDa can be covered. Sedimentation velocity experiments can also be used to study of the complex composition from multi-signal analysis exploiting differences in each components spectroscopic signal, or by following the composition dependence of the sedimentation rates of the macromolecular system, as described in Gilbert-Jenkins theory.

Sedimentation equilibrium experiments are not concerned only with the final steady-state of the experiment, where sedimentation is balanced by diffusion opposing the concentration gradients, resulting in a time-independent concentration profile. Sedimentation equilibrium distributions in the centrifugal field are characterized by Boltzmann distributions. This experiment is insensitive to the shape of the macromolecule, and directly reports on the molar mass of the macromolecules. Information that can be obtained from an analytical ultracentrifuge include the gross shape of macromolecules, the conformational changes in macromolecules, and size distributions of macromolecular samples. For macromolecules, such as proteins that exist in chemical equilibrium with different non-covalent complexes the number and subunit stoichiometry of the complexes and equilibrium constants can be studied.

2.4.3 Mass spectrometry

Mass spectrometry (MS) is an analytical technique for the determination of mass to charge ratio (*m/z*) of the charged sample molecules. Further more it could be used for for elucidation of protein or peptide structure. The principle of this technique is ionization of sample molecules and determining the mass-to-charge ratio of created ions in the gas phase from their behavior in the electromagnetic field. Furthermore the complex molecules could be fragmented and *m/z* ratio of these fragments could be measured.

The history of mass spectrometry dates back more than one hundred years and has its roots in physical and chemical studies regarding the nature of matter. The study of gas discharges in the middle 19th century led to the discovery of anode and cathode rays, which turned out to be positive ions and electron. Improved capabilities in the separation of these positive ions enabled the discovery of stabel isotopes of the elements. The first such discovery was with the atom neon, which was shown by MS to have at least two stable isotopes: neon-20 with 10 protons and 10 neutrons and neon-22 with 10 protons and 12 neutrons. In 1950s was mass spectrometry used as the detector in gas chromatograhy. Chemical ionization was developed in 1960s, field desorption ionization in 1969s. Fourier transform ion cyclotron resonance mass spectrometry was developed by Alan G. Marshall and Melvin B. Comisarow at the University of British Columbia in 1974. The inspiration was earlier developments in conventional ICR and Fourier Transform Nuclear Magnetic Resonance (FT-NMR) spectroscopy, following electrospray ionization and MALDI ionization. To data the most powerful mass spectrometers are able to determine the mass of large proteins (>100 kDa) with resolution of less than 1 Da.

Each mass spectrometer is composed from three parts: ion source, mass analyzer and detector. In protein science the most common ion sources are matrix-assisted laser desorption ionization (MALDI), when the sample is co-crystallized on the target plate with matrix, and electrospray ionization (ESI), when the sample is ionized and transferred into the gas phase using electric field. The most useful analyzers are time-of-flight (TOF), quadrupole, quadrupole ion-trap and Fourier transform ion cyclotron resonance.

Detergents and mass spectrometry

Detergents represent indispensible tools in protein research. They are widely used for isolation and solubilization of membrane proteins, isolation of lipid rafts, or as antiaggregating agents in protein *in vitro* folding and crystallization. Even though they can prove extremely useful, their presence may also have tremendous impact on the protein sample or its subsequent analysis by mass spectrometry (MS). Depending on the concentration, detergents may cause signal suppression, adduct formation, and shape distortion of protein signals. Polyoxyethylene based detergents (POE) negatively affect the analysis even at very low concentrations. Alkyl-glycosides are less harmful and reasonable spectra can be acquired in their presence, although they may cause charge state shifts in the protein spectra.

Removal of detergents is mainly based on two principles, chromatography or extraction. Chromatographic methods separate the proteins/peptides from detergent based on their different

hydrophobicity or charge. Ion exchange chromatography works well for ionic detergents but can be also used for selective trapping of proteins/peptides from mixtures with nonionic detergents. Specific arrangements include mixed bed resin (e.g., NID trap from MichromBioresources). Separation based on hydrophobicity utilizes HILIC columns from which the NID elutes first and the proteins/peptides are eluted by decreasing percentage of organic solvent. Despite several advantages, HILIC columns do not offer separation efficiency comparable to reversed phase (RP) chromatography. Yeung et al. have used ethyl acetate for quantitative removal of alkyl-glycosides and for partial removal of POE-based detergents like Triton X-100 and NP-40. Another work, also based on ethyl acetate extraction, described removal of ionic sodium desoxycholate. Extraction protocols similar to those mentioned above for intact proteins can be applied to peptide mixtures as well. One advantage is that due to the digestion large hydrophobic polypeptides are very often split into shorter and less hydrophobic peptides. Nevertheless, all techniques for detergent removal are suffering from some drawbacks. As stated above, dialysis and adsorption techniques are more suitable for detergent exchange or for lowering their concentration. Precipitation can be problematic with small sample amounts and some proteins may be difficult to redissolve. The use of polyacrylamide gels may result in lower recovery or loss of larger and hydrophobic peptides. Specific precipitations may cause losses of hydrophobic protein/peptides, as these can coprecipitate with the detergent. Chromatographic techniques can, especially in the ion exchange mode, also suffer from losses of specific peptides. This behavior depends on the conditions of chromatographic separation and represents a problem mainly for complicated peptide mixtures covering a broad range of isoelectric points. In the HILIC mode, phase separation may occur at higher salt concentrations. Extraction with organic solvents is not very efficient and several extraction steps must be performed to remove the detergent quantitatively. In addition, hydrophobic peptides can easily be lost as they may partition into the organic phase and, with lower volumes, it might be difficult to aspirate the solvent from the organic phase due to low viscosity and convex meniscus. But the major disadvantage of extraction techniques remains in their “offline” arrangement (Rey et. al. 2010).

2.4.4 Electron microscopy

An electron microscope is a type of microscope that uses a particle beams of electrons to illuminate the sample and produce a magnified image. Electron microscopes (EM) have a greater than a optical microscope, because electrons have wavelengths about 100,000 times shorter than

photons, and can achieve better than 50 pm resolution and of up to about 10,000,000 x. The electron microscope uses electrostatic and electromagnetic "lenses" to control the electron beam and focus it to form an image. These lenses are analogous to, but different from the glass lenses of an optical microscope that form a magnified image by focusing light on or through the sample. In transmission, the electron beam is first diffracted by the sample, and then, the electron microscope “lenses" re-focus the beam into a Fouriere-transformed image of the diffraction pattern for the selected area of investigation. The real image thus formed is magnified by a factor ranging from a few hundred to many hundred thousand times, and can be viewed on a detecting screen or recorded using photographic film or plates or with a digital camera. Electron microscopes are used to observe a wide range of biological and inorganic samples including microorganism, cells, large molecules, metals and crystals. The history of electron microscopy began in 1931s by German physicists Ernst Ruska and Max Knoll, who constructed first prototype of electron microscopy. Five years later (1936) developed Helmut Ruska applications for the microscope, especially for biologic samples. Manfred von Ardenne constructed the first practical electron microscope in 1938, at the University of Toronto. Two years later company Siemens produced the first commercial transmission electron microscope (TEM).

We can divide electron microscopy to four groups: transmission electron microscope (TEM), using transmitted electrons, scanning electron microscopy (SEM), using secondary electrons. Unlike the TEM, where electrons of the high voltage beam carry the image of the sample, the electron beam of the scanning electron microscopy does not at any time carry a complete image of the sample. The SEM produces images by probing the specimen with a focused electron beam that is scanned across a rectangular area of the sample (raster scanning). The third group is reflection electron microscopy (REM), where same as in TEM and electron beam is incident on the surface of the sample, but instead of using transmission or secondary electrons, the reflected beam of elastically scattered electrons is detected. The last group is scanning transmission electron microscopy (STEM). The STEM rasters a focused incident probe across a sample has been thinned to facilitate detection of electrons scattered through the sample. The high resolution of the TEM is thus possible in STEM.

Aims of the thesis

- To determine 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 enzymatical properties of α-NAGA.
- To determine *N*-glycosylation sites of α-NAGA.

3. Methods

3.1 Molecular biology methods

DNA cloning is a process of obtaining multiple copies of desired DNA (usually a gene of interest) *in vitro*. It could be include processes such as RNA isolation, reverse transcription, PCR amplification or DNA manipulations such as restriction, ligation into vectors. In this thesis DNA cloning usually followed this workflow: RNA isolation using RNeasy Plant Mini Kit (QIAGEN), reverse transcription (SuperScript III, Invitrogen), PCR using DeepVent DNA polymerase, restriction digestions and ligation into pBlueScript SK+ (Strategene) was used for routine cloning in XL-1 MRF' Blue (Stratagene) or NovaBlue competent cells (Merck). Vector pET30a+ was used as expression vector for *E. coli* expression, while pPICZα and pYES2CT were used for electroporation of yeast expression system *Pichia pastoris* and *Saccharomyces cerevisiae*.

Isolation of poly(A) RNA and construction of a cDNA library

A submerged culture of *Aspergillus niger* was cultivated for 48 hours. Fresh mycelium (4 g, wet weight) was frozen with liquid nitrogen and homogenized. Total RNA was prepared using an RNeasy Plant Mini Kit (QIAGEN). First-strand cDNA was synthesized directly from the total RNA by Moloney Murine Leukemia Virus reverse transcriptase (Invitrogen), and the second strain was synthesized by DeepVent DNA polymerase (Invitrogen). These cDNAs were purified using a DNA purification kit (Genomed).

Cloning the cDNA encoding α-NAGA from Aspergillus niger

I have to divide the preparing of cDNA encoding α-NAGA to three parts according to which expression system was used.

Firstly, the prokaryotic expression system *E. coli* was used. One set of primers was designed to amplify the α-NAGA gene. A sense primer (5′-TTCCATATGGGCTTCAA-CAATTGGGCCCGC-3′) containing the NdeI restriction site (underlined) and antisense primer (5′-CCAAGCTTATCAGAGCCGGAAGACTGCTGTCGC-3′) containing HindIII site (underlined). The fragments were amplified using Pfu DNA Polymerase (Invitrogen) and a Thermal Cycler GeneE (Techne). The PCR product (fragment 1462 bp) was amplified and ligated into the Bluescript II SK- vector (Invitrogen) and transformed into *E. coli* BL21-DE3 (Strategene). The resultant plasmid, Bluescript II SK/α-NAGA was confirmed by restriction digestion and

sequencing. The Bluescript II SK/α-NAGA plasmid was digested by NdeI and HindIII and the gene product was ligated into pET30a+ (Invitrogen).

Pichia pastoris as second expression system was used. These primers were designed for this expression system: a sense primer (5´-CCGGAATTCATGGGTTTCAACAATTGGGCCCGC-3´) containig EcoRI restriction site (underlined) and antisense primer (5´-ATTTCTAGATTAGCCATCCCTCTCATAAAGACACGACTT-3´) containing restriction site XbaI (underlined).

Saccharomyces cerevisiae was used as last expression system. Primers for this expression system are: A sense primer (5´-TATTCTAGAATGGGTTTCAACAATTGGGCCCGC-3´) containing the EcoRI site (underlined) and antisense primer (5´-ATTGAATTCTTAGCCATCCCTCTCATAAAGACACGACTT-3´) containing the XbaI restriction site (underlined) were designed on the basis of the known amino acid sequence.

The fragments were amplified using the same DNA polymerase and the same conditions as in *E.coli* expression system described above. The PCR product was amplified and ligated into the Bluescript II SK- vector (Invitrogen) and transformed into *E. coli* BL21-DE3 (Strategene). The resultant plasmid, Bluescript II SK/α-NAGA, was confirmed by restriction digestion and sequencing. The Bluescript II SK/α-NAGA plasmid was digested by EcoRI and XbaI and the gene product was ligated into pPICZα (Invitrogen) in the case of *P. pastoris* expression system and into pYES2CT (Invitrogen) in the case of *S. cerevisiae* expression system.

The expression plasmids containing gene of interest were sequenced and a stocks of solution of purified DNA were prepared using JETSTAR 2.0 Maxi kit (Genomed).

The expression plasmid pET30a+ was transformated into prokaryotic expression system *E.coli* BL-21 (DE3) Gold by methods of heat shock.

Pichia pastoris GS115 was transformed by pPICZα expression plasmid by electroporation (25 μF, 30 ns, 2.0 kV, MicroPulserTM Electroporator, Bio-Rad) and the transformed cells were selected on the Zeocin selective plates.

Expression vectors pYES2CT was transformed into *S. cerevisiae* W303-1A (Invitrogen) by electroporation (25 μF, 22 ns, 1.2 kV, MicroPulserTM Electroporator, Bio-Rad) and the transformants were selected on an SC plate supplemented with tryptofan, leucine, adenine and histidine, without uracil.

3.2 Protein expression and purification

3.2.1 Recombinant expression in *E. coli*

In this thesis proteins were produced in *E. coli* BL-21 (DE3) Gold or in ArcticExpress *E. coli* system with cold-active chaperones (Strategene). ArcticExpress competent cells have been engineered for improved protein processing at low temperatures. These cells co-express the cold-adapted chaperonins Cpn10 and Cpn60 from the psychrophilic bacterium, *Oleispira antarctica*. The Cpn10 and Cpn60 chaperonins from *O. antarctica* have 74% and 54% amino acid identity to the *E. coli* GroEL and GroES chaperonins, respectively, and show high protein refolding activities at temperatures of 4-12°C. When expressed in ArcticExpress cells, these chaperonins confer improved protein processing at lower temperatures, potentially increasing the yield of active, soluble recombinant protein.

E. coli BL-21 (DE3) Gold

Transformed competent cells were cultivated in LB medium containing tetracycline and kanamycine as a selection marker in Erlenmeyer flasks at 37°C until the optical density of 0.6-0.8 (measured as absorbance at 550 nm), then the production was induced by adding 0.1 mM IPTG. After 4-8 h of production the bacteria were harvested by centrifugation 5000 x g/15 min.

ArcticExpress BL-21 (DE3) E. coli system

Transformants were cultivated in LB medium containing streptomycine and kanamycine as a selection marker in Erlenmeyer flask at 37°C until optical density of 0.6-0.8 (measured as absorbance at 550 nm). The bacterial culture was cold to 12°C and induced by adding 0.1mM IPTG. After 4-12 h of production the bacteria were harvested. The small amount of bacterial suspense was disrupted and analyzed for enzymatic activity. Unfortunately, enzymatic activity was not measured. The protein was produced in form of inclusion bodies. Inclusion bodies were found in *E. coli* BL-21 (DE3) Gold expression system, too.

The inclusion bodies were isolated by several freeze - thaw by 5 sonication cycles in 25% sucrose buffer. Bacterial suspense was centrifugated 5000 x g and the pelet was resuspended in 0.1% Triton X-100 and centrifuged. The last step of isolation of inclusion bodies was washing in 50 mM Tris, 150 mM NaCl, pH 7.4.

3.2.1.1 Protein refolding *in vitro*

α-NAGA produced in *E. coli,* was produced in form of inclusion bodies. For these reasons was carried out renaturation *in vitro.* The enzyme was renaturated using two methods, rapid dilution and dialysis.

Rapid dilution

Isolated inclusion bodies were dissolved in denaturating buffer (6 M guanidinium hydrochloride pH 8.0 or 8 M urea both with addition of 100 mM DTT). The refolding was done by adding dissolved inclusion bodies to appropriate pre-chilled refolding buffer. The refolding buffer consisted arginine as a low molecular chaperon, cysteamine/cystamine or *ox/red* glutathione as a redox pairs and ihibitors of proteases as PMFS, leupeptin, pepstatin. The concentration of arginine and redox pairs was changed in developing of renaturating protocol.

Dialysis

Isolated inclusion bodies were solubilized in the same denaturating buffers as in rapid dilution methods. The solution was cleared from insoluble material by centrifugation at 30,000 x g for 30 min. The mix was dialyzed twice at 4°C against 1 L of renaturating buffer.

For finding of optimal conditions of renaturating of α-NAGA the i-Fold refolding kit (Merck) was used, too.

3.2.2 Recombinant expression in yeast expression system

For expression of α-NAGA two expression systems *Pichia pastoris* and *Saccharomyces cerevisiae* was used.

Pichia pastoris

Using a single colony 25 ml of BMGY in a 250 ml baffled flask was inoculated. Cell suspense was cultivated at 28°C in a shaking incubator (300 rpm) until culture reaches an OD_{600} 4-6 (~16-18 hours). The cells will be in log-phase growth. Cells were harvested by centrifuging at 3000 x g for 5 minutes at room temperature. Supernatant was decanted and pellet was resuspended to an OD_{600} of 1.0 in BMMY medium to induce expression. Culture was placed in 1 liter baffled flask. 100% methanol to a final concentration of 0.5% was added every 24 hours to maintain induction (Pichia expression kit, Manual, Version M, Invitrogen).

The concentration of methanol as an inductor, time of cultivation, volume of cultivation media and the shape of cultivation flasks were changed for getting the recombinant enzyme.

Saccharomyces cerevisiae

Precultures were prepared by inoculating 15 ml of SC selective medium containing 2% glucose, tryptophan, leucine, adenine and histidine, without uracil with a single colony of S. *cerevisiae* W303-1A containing the pYES-2CT/ α-NAGA construct. After overnight growth at 30˚C and 220 rpm, the pellets were collected by centrifugation, resuspended in 200 ml of SC selective medium as above but with the 2% glucose replaced by 2% galactose, and the culture was incubated at 30˚C and 220 rpm. The cell suspenze was harvested by centrifugation. After centrifugation, the yeast pellet was resuspended in lysis buffer composed of 50 mM citrate phosphate (pH 3.5), 5% glycerol, 1 mM PMSF and 1mM dodecyl maltoside. An equal volume of acid-washed glass beads (r = 0.25-0.5 mm, Pierce, DE) was added. The cells were disrupted by vortexing for 30 s, followed by 30 s on ice (repeated four times for complete cell lysis). The mixture was centrifuged and the supernatant transferred into a fresh microcentrifuge tube for the enzyme activity assay.

3.2.2.1 Enzyme activity assay

α-NAGA and α-GA activities were assayed using *o*-nitrophenyl *N*-acetyl-α-D-galactosaminide (*o*-NP-α-GalNAc) and *p*-nitrophenyl α-D-galactosidase (*p*-NP-α-Gal) as a substrate, respectively. One unit of enzymatic activity was defined as the amount of the enzyme releasing 1 μmol of *o*-nitrophenol per minute at 50 mM citrate-phosphate buffer at pH 3.0 and 35°C. After incubation of the reaction mixture at 35°C for 10 min, liberated *o*-nitrophenol was determined spectrophotometrically at 420 nm under alkaline conditions (50 μl of the reaction mixture was added to 1 ml of 0.1 M Na_2CO_3); 0-2 mM *o*-nitrophenol was used for calibration. □-GA activity was assayed as described above, *p*-nitrophenyl-α-D-galactopyranoside was used as a substrate (*p*-NP-α-Gal, Sigma, Czech Republic).

Kinetic values were determined as described above using *o*-NP-α-GalNAc within the concentration 0.5-5.0 mM. All data were measured in triplicates. The effect of pH on α-NAGA activity was measured at 35°C as described in the enzyme assay, 50 mM HCl–glycine buffer was used for pH 1-3 and 50 mM citrate-phosphate buffer for pH 3-7. Every assay of the enzyme at different pH values was paralleled with appropriate blank void of enzyme. No differences of the

solubility of the *o*-NP-α-GalNAc (final concentration 2 mM) with respect to pH values were observed. All data were measured in triplicates.

Protein concentration was determined according to Bradford using a Bio-Rad Protein Assay kit, with bovine serum albumine as the standard.

3.2.3 Purification of the recombinant enzyme

A BioSys HPLC System with a UV and conductivity detector (Beckman-Coulter) was used for all chromatography steps. The purification of the enzyme stared by HIC chromatography on the column Phenyl-Sepharose HR column (2.6 × 10.6 cm, Merck), followed by cations exchange chromatography S-Sepharose FF column (1.6 × 12.5 cm, Merck) and as the last chromatography Superdex 200 10/300 GL gel filtration column (1.0 × 30 cm, Amersham Bioscience) was used.

The best yield of the active enzymes was determined by the combinations of purification steps. Firstly, the chromatofocusing column (Mono P 5/200 GL Amersham Bioscience) instead of Superdex 200 10/300 GL gel filtration column was used. It was observed that the change of this step provide the same yield of recombinant enzyme and the buffer for gel filtration column is cheaper then the buffer for chromatofocusing.

3.3 Sedimentation velocity and sedimentation equilibrium measurements

Sedimentation analysis was performed using a ProteomeLabXL-I analytical ultracentrifuge equipped with an An50Ti rotor (BeckmanCoulter). The protein (0.4 mg/ml) was dialyzed against 50 mM sodium-citrate buffer (pH 3.5, used also as a reference and sample dilution buffer). The sedimentation velocity experiment was carried out at 40,000 rpm and 20°C, absorbance scans were recorded at 280 nm in 5 min intervals with 30 μm spatial resolution. Buffer density and α-NAGA partial specific volume were estimated in SEDNTERP 1.09 (www.jphilo.mailway.com). Data were analyzed with SEDFIT 12.1 (Schutz, 2000). The sedimentation equilibrium experiment was performed with α-NAGA concentration of 0.13 mg/ml at 10-12-14-16-18-20-22,000 rpm at 4°C. Absorbance data were collected at 280 nm by averaging 20 scans with 10 μm spatial resolution after 30 h (first scan) or 18 h (consecutive scans) of achieving equilibrium and were globally analyzed with SEDPHAT 8.2 using a non-interacting discrete species model (Schutz, 2003).

3.4 Electron microscopy

Negatively stained samples of wild-type and cloned enzyme complexes were prepared in parallel as follows: drops of enzyme solution in 50 mM citrate-phosphate buffer (pH 3.5), diluted to approx. 50 μg/ml were applied onto glow discharge-activated carbon coated grids. After adsorption for 30 s, the grids were negatively stained with unbuffered 2% (wt/vol) uranyl acetate or unbuffered 0.5% (wt/vol) uranyl formate in water, respectively. The samples were analyzed in a Philips CM100 electron microscope (FEI, formerly Philips PEO) at 80 kV. Digital images were recorded using a MegaView II slow-scan camera (Olympus, formerly SiS GmbH) at a primary magnification of 64,000, giving a pixel size of about 1 nm. All image processing was done using AnalySis 3.2 software.

3.5 Mass spectrometry

Sample preparation for amino acids sequencing

For protein identification mass spectrometry and N-terminal sequencing were used. MALDI TOF MS was used for sequence mapping of tryptic digests. The Coomassie Blue-stained protein band was excised from SDS-PAGE gel and washed with 0.1 M 4-ethylmorpholine acetate pH 8.3 buffer containing 50% acetonitrile several times. After complete desalting, the protein was reduced by tris[(2-carboxyethyl)phosphine hydrochloride] at 75˚C for 30 min. After reduction the alkylation with iodacetamide at 25˚C for 60 min in dark was carried out. The gel pieces were further washed with acetonitrile and rehydrated with water. This step was repeated five times. The gel pieces were dried in vacuum, rehydrated in 50 mM 4-ethylmorpholine acetate cleavage buffer containing 10% acetonitrile and 1 μl of trypsin (5 ng/ml) was added. Positive ion mass spectra were measured on a Bruker ULTRAFLEX III TOF mass spectrometer. The spectra of the peptides were measured in reflecton mode by using α-cyano-4-hydroxy-cinnamic acid in aqueous 40% acetonitrile containing 0.2% FA (10 mg/ml) as a MALDI matrix.

For N-terminal sequencing, α-NAGA was transferred from 10% SDS-PAGE gel onto PVDF membrane and analysed by automated cycles of Edman degradation in a Procise 491 Protein Sequencer (Applied Biosystems).

Analysis of the N-glycosylation sites

α-NAGA was resolved on 12% SDS polyacrylamide gel electrophoresis under reducing

conditions. The Coomassie blue-stained protein band was cut and treated as described above. The gel pieces were dried in vacuum and rehydrated in the kit deglycosylating buffer; for deglycosylation of α-NAGA PNG-ase F PO704S Kit and Endo Hf Kit (both New England BioLabs) were used for 12 h at 37˚C. After deglycosylation the gel pieces were prepared for digesting by trypsine, chymotrypsine, GluC and AspN, proteinases (sequencing grade, Roche). The procedures of enzymatic digestion were described previously (Ma et. al. 2003). Positive ion mass spectra were measured on a Bruker ULTRAFLEX III TOF mass spectrometer.

N-glycosylation sites were determined by LC/ESI/FTMS. Peptides were separated by reverse phased μHPLC system (Agilent Technologies) equipped with a MAGIC C18 column (0.2 × 150 mm, 5 μm) (Michrom Bioresources). Peptides were eluted at flow rate of 4 μl/min under gradient conditions. The column was directly connected to the mass spectrometer. Mass spectra were acquired on APEX-Qe FTMS instrument equipped with a 9.4 T superconducting magnet (Bruker Daltonic).

4. Results

4.1 The α-galactosidase type A gene *aglA* from *Apergillus niger* encodes a fully functional α-NAGA

Among the glycosidases necessary for the synthesis of valuable mimetics for NK cell receptors are also α-GA and, in particular α-NAGA. This may appear at the first site surprising since the primary specificity of NK cell receptors is towards the β-linked HexNAc. However, both our recent work (Veprek et. al. 2006) and the complete analysis of the binding specificity using oligosaccharide arrays (not published) indicate that these receptors bind also α-linked HexNAc with nearly equal affinity. Therefore, in collaboration with the Laboratory of Biotransformations at the Institute of Microbiology v.v.i. we have initiated search for fungal α-GA and α-NAGA.

As a starting point we have used the genome data available from the recently sequenced genome of *Aspergillus niger.* α-NAGA are quite rare in nature, but they are known to be evolutionarily related to α-galactosidases that have been much more widely studied. A BLAST search for α-GA primary sequences within the *A. niger* genome revealed five distinct protein-coding genes. Apart from genes *AglA* and *AglB* that are well characterized, three additional sequences with sequence identity higher than 33% could be found. These three sequences have not been characterized well but have hypothetical open reading frames similar to *aglB* with length varying from 391 to 431 amino acids. From the point of view of evolutionary relation with other enzymes it was interesting to note that while *aglB* has more than 70% identity with many α-GA from GH 27 family, *aglA* had nearly similar identity (about 64%) to fungal α-GA as it had to α-NAGA from *Acremonium sp. 413.* This relationship opened up a possibility that this gene in fact encodes α-NAGA (these enzymes are often able to cleave α-GA as well because of the smaller size of Gal then GalNAc).

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 ability of this strain to hydrolyze α-linked GalNAc was ten times higher than its ability to hydrolyze α-linked Gal. Additional biochemical investigations revealed the large native molecular size of this enzyme (440 kDa) and its pI (4.8). The Km for *o*-NP-α-GalNAc was found to be 0.73 mM, and the optimum activity was achieved at pH 1.8 and 55°C. The enzyme belongs to retaining glycosidases as proven by NMR investigations (Weignerova et. al. 2008). 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. Both spots were analyzed by N-terminal sequencing and were shown to differ only in three N-terminal amino acids including one glutamate that was able to shift this isoform to more acid pI (Figure12A).

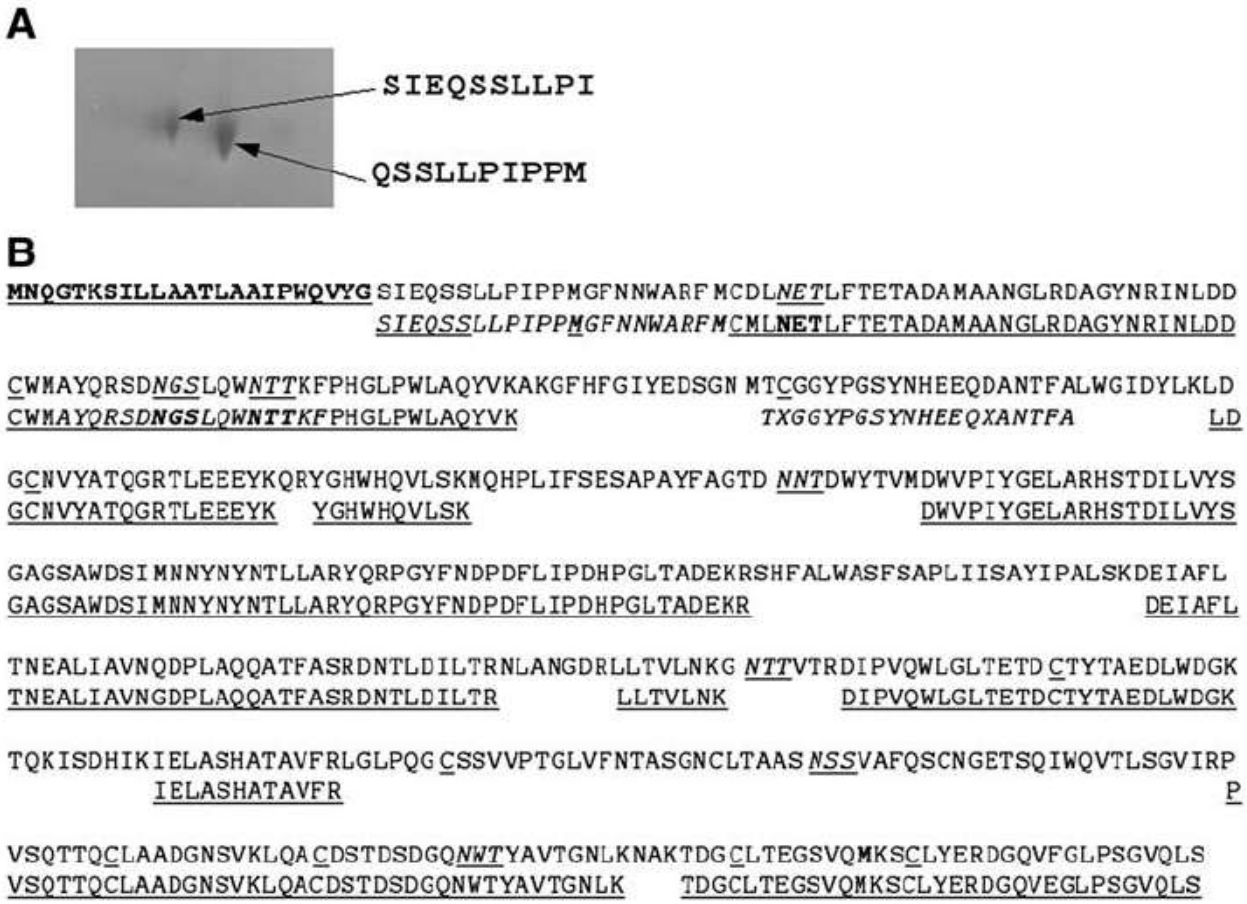


Figure 12. Summary of sequencing data for enzyme acting as α-NAGA. (A) Separation of prepared enzymes differing in their indicated N-terminal sequences by 2D electrophoresis, indicating minor heterogeneity in enzyme preparation confirmed by N-terminal sequencing. (B) Upper lane indicates amino acid sequence 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.

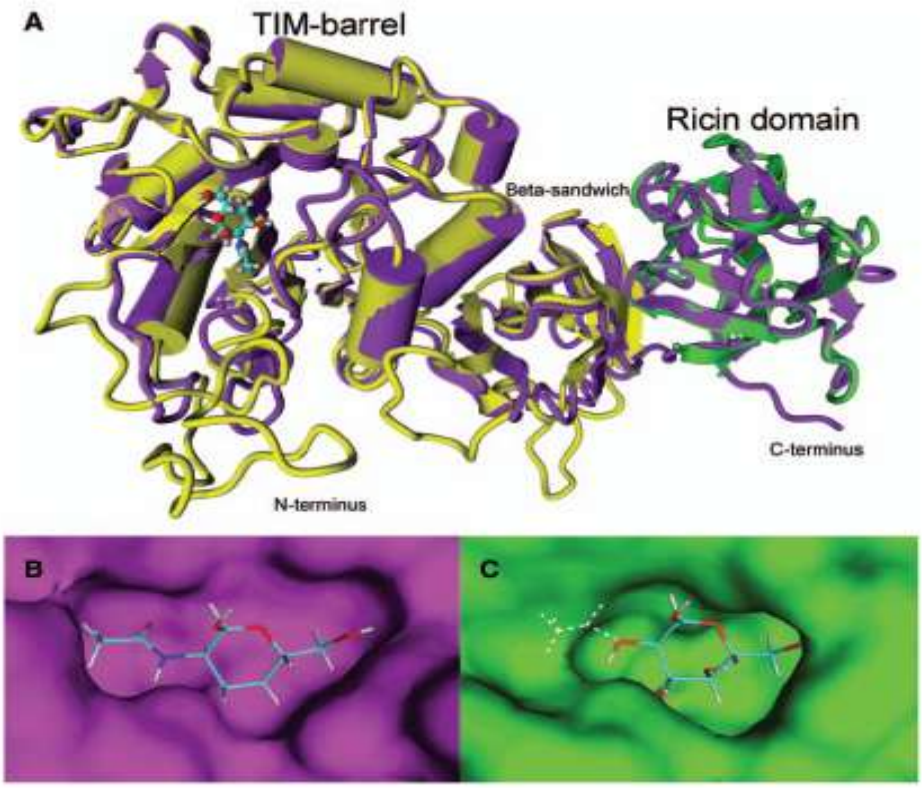


Figure 13. Structure of the α-NAGA from *Aspergillus niger*. (A) Overall fold shows a TIM-barrel with the active site at the N-terminus, a small domain of eight antiparallel β-strands packed in β-sandwich in the middle, and a ricin-like domain on the right. The generated model (magenta) is overlaid with the crystal structure of the homologs α-NAGA from chicken (yellow), with GalNAc 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.

Alignment of the found *AglA* sequence against structurally solved α-NAGA revealed high sequence identity allowing to construct a molecular model of the enzyme (Figure 13). Comparison of experimentally solved three dimensional structures with our molecular model revealed a remarkable similarity spaning over evolutionary unrelated species. This finding points to the fact that all α-NAGA are very similar independenty of their source with regard to both domain arrangement as well as to the position of individual catalytical amino acid residues. This fact allows to use the model not only to view the general domain architecture, but even to consider the

mechanistic detail sof the catalytical process. 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. Considering the high evolutionary conservation of these family of enzymes we could also construct the molecular model of *AglB*. Because both these models were of a very high duality as discussed above, we could perform substrate docking into their active sites (Figure 13B and C).

Docking of *o*-NP-GalNAc occurred rapidly a remained stable in 2 ns, although the entire simulation was run for 10 ns. On the other hand, an attempt to dock the same substrate into the active site of *AglB* led to rapid repulsion, although stable docking in 2 ns was possible for *o*-NP-Gal. This confirms that *AglB* is a pure α-GA not able to cleave α-GalNAc.

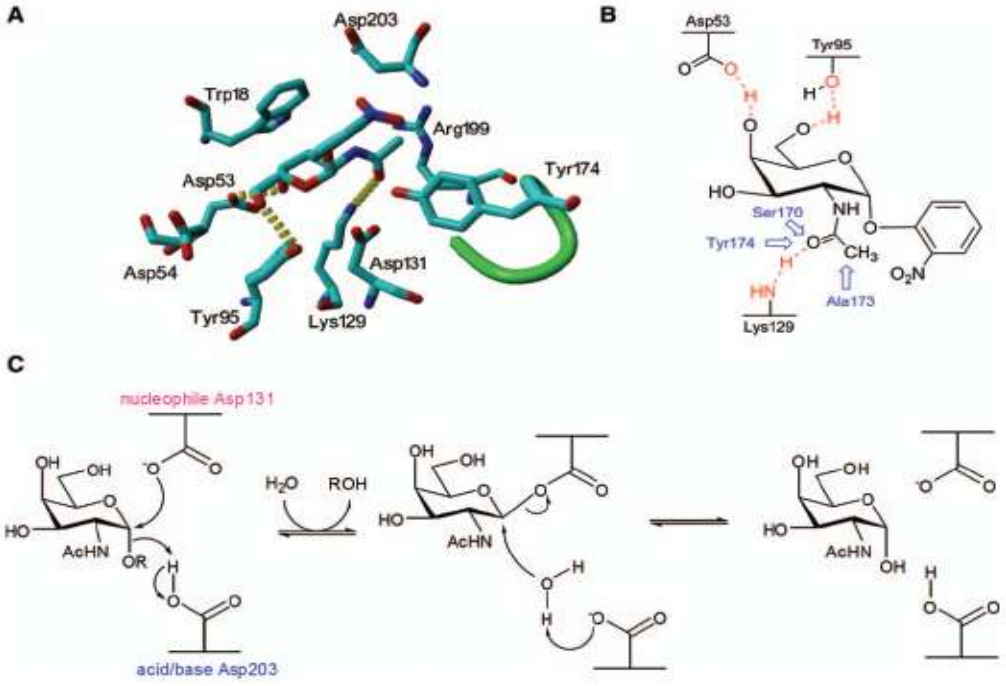


Figure 14. (A) Active site amino acids identified for *aglA* enzyme with bound 2-NP-α-GalNAc after 10 ns of MD with so-called *N*-acetyl recognition loop (green), which extends the binding pocket of *aglA* enzyme so that it can accept the amino group at the C2-atom. (B) Scheme of hydrogen bonds, created by *o*-NP-α-GalNAc with *aglA* enzyme. (C) Scheme of catalytic mechanism proposed for *aglA* enzyme, where Asp203 acts as acid/base and Asp131 is responsible for nucleophilic attack at C1.

The above substrate docking experiments could repeal further mechanistic details about the catalytic process including amino acids involved in substrate binding and hydrolysis. The overall size of the substrate binding pocket is usually significantly larger in α-NAGA in order to allow to accomodate much larger *N*-acetyl group. Accordingly, small side chain amino acids would be expected to occur around the acetyle binding site. Indeed, in the case of *AglA* these amino acid were Ser170 and Ala173. In addition, there are some amino acids establishing positive interactions with GalNAc but not with Gal allowing to undertand the much higher afinity for the formel substrate. In this case such an interaction occured with Lys129 that forms a hydrogen bond with carbonyle oxygen in GalNAc, and additional Tyr174. Other amino acids are responsible for the specificity of the enzyme for *galacto* configuration: Asp53 forms a hydrogen bridge to C4 hydroxyle defining this configuration, and Tyr95 bind the sugar unit. Finally, two aspartates represent the catalytic residues directly involved in hydrolysis of the glycosidic bonds in the way that is typical for retaining mechanism: acid/base Asp203 and nucleophile Asp131. Multiple sequence alignment confirmed the general validity of these findings.

4.2 Facile production of *Aspergillus niger* α-NAGA in yeast

The development of an appropriate system for expression of recombinant α-NAGA appeared critical for the preparation of sufficient amount of material necessary for the detailed biochemical and structural studies of the enzyme, and was essential in order to initiate structure-activity investigations using molecular mutagenesis as well as studies of the role(s) of enzyme glycosylation. These tasks proceeded in several stages: (a) cloning and sequencing of α-NAGA from the producing strain of *A. niger*, (b) attempts to express active enzyme in bacteria (*E .coli*), (c) attempts to express active enzyme in the yeast *Pichia pastoris*, (d) attempts to express active enzyme in the yeast *Saccharomyces cerevisiae.*

The gene fragment coding for α-NAGA was obtained from total RNA that was isolated from cultures of *A. niger,* in which the production of an enzyme was induced by soya flour (Weignerova et. al. 2008). Degenerate oligonucleotide primers were designed based on the sequenced native α-NAGA (see the previous chapter) using sequences specific for *AglA*, and contained restriction sites for NdeI and HindIII, respectively, used for all subsequent manipulations. RT-PCR yielded a single DNA fragment having approximately 1500 bp containing the entire open reading frame of the enzyme. This fragment was cloned into Bluescript II SK+ vector, sequenced, and the DNA sequence was identical to the one corresponding to *AglA* gene.

The cloned fragment was transferred to bacterial expression vector pET-30a^{+} assuring the presence of the initiation methionine created at the NdeI site. The double stop codon was placed at the HindIII site to assure efficient termination of translation. The production was tested in several bacterial strains using different conditions. Conditions optimal for the production of the target protein were established (0.1 mM IPTG, 37°C, 4 h). However, all produced protein precipitated into inclusion bodies. Iclusion bodies were isolated, dissolved in denaturating reagens, and numerous experiments were performed in order to find optima conditions for *in vitro* refolding. However, all attempts resulted in precipitation of the enzyme protein, and no soluble protein was obtained. Further attempts to use commercial iFOLD kits did not provide successfull results, either.

Next we attempted to produce the active enzyme in a more complex, eukaryotic system based on a methanotrophic yeast *Pichia pastoris.* DNA fragment was cloned into the comercial pPICZα from Invitrogen. The DNA construct was verified by sequencing, transferred into the production strain by electroporation, and the mixture was plated onto the selection plate with Zeocin. 100 colonies were selected, produced on liquid medium, induced, and enzymatic activity was measured both in the supernatant and inside the broken cells. However, no measurable activity could be obtained in any of the tested sample indicating the lack of active enzyme. Many additonal

optimization tests were performed with regard to methanol concentration used for induction, volume of culture, the shape of culture flaks, temperature of production etc. However, none of these optimization provided any condition for the production of active enzyme.

Since the literature search data indicated examples of successfull production of enzymatically active glycosidases in *Saccharomyces cerevisiae* (strain W303-1A) (Ashida et. al. 2000) we finally attempted to produce α-NAGA using this expression system. DNA fragment coding the enzyme was transferred to pYES-2CT expression vector, and the construct was again verified by DNA sequencing. This strain was transformed by electroporation, and the mixture was plated onto SC plates based on uracil complementation. Positive colonies were selected, and rapid screening was performed in 5 ml of medium withdrawing aliquots after 24 h. Aliquots were spun, and activity was assayed in supernatant and in broken cells. No activity could be detected in the medium, but we could observe a significant activity in the broken cells. Extensive optimization of production conditions (volume of medium, volume and shape of culture flask, concentrations of inducers (Gal), production times and cultivation temperatures. The highest production of intracellular α-NAGA (approximately 0.4 U/mg protein) was achieved in 2 L cultivation flasks containing 200 ml of medium at 30°C with 2% galactose and 72 h of cultivation.

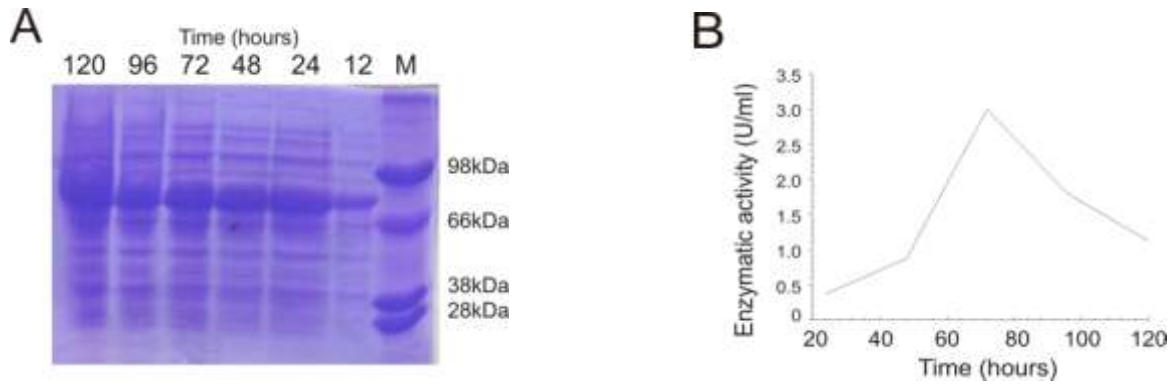


Figure 15. Time profile of the recombinant intracellular α-NAGA production by *S. cerevisiae* W303: (A) SDS-PAGE electrophoresis (B**)** enzyme activity. Time optimization was carried out at 30˚C. Aliquots of the cell culture were harvested 12, 24, 48, 72, 96 and 120 hours after transfer to the SC medium with Gal 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.

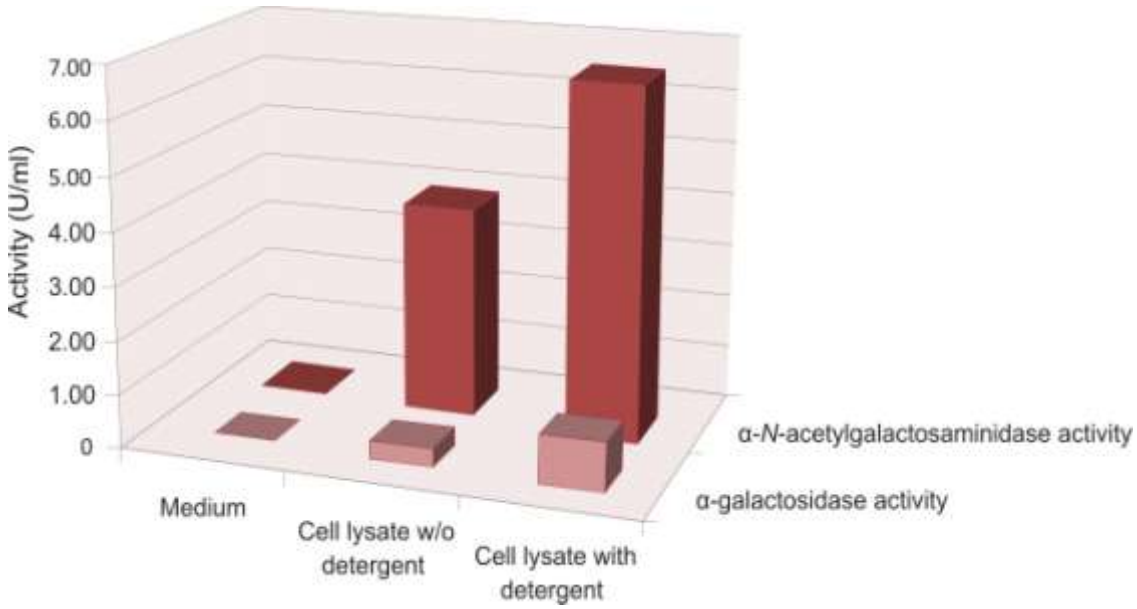


Figure 16: α-NAGA and α-GA activity measured in cell lysates with and without dodecyl maltoside (detergent) in lysis buffer.

The enzyme purification was achieved with good recovery (Table 2). The enzyme purification was carried out by four-step chromatography. The majority of the contaminating proteins were removed by hydrophobic chromatography (Phenyl-Sepharose HR). The subsequent purification steps were used to remove the other proteins with similar biochemical properties. The enzyme was purified with a final yield of 12.1%. The specific activity of the recombinant enzyme against *o*NP-α-GalNAc was determined to be 42.3 U/mg.

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

SDS-PAGE of the purified α-NAGA revealed one single protein band with an estimated molecular mass of 76 kDa. This value is higher than the molecular mass calculated from the amino acids sequence (53961.2 Da) because of extensive protein glycosylation. The enzymatic properties of α-NAGA were assayed with *o*-NP-α-GalNAc as the substrate. The recombinant α-NAGA exhibited a K_M value of 0.56 mmol/l and pI 4.4 in 50 mM citrate-phosphate buffer (pH 3.5) at 37˚C. The purified enzyme has a pH optimum at 2.0-2.4 (at 37˚C) and temperature optimum of 50-55˚C (at pH 3.0). The recombinant enzyme was stable in 50 mM citrate-phosphate buffer within the pH range 2-4 and at 4˚C for several weeks without any significant loss of activity. A loss of activity (35%) was observed after 2 days of incubation at 37˚C. Its enzymatic properties are consistent with the wild extracellular α-NAGA isolated from *A. niger* CCIM K2 (Weignerova et. al. 2008).

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 group A transformation in to the group of H(0), beeing a universal donor.

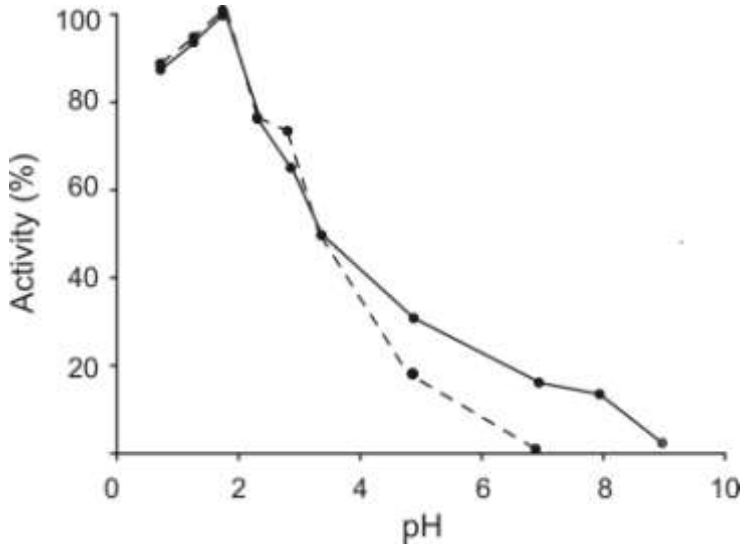


Figure 17. Effect of pH on the activity of wild α-NAGA (- -●- -) and recombinant α-NAGA (—●—).

The molecular weight of the native and recombinant α-NAGA was investigated using gel filtration and analytical ultracentrifuge. 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 monomers. The analytical centrifugation confirmed the molecular mass of both enzymes. All experiments were performed at four various proteins concetrations, at 4°C, and the data was analyzed using the software SEDFIT. The sedimentation coefficients were

determined of 7.65 and 5.07 in causes of the recombinant α-NAGA. These values corresponded well with the mass of the dimer and the monomeric form, respectively. Moreover, both monomeric and the dimeric enzyme forms displayed both α-NAGA and α-GA activity in same ratio.

Finally, the occurrence of the monomer and the dimer species and their ratio remained constant under various protein concentrations. Both sedimentation velocity and equilibrium experiments were performed with the protein at 2/3, 1/3 and 1/6 of the initial concentration. There is no evidence for equilibrium between the two protein forms (Figure 18).

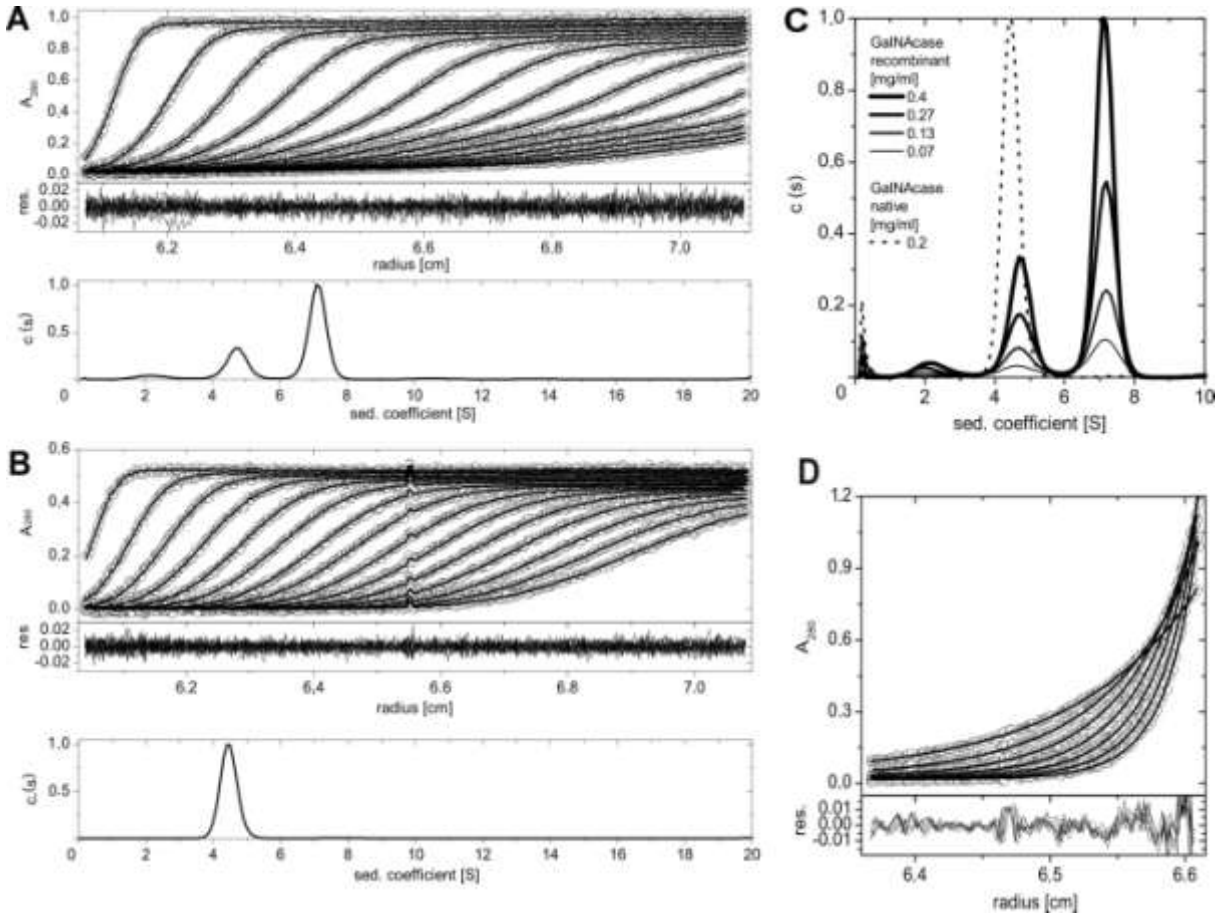


Figure 18: Sedimentation velocity experiments. The recombinant α-NAGA (A), and native α-NAGA (B), were analyzed in an analytical ultracentrifuge using sedimentation velocity. Fitted data (upper panel) with residual plots (middle panel) showing the accuracy of the fit are shown together with the calculated continuous size distribution c(s) of the sedimenting species (lower panel) at various α-NAGA loading concentrations. (C**)** The sedimentation velocity experiments showed that there is no apparent equilibrium between the two α-NAGA forms, since the decrease in both forms corresponds solely to the decrease in enzyme loading concentration. (D) Equilibrium sedimentation distribution of recombinant α-NAGA. Upper panel shows absorbance data with fitted curves (non-interacting discrete species model, lines), lower panel shows residuals derived from fitted data.

Electron microscopy technique was used for visualization of the wild-type and recombinant α-NAGA. Uranyl acetate and uranyl formate was used for staining. The wild-type α-NAGA sample showed a distribution of small particles with a mean diameter of 6-7 nm and the shapes ranging from almost isomeric to elongated. The recombinant α-NAGA showed particles with broader size distribution, with a mean diameter of 7-10 nm. This data are in good agreement with the sedimentation velocity and gel filtration experiments and confirmed that recombinant α-NAGA is present in two active forms.

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 GlcNAc 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 5b. 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. For example of determination of the *N*-glycosylation sites (Figure 19).

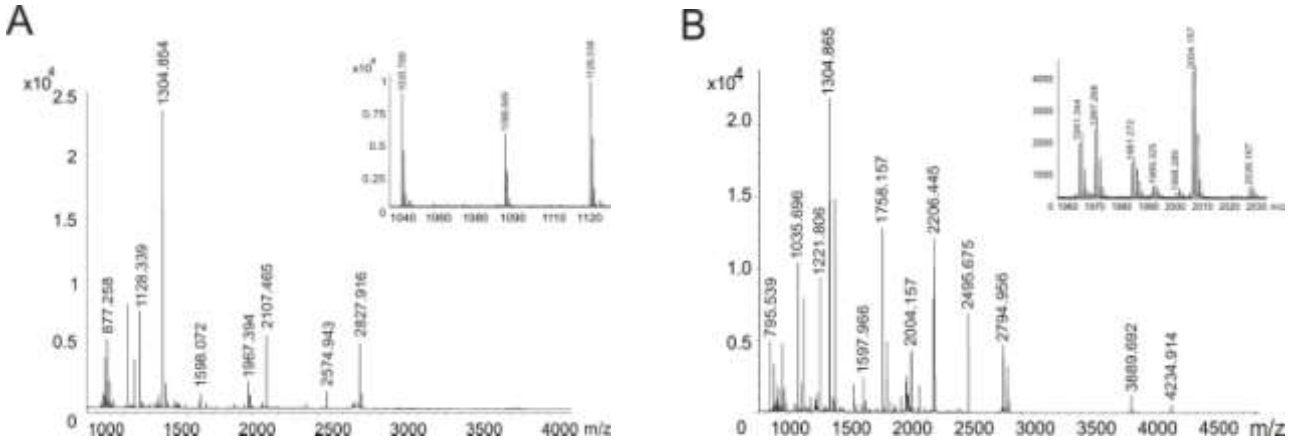


Figure 19. Determination of *N*-glycosylation sites. (A) α-NAGA was digested with Asp-N and treated with PNGase F. MALDI-TOF mass spectrometry analysis showed that the high mass peaks had disappeared, while a new peptide signal at *m/z* 1128.337 appeared. This peak corresponded to the peptide residue 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 *S. cerevisiae*. The notable advantage of our expression system is in shorter production times, and, up to four fold 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.

4.3 Patent application

Czech patent application was preparated on the basic of results with expression of α-NAGA in yeast expression system *Saccharomyces cerevisiae*.

The name of patent application is: Expression of active form of α-NAGA from filamentous fungi *Aspergillus niger* and its expression in yeast expression system *Saccharomyces cerevisiae.*

This application refers to α-NAGA enzyme with α-GA activity and preparation of active form of this enzyme in yeast expression system *Saccharomyces cerevisiae*.

The patent claims were divided into four points: the first point refers recombinantly prepared active form of α-NAGA with dual α-GA activity in yeast expression system *Saccharomyces cerevisiae*. The second point of the patent claim refers of isolated molecule of nucleic acid coding amino acids sequence of active form of α-NAGA. The third one was claimed method of preparation of the nucleic acids and primers (isolation of RNA, transcription of RNA to cDNA, and PCR amplification with prepared cDNA and designed primers). Production, isolation and purification of α-NAGA were forming the last claim of this patent application.

The expression and charactezation of α-NAGA has been described above. α-NAGA has approximately 20% of enzymatic activity in neutral pH. This fact can be used for the red blood type A transformation in to type H(0), being a universal donor. The next usage can be in preparing of blood derivates and synthesis of oligo/poly carbohydrate strucutre. All of these applications can be used in pharmacetutic industry. Due to dual activity of α-NAGA we can use this enzyme as a model protein for preparing of α-GA with enzymatic activity in lower pH.

4.4 Cooperation between Subunits Is Essential for High-Affinity Binding of *N*-Acetyl-D-hexosamines to Dimeric Soluble and Dimeric Cellular Forms of Human CD69

CD69 was originally described in 1986 as EA 1 on the surface of human T cells after 12-o-tetradecanoyl phorbol-13-acetate treatment (Hara et. al. 1986) and later as AIM (activation inducer molecule) (Cebrian et. al. 1988). CD69 is an important member of C-type animal lectin family closely related to molecules NKR-P1 and Ly49. CD69 was observed on other immune system cells (Testi et. al. 1994) and after cloning (Hamann et. al. 1993) it was recognized as C-type lectin. The CD69 gene is located within the NK gene complex on human chromosome 12. It codes a type II calcium-dependent membrane lectin. The identification of the natural ligand for CD69 is a key critical step for further advancement of our knowledge on the biology of this receptor. The initial findings that CD69 binds to calcium and certain HexNAc could not be later reproduced using a somewhat different expression construct. Since then, these discrepancies have been at least partially explained by careful structural evaluations of the recombinant proteins used for binding studies, as well as by establishing a direct link between the binding of calcium and carbohydrates. Themost recent development of efficient structural mimetics of the high-affinity ligand for CD69 opened the way for manipulating with numerous activities of CD69 at the molecular and cellular level and provided efficient compounds for further in vivo testing of their immunomodulating properties.

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 and 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. The homogeneity and monodispersity were routinely evaluated for each batch of the produced soluble dimeric CD69 using SDS electrophoresis under both reducing and nonreducing conditions and gel filtration. Moreover, the identity, quality, and proper refolding of each batch of the produced protein were also verified using high-resolution ion cyclotron resonance mass spectrometry, one-dimensional proton NMR, thermal stability experiments, and tests of the biochemical stability (Table 3).

Protein	**Characteristics**	**Td[a]**	**Td[b]**	**Td[c]**
CD6CD69WT	noncovalent dimers	65	67	65
CD69Q93A	dimer/monomer equilibrium	63	62	64
CD69R134A	dimer/monomer equilibrium	62	60	60
CD69RDM	monomeric	60	62	61

Table 3. Summary of Stability Properties of Wild-Type Dimeric CD69 and CD69 Dimerization Mutants. (a) Determined from thermal UV denaturation measurements. (b) Determined from differential scanning calorimetry. (c) Determined from FTIR spectroscopy.

The D-GlcNAc as the high affinity carbohydrate ligand, together with D-ManNAc were performed for using of detailed binding studies with soluble dimeric CD69. D-Gal was used as negative controls in some experiments. The initial evidence for the interaction of the soluble dimeric CD69 with GlcNAc was obtained by NMR titration. The results of this experiment confirmed the specific binding of GlcNAc to the dimeric receptor and provided affinity estimation in the low micromolar range (Kd $= 4.0 \times 10^{-7}$). On the other hand, no interaction could be seen with ManNAc under the same experimental conditions. The results from NMR titration were confirmed by direct binding experiments. When the bound and unbound ligands had been separated by dialysis under equilibrium. These experiments revealed two binding sites per receptor dimer. The direct binding experiments enable the degree of saturation at each particular ligand concentration to be calculated. The resulting saturation curve clearly revealed a striking cooperativity in the highly specific (Kd = 4.0×10^{-7} M) binding of GlcNAc to the receptor (Figure 20C). These results were also confirmed by the flourescent titration (Figure 20D).

.

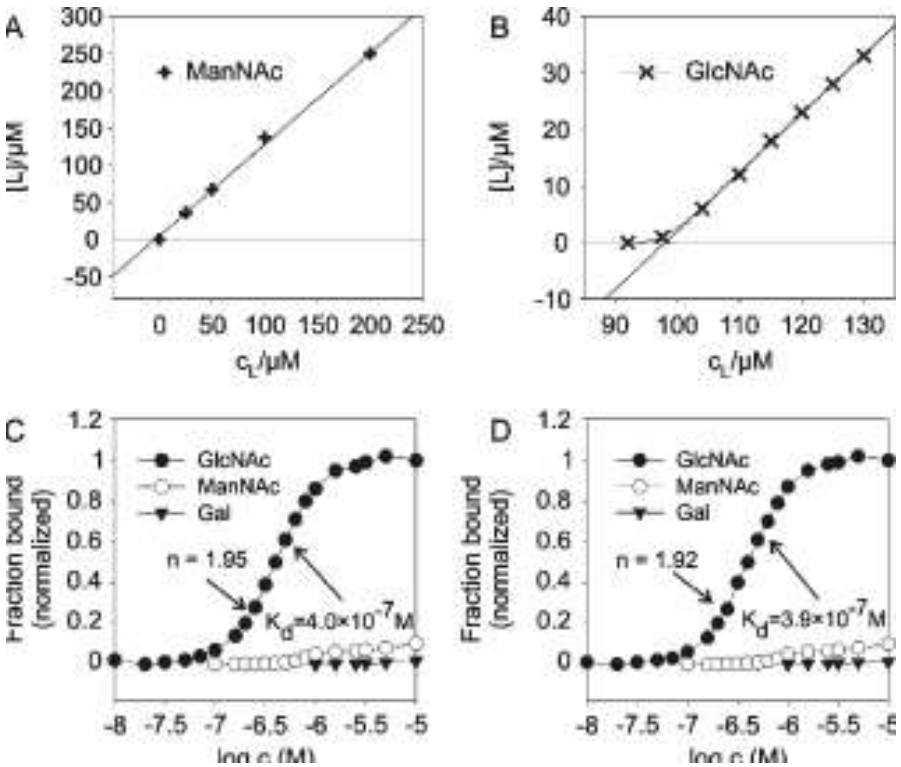


Figure 20. Measurements of direct interaction of soluble noncovalent dimer CD69 with ManNAc and GlcNAc. (A, B) NMR titration of soluble CD69 with ManNAc and GlcNAc, respectively. (C, D) Concentration dependence of receptor saturation measured by equilibrium dialysis and tryptophan fluorescence quenching, respectively, using GlcNAc, ManNAc, and Gal as indicated.

The structure changes of soluble CD69 upon ligand binding were investigated by variations in the hydrodynamic properties of the receptors. These variations were studied by gel filtration (Superdex 200HR) and analytical ultracentrifugation and compared with the size and structure of the receptors preincubated with ManNAc as a control. This analysis did not revealed any changes in hydrodynamic properties after binding of carbohydrate unit into receptor.

In the next step, the interactions of GlcNAc with monomeric subunit of CD69 were studied. The crystale structure of CD69 dimer was used for construction of CD69 monomeric form. Firstly, were analyzed the dimer interface for critical residues participating in the dimerization. Two such residues Q93 and R134, both interacting with residues of the other subunit, could be found. These residues were mutated to alanine, singly or in combination. The stability of this protein was comparable to that of the wild type protein. This indicated that the mutations did not results in any decrease of protein stability. This protein was used for investigation of binding affinity with GlcNAc too. The results showed that binding to the monomeric subunit of CD69 was much weaker and noncooperative (Figure 21).

.

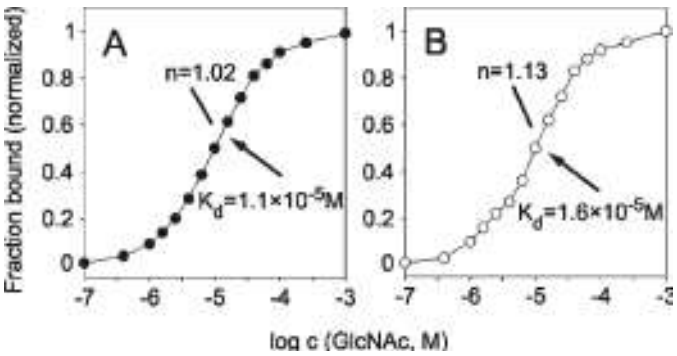


Figure 21. Binding of GlcNAc to the monomeric subunit of CD69. (A, B) Binding of GlcNAc to monomeric CD69 analyzed using equilibrium dialysis and fluorescence titration, respectively.

The construct for non-covalently dimer receptor contained an extended dimer interface involved in contacts between the ligand binding domains, as well as the neck regions.However, it did not contain the C68 residue that participates in the covalent dimerization of the natural form of CD69 found at the surface of leukocytes. In order to trigger the CD69 is required an efficient ligand. Interaction of receptors with this ligands leads to receptor cross-linking. Alternatively, we can used specific antibodies for aggregation. Both forms of activation were used in this experiments. It was used two specific antibodies against CD69 as well as the HexNAc disaccharide dimer, which has been previously described as the most efficient carbohydrate ligand. CD69 gene of interest was cloned into vector suitable eukaryotic expression vector. Jurkat T-lymphocyte leukemic cell line was transfected with this vector. These clones displaying identical surface expression of mutated and wild-type forms of receptors. The cellular CD69 with single mutation in amino acids responsible for the covalent (C68) or noncovalent (Q93, R134) dimerization remained in dimeric form. The same situation was observed in the double mutants (C68A/Q93A, C68A/R134A). The triple mutant (C68A/Q93A/R134A) was found as monomeric. This monomeric form is able to be dimerization by the HexNAc dimer. This dimerization causes extensively cross-linking and formation of high molecular weight aggregates.

In the last experiment was observed the influence of mutations affecting the dimerization of cellular CD69 on the ability of this receptor to activate the Jurkat cell line. This activation was analyze by the increasing of the concentration of the intracellular calcium.

The single mutations in amino acids at the noncovalent dimer (Q93A or R134A) increased the efficiency of these receptors in cellular signaling. On the other hand, when using ligand cross-

linking with the *N*-acetylhexosamine disaccharide dimer, a very low efficiency of cellular signaling was observed compared to the wild-type.

4.5 Synthetic GlcNAc Based Fully Branched Tetrasaccharide, a Mimetic of the Endogenous Ligands fo CD69, Activates CD69, Killer Lymphocytes upon Dimerization via a Hydrophilic Flexible Linker

The complex saccharide structures are involved in many biologically important signal transduction processes. They play key role in molecular recognition events contributing to cell-cell, cell-bacteria, and cell-virus interactions. The lectin receptors are able to recognize carbohydrate structures present on the surface of the tumor cell. This interaction can initiate lysis of the tumor transformated cells by the cell of the immune system. In this work we investigated two NK cells lectin activation receptors, rat NKR-P1 and human CD69. These receptors are unique for their ability to distinguish between closely related carbohydrate structures and to recognize the HexNAc in both gluco and galacto configurations. Carbohydrates interact with these lectins over an extensive surface area, but the structure and position of the oligosaccharide binding sites are unique for each of the two receptors. Rat NKR-P1 has a binding groove that accommodates the linear oligosaccharides, whereas sugar-binding sites in human CD69 are at three separate locations, and thus branched carbohydrates seem to be preferred. Three series of carbohydrate ligands for NK cells receptors NKR-P1 and CD69 were prepared (Figure 22).

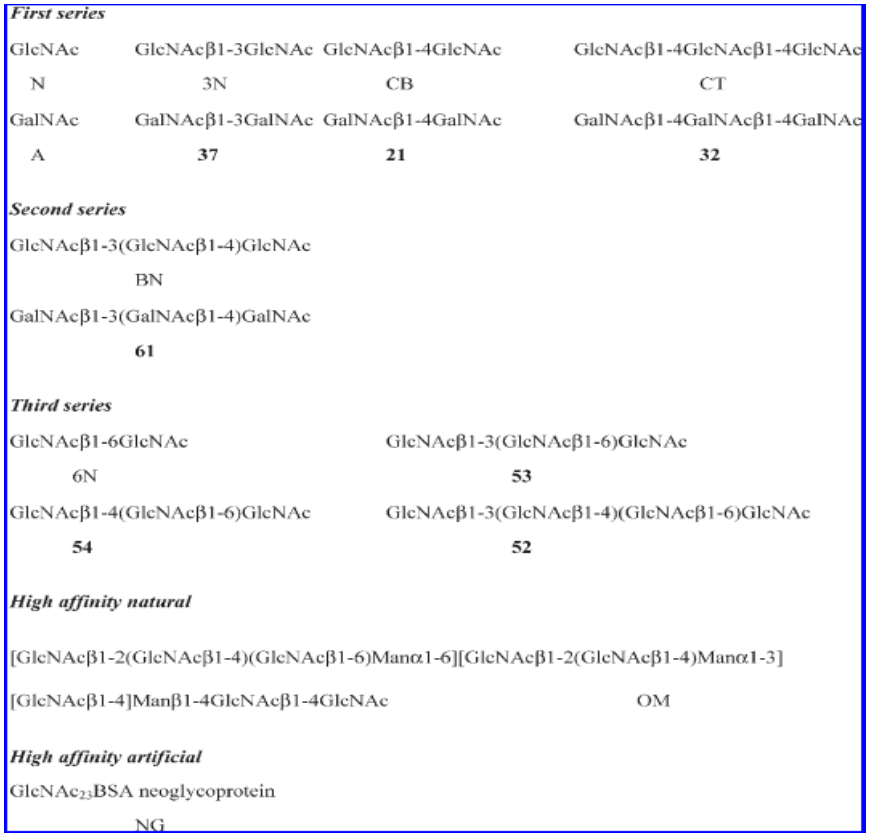


Figure 22: Structures of the Synthesized HexNAc Based Oligosaccharides.

These 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 complared to the GlcNAc control and the *β1-4* linkage is preffered 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.

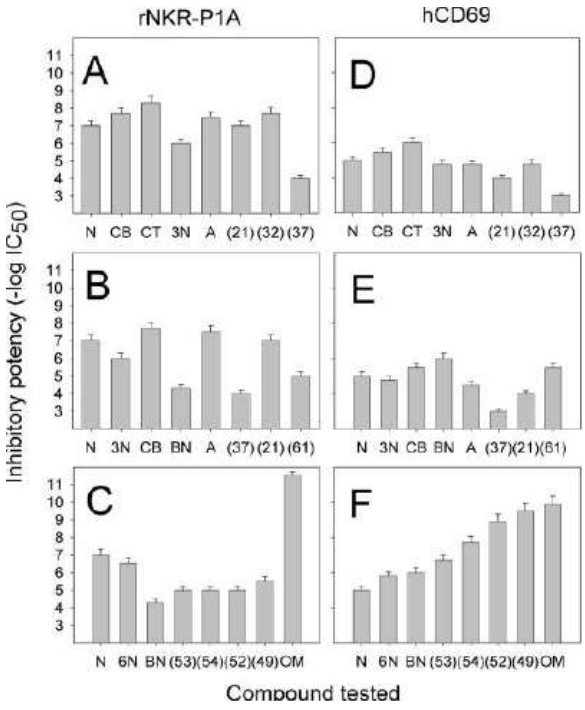


Figure 23. Biological testing of the synthesized HexNAc based oligosaccharides using inhibition assay. Indicated compounds were tested as the inhibitors of binding of the radiolabeled rNKR-P1A (left) or hCD69 (right) to the high affinity GlcNAc23BSA ligand. From the complete inhibition curves, IC% values were calculated.

The tetrasaccharide 52 achieved 10 000 times better inhibitory activity than the GlcNAc control, therefor has been selected for further experiments as an efficient mimetic.

^{3}H-labeled tetrasaccharide 52 was used for the details of binding parameters of CD69 receptors. In the receptor was one high affinity binding site per receptor subunit with Kd = 3.2 x10^{-9} M.

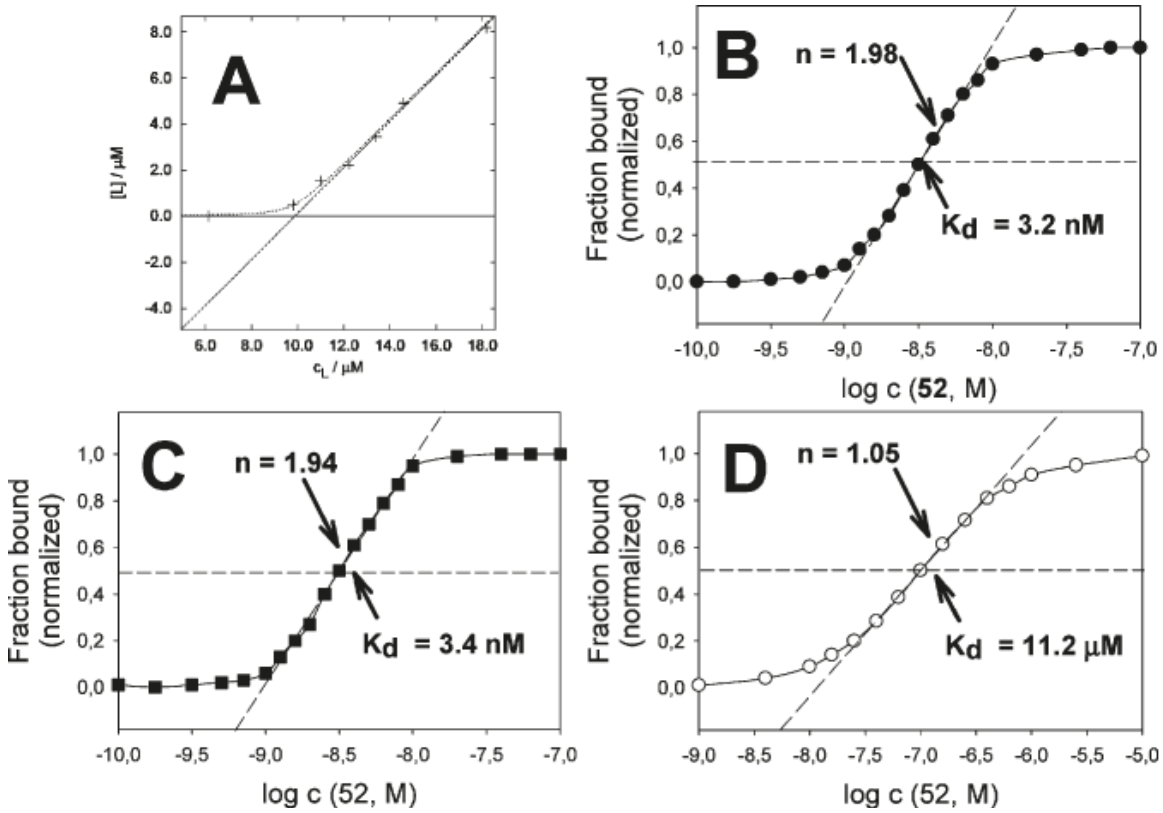


Figure 24. Direct binding of compound 52 to human CD69 receptor: (A) NMR titration of 10 µM soluble CD69 with the compound 49; (B) binding of 3H-labeled 52 to CD69 followed by equilibrium dialysis; (C, D) binding of 52 to the dimeric and monomeric form of CD69 followed by tryptophan fluorescence quenching.

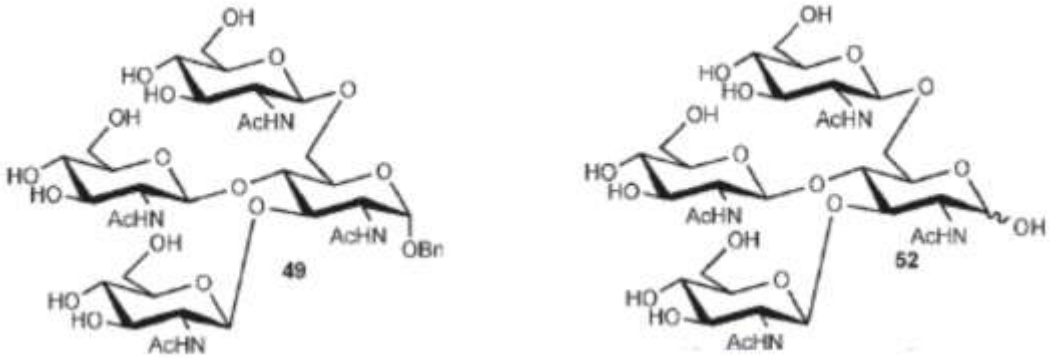


Figure 25. The structure of compounds number 49 and 52.

The compound 52 has been shown as a suitable mimetic of the complex physiological ligand for the receptor, however the ligand mimetic must be present in a multivalent (at least bivalent) form. The tetrasaccharide 52 was efficiently dimerizated using standard chemistry for peptide or protein cross-linking (Figure 26).

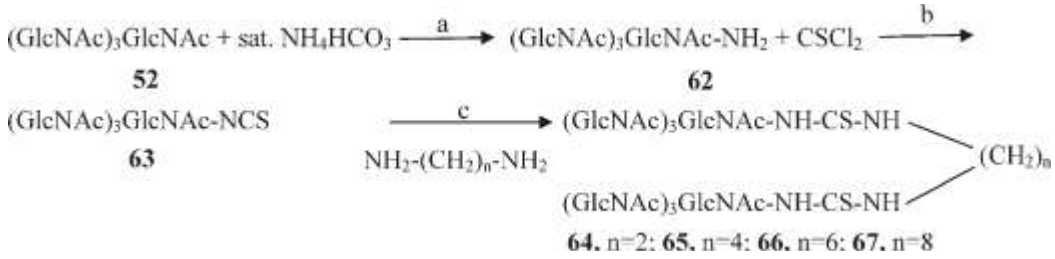


Figure 26. Reagents and conditions: (a) sat. aq NH4HCO3, 30˚C, 7 days; (b) CSCl2, NaHCO3, acetone H2O; (c) 2.4 equiv of 63 per 1 equiv of diamine, CH2Cl2, room temp, 2 days.

While the monomeric tetrasaccharide 52 was not active in the test, the dimeric compounds 64-67 proven positive results in precipitation of soluble form of CD69 receptor. The best activity was achieved with compounds 65. The activity of compounds 68 was comparable with the monoclonal antibody positive control.

The ability of the dimerized tetrasaccharides enhance the anti tumor potential of the immune system, was tested short-term cytotoxicity assay. The compound 65 and 68 increased the killing of human erythroleukemic cells line K562 and the target cell line known to be sensitive for natural killing. Moreover compound 68 was also active in the case of NK resistant tumor cell line RAJI in the situation when the other compoun had not effect.

In our last experiments we investigated the influence of tetrasaccharide 68 on suppress of the growth of experimental tumors and activating of the tumor infiltrated lymphocytes. For this attempt was preformed using experimental model of mouse B16 melanoma a low metastasis variant. Compound 68 had the most efficient decreasing the size of tumor at day 26 to day 30. The same effect had the compounds number 65 and 70 at day 30 after injection of tumor cells. For screening of killer cells operating inside tumor cytotoxic activity assay was performed. In this assya were used tumor infiltrating lymphocytes isolated from animals treated the individual compounds and assayed after isolation of these cells *ex vivo*. The compounds 65 and 68 were tested. The compounds 68 was effectived in both B16 melanoma and NK cell resistence P815 mastocytoma (Figure 28A, B). The compound 65 had statistically significant effect only in case of P815 mastocytoma tumor line.

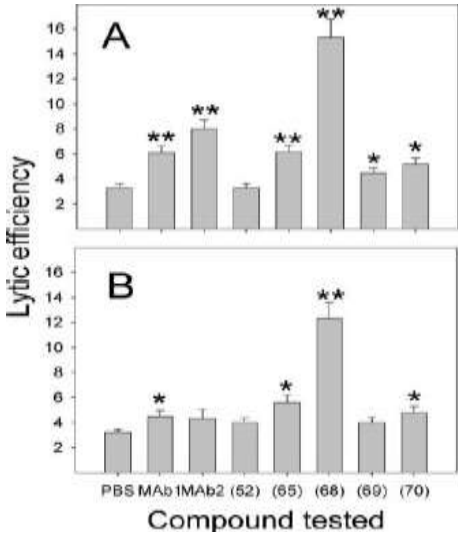


Figure 27. Natural killing assays in the presence of the tested compounds using sensitive human cell line K562 (A) and resistant human cell line RAJI (B).

.

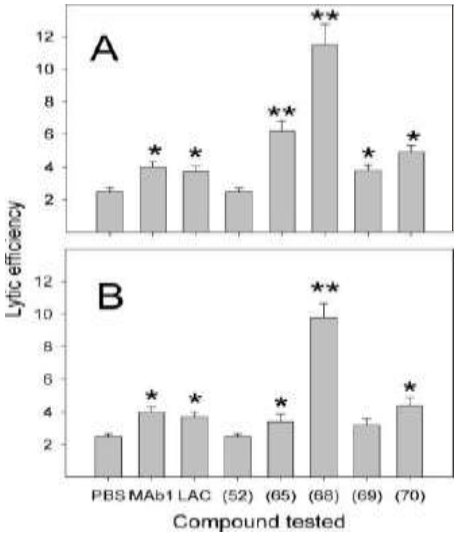


Figure 28. Natural killing of tumor infiltrating lymphocytes isolated from mice treated with the indicated compounds. Killing of B16 melanoma (A) and NK resistant mastocytoma P815 (B) targets is shown.

On the base of these experiments were found two active compounds number 52 and 68. Compound 52 is very most efficient ligands for CD69 receptor, which is able to develop effective activation of NK cells. The compound 68 increases the natural killer activity of the NK cells and decreases the growth of tumor cells *in vivo*.

5. Discussion

The main focus of this thesis was to develop expression system for recombinant production of fungal α-NAGA. α-NAGA belongs to family of glycosidase.

Glycosides are classified as enzymes catalyzing the hydrolysis of *O*- or *S*-glycosides giving smaller sugar moieties. The next classification of glycoside can also be 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. Glycoside was found in essentially all domains of life. In prokaryotes they are found both as intracellular and extracellular enzymes that are largely involved in nutrient acquisition. For example, *Bifidobacteria* posses neuroaminidases, α-glycosidases, α-GA. These enzymes allow the organisms to utilize different types of carbohydrates to adapt and compete in an environment with changing nutritional conditions.

In higher organisms glycosides are found in the endoplasmatic reticulum and Golgi apparatus where they are involved in processing of *N*-linked glycoproteins, and in the lysozyme as enzymes involved in the degradation of carbohydrate structures. Deficiency in specific lysozomal glycoside hydrolases can lead to a range of lysosomal storage disorders that result in developmental problems or death. Glycosides are found in the interstinal tracts and saliva where they degrade complex carbohydrates such as lactose, starch and sucrose. The enzyme lactase is required for degradation of the milk sugar lactose and is present at high levels in infants, but in most populations will decrease after weaning or during infancy, potentially leading to lactose intolerance in adulthood. The enzyme *O*-GlcNAc-ase is involved in removal of GlcNAc groups from serine and threonine residues in the cytoplasm and nucleus of the cell.

Some glycosidases are very extended, for example the glycosidases of higher organism occurring in ER or Golgi apparatus. However, exist glycosidases which occur very rarely and they have very often interesting properties. For example, the glycosidases from filamentous fungi have very low optimum, dual enzyme activity, enzyme activity at high temperature etc. These types of glycosidase are very often extracellular and inducible. This fact can bring the problem with finding of new type of glycosidase, because we do not know which organism produce the specific glycosidase and we do not know the inductor. On the other hand these glycosidases can be used in many industrial or pharmaceutical applications. Therefore is very important to search the new type of these enzymes, which can be useful for development of new compounds, drugs or applications. By detailed studies of these enzymes we can find new relationship between function and structure or between organisms.

For a typical screen for any given enzyme or other gene product, a functional assay is

applied: for example, proteases are screened in an assay specific for proteases, amylases are screened in an assay specific for amylases etc. The existing methods for traditional functional screening for extra-cellular enzymes are substantially limited to the applied screening assays. This means that the screening provides only those enzymes for which a functional assay exists. This method for screening for enzymes or other proteins has the disadvantage of being both time-consuming and expensive.

A screening for enzymes in fungi can be based on an expression-cloning method, which combines the ability of *S. cerevisiae* to express heterologous (fungal) genes with the utilization of enzyme assays. The fungus of interest is cultivated under conditions that give high-level enzyme activity; mRNA is prepared from the resulting biomass and a cDNA library is constructed in *E. coli* . Plasmid DNA is isolated from subpools of this library and transformed into *S. cerevisiae* . Subsequently, the yeast transformants are screened for enzyme activity.

For example (Weignerova et. al. 2008), they reported the screening of collection of filamentous fungal strains for extracellular α-NAGA activity. In this work were tested 42 strains of filamentous fungi and 6 inductors under various cultivation conditions. The existing methods for traditional functional screening for extra-cellular enzymes are substantially limited to the applied screening assays. This method for screening for enzymes or proteins has the disadvantage of being both time-consuming and expensive.

The second approach of discovering of new proteins or enzymes can be genome mining. The principle reason for gene mining is to identify and isolate genes that are characterised and have special properities. The widespread use and availability of molecular biological techniques have allowed for the rapid development and identification of nucleic acid sequences. With the availability of integration of laboratory equipment with advanced computer software, researchers are able to conduct advanced quantitative analyses, database comparisons and computational algorithms to seek and identify gene sequences. Genetic databases for organisms such as *E. coli*, or *S. cerevisiae* are available for public. These biological databases store information that is searchable and from which biological information can be retrieved. The main database is GenBank at the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov).

Using DNA databases and computer softwares we can compare several DNA sequences of related proteins/enzymes from different sources. By this comparison we can find evolutionary conserved sequences of these proteins. On the basic of this conserved sequences we can designed specific primers and isolated gene of interest. The gene of interest can be cloned into expression vector and expressed in suitable expression system.

α-NAGA from filamentous fungi *Aspergillus niger* is one of glycosidase having unique properties such as low pH optimum or dual α-GA activity According to enzyme nomenclature of IUB-MB (International Union of Biochemistry and Molecular Biology) this enzyme belongs 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. The CAZy database described the families of structurally related catalytic and carbohydrate binding modules or functional domains of enzymes that degrade, modify or create glycosidic bound. In the CAZy are enzymes divided to four classes Glycosidase-Hydrolase (GHs), Glycosyl-Transferase (GTs), Polysaccharide-Lyases (PLs) and Carbohydrate-Esterase (CEs). These classes are divided to individual families according the specific enzyme activities, evolutionary similarity and mechanism of enzymatic reaction.

The phylogenetic mapping showed that the evolution of genes for α-NAGA from vertebrates is distinct from evolution of genes for α-NAGA from fungi. α-NAGA were evolved from fungal α-GA occurring in fungi growing on solid substrates, as a way for using compounds containig terminal α-linkage GalNAc. α-NAGA can occur in organisms in various isoforms, which can be different in their catalytic activities. The catalytic activity is depend on the place where α-NAGA occur.

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 K_M value of 0.73 mmol/l for *o*-NP-α-GalNAc in 50 mM citrate-phosphate buffer (pH 3.5) at 35°C. 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 reverse transcripted into DNA sequence. The gene of interest was cloned into prokaryotic expression vector. Two prokaryotic

expression strains was 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 was 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*. 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 enviroment. 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.

The enzyme activity in neutral pH can be used for modification of blood cell type A to H(0) universal donor, because α-GalNAc is present in A and AB human blood groups. The selective deglycosylation of α-GalNAc residues from the blood agglutinogen A results in its change to the agglutinogen H(0). The problems which can occur are several. Firstly, the enzymatic cleavage have to be very specific. Specificity of enzyme cleavage can be investigated for themselves blood epitopes. For analysis of epitopes the mass spectrometry, think layer chromatography or HPLC-PAD can be used. In the case of successful with specific cleavage of themselves epitopes can be designed the next experiments with blood cells. In these experiments will be necessary to find

optimal conditions and time for treatment of blood cells, because no cell can stay with epitop A on the surface. The enzyme can be in soluble form or as immobilized on the surface of some suitable particulars, which can be placed in column or in plate. The efficiency of the treatment can be investigated by agglutination test or by flow cytometry.

In case of overcoming of all of these problems, the recombinant α-NAGA would be used as a very strong tool for preparing of blood derivates or blood cell type H(0) as a universal donor.

In vivo glycosidase catalyse the hydrolysis of glycosidic linkages. *In vitro* they can catalyse the formation of a new glycosidic bond either by transglycosylation or by reverse hydrolysis (condensation). The kinetically controlled transglycosylation requires a glycoside activated by a good leaving group, whereas the thermodynamically controlled reverse hydrolysis uses high concentrations of free sugars. Transglycosylation is used more frequently, since with this method it is more straightforward to influence the regioselectivity of the enzyme and the yield of the reaction by modifying the reaction conditions. A good choice of enzyme source, suitably modified glycosyl donors and acceptors, addition of co-solvents or cyclodextrins are good tools for the synthesis of glycosidic structures, which are hardly achievable otherwise. In comparison with glycosyltransferases, β-*N*-acetylhexosaminidases have the advantage of broad substrate specificity, robustness, accessibility and, not least, much cheaper substrates. Until now, β-*N*-acetylhexosaminidases from glycoside hydrolase family 20 have been used for synthetic purposes, but no one α-NAGA.

The synthesis of a novel immunoactive disaccharide β-D-GalNAcα-(1→4)-D-GlcNAc is the variation of a reaction with substrate modified at C-6. This carbohydrate exhibits a high affinity to the rat NK cell activation receptor NKR-P1 and human NK cell activation receptor CD69. It was prepared via a transglycosylation reaction and subsequent chemical oxidation. 2-acetamido-2-deoxy-β-D-galactohexodialdo-1,5-pyranosyl was transfered onto GlcNAc with catalysis by β-*N*-acetylhexosaminidase from *Talaromyces flavus*.

The second example of enzymatic synthesis can be ManNAc and oligosaccharides containing ManNAc at the reducing end were identified as strong ligands of the rat natural killer cell activation protein NKR-P1 (Krist et. al. 2001). However, both enzymatic and chemical syntheses of these compounds failed. The mixture of β-D-GlcNAc-(1→4)-D-GlcNAc and the desired β-D-GlcNAc-(1→4)-D-ManNAc is easily obtainable by Lobry de Bruyn-Alberda van Ekenstein epimerization of β-D-GlcNAc-(1→4)-D-GlcNAc (N, N′-diacetylchitobiose), but the problem is separating the epimers. β-*N*-acetylhexosaminidase from *Aspergillus oryzae* displayed a surprising ability to selectively discriminate the disaccharide substrates according to their reducing

saccharide unit. Thus, N,N´-diacetylchitobiose was selectively hydrolysed by β-*N*-acetylhexosaminidase, whereas its C-2 epimer (β-D-GlcNAc-(1→4)-D-ManNAc) was completely resistant to the enzyme hydrolysis.

Using the recombinant α-NAGA in enzymatic synthesis of glycomimetic should bring new possibility in preparing of mimetic and new structures, which can be used as a ligands of many proteins. Rationale for these attempts are provided by our detailed knowledge of the binding specificity of NK cell receptors. In particular, in 2005 a paper from our laboratory reported a surprising finding indicating that rat NKR-P protein bound to comb-like dendrimers bearing α-linked terminal GalNAc residues (Veprek et. al. 2006). This finding appeared initially very unusual, since an absolute majority of carbohydrate binding proteins (lectins, enzymes, antibodies etc.) absolutely distinguishes between α- and β-anomeric conformation. However, the recent data obtained with mouse NKR-P1C protein using oligosaccharide arrays at CoreH facility of Consortium for Functional Glycomics appear to confirm this unusual assignment (Figure 30). Among the oligosaccharides displaying the highest binding of this protein there was, for instance, glycan 174 having the sequence **GlcNAcα1-3**Galβ1-4GlcNAcβ-Sp8 (6328 ± 925 f.u.), glycan 339 with a sequence **GlcNAcα1-4**Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0 (4568 ± 187 f.u.), glycan 87 bearing a sequence **GalNAcα1-3**(Fuca1-2)Galβ1-4GlcNAcβ-Sp8 (3155 ± 479 f.u.), and glycan 94 with a sequence **GalNAcα1-4**(Fuca1-2)Galβ1-4GlcNAcβ-Sp8 (2887 ± 260 f.u.). α-NAGA does not cleave β-linked GlcNAc (unpublished observations), and most it will most probably not transfer this carbohydrate residue, either. However, a way is now open for possible synthesis of GalNAc α-terminated glycomimetics using this enzyme. These syntheses will be the subject of our future studies.

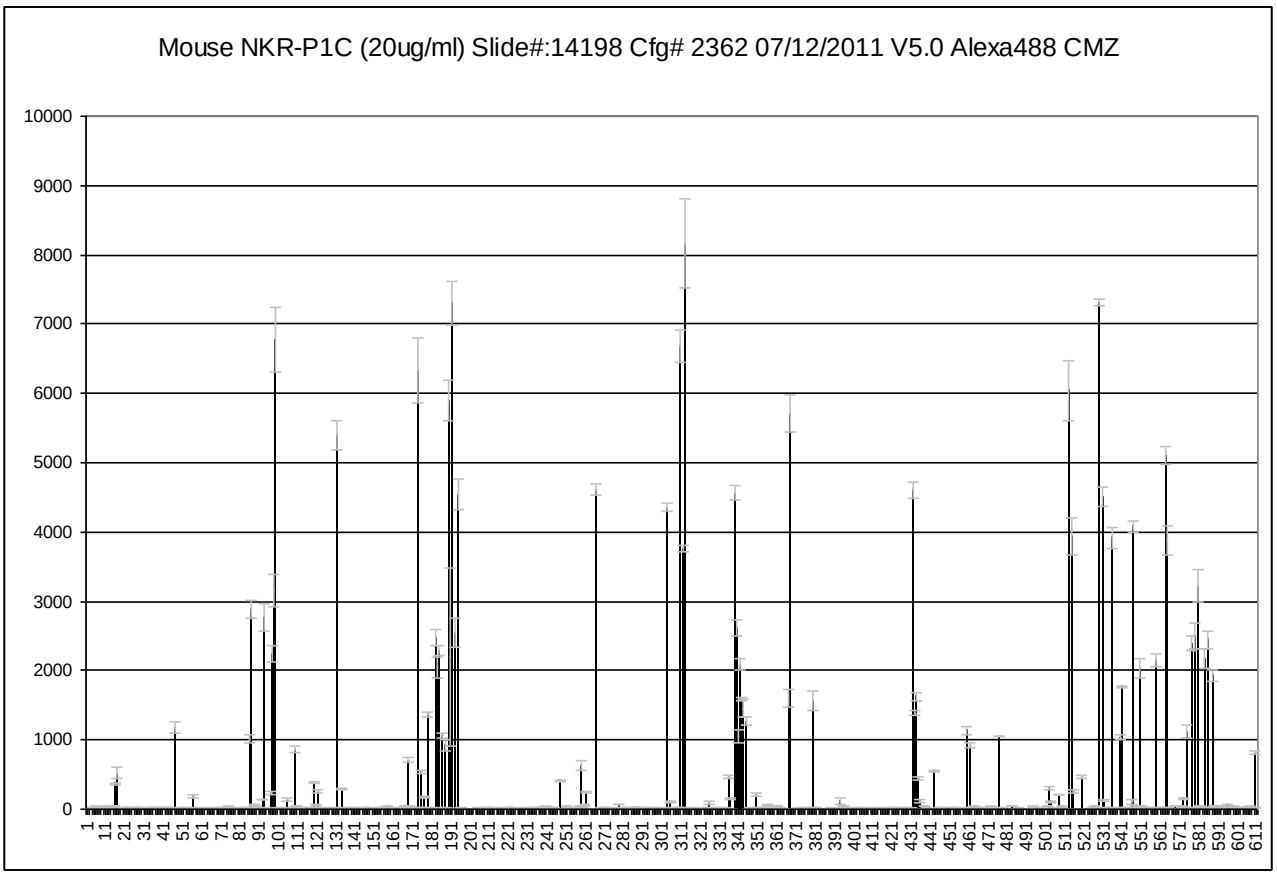


Figure 30. Binding of dimeric mouse NKR-P1C(B6) obtained using eukaryotic expression in HEK cells. This array was recorded with protein labeled using Alexa Fluor 488 at 20 μg/ml. Identification of individual binding results is provided by Core H facility.

Production of recombinant forms of soluble NK cell receptors is a long-term program in our laboratory. Ideally, it would be suitable to express the entire extracellular parts of these receptors including the cysteine residues involved in the dimerization of natural forms of these receptors that are usually present in the membrane proximal segment of at the very C-terminal end. However, for practical reasons this is usually not possible, and defining the proper extracellular segment for expression is a matter of trial and error. The problem is the presence of mutliple cysteine residues that are very abundant in these proteins which makes their expression very difficult. Usually, either the short CRD stabilized by two intrachain disulfide bridges or the long CRD having three of these bridges is expressed together with the flanking segments of the polypeptide. Protocols describing the production of full-length human CD69 containing the entire extracellular portion including the membrane proximal dimerization cysteine have been published, but we could not reproduce this protocol in our laboratory. For these reasons, we prefer to omit the extra cysteine residues (except

the 6 involved in intrachain disulfide bridging) from expression constructs for bacterial expression. After *in vitro* refolding, such protein tend to arrange either as noncovalent dimers (such as CD69, chapter 4) or as monomers (many other proteins including mouse NKR-P1C). The outcome of this refolding is most probably the inherent characteristic of the individual protein, and our possibility to influence the monomeric or dimeric forms are very limited. For instance, our recent attempts to produce covalent or even noncovalent mouse NKR-P1 dimers using bacterial expression system had limited success (David Adamek, personal communication), and such dimers were only obtained using eukaryotic expression (Figure 30). It would appear that dimeric arrangement of the receptor is important for ligand identification studies: we could not detect binding of carbohydrates to monomeric mouse NKR-P1C (Daniel Rozbesky, unpublished), but the identical dimeric protein provided very nice binding data (Figure 30). Apparently, the presence of multivalent interactions of the oligosaccharide ligands with the dimeric receptor having multiple carbohydrate binding sites is needed. This effect should not occur for the monomeric carbohydrate ligands. However, our study presented in chapter 4 provided evidence for high affinity cooperative binding under this situation (after dissociation of the subunits using molecular mutagenesis this was effect was lost). We currently investigate if this interesting effect holds true only for CD69 or if its a general feature of the entire family of C-type lectin NK cell receptors.

The potential advantage of using enzymes (such as α-NAGA) for the synthesis of carbohydrate-based mimetics is evident when one considers the complexity of total chemical synthesis of such compounds. In particular, to prepare highly branched oligosaccharides using total chemical synthesis is very difficult (chapter 5), and requires specific and efficient synthesis procedures in order to overcome steric hindering. From this point of view, the availability of enzymes including α-NAGA would be good way how to overcome this problem, and prepare highly branched mimetics with terminal GalNAc α1- sequences.

In conclusion, I 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.

References

Albericio F., Kneibcordonier N., Biancalana S., Gera L., Masada R. I., Hudson D., Barany G. (1990). Preparation and application of the 5-(4-(9-fluorenylmethoxycarbonyl) aminomethyl-3, 5dimethoxyphenoxy)-valeric acids (PAL) handle for the solid-phase synthesis of C-terminal peptide amides under conditions. *J. Org. Chem.* **55**: 3730-3743.

Anderson P., Caligiuri M., Ritz J., Schlossman S. F. (1989). CD3-negative natural killer cells express zeta TCR as part of a novel molecular complex. *Nature* **341**: 159-162.

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.

Ashkenazi A. (2002). Targeting death and decoy receptors of the tumour-necrosis factor superfamily. *Nat. Rev. Cancer* **2**: 420-430.

Avery O. T., Macleod C. M., McCarty M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction the transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type II. *J. Exp. Med.* **79**: 137-157.

Ballou C.E. (1990). Isolation, characterization, and properties of Saccharomyces cerevisiae mnn mutants with nonconditional protein glycosylation defects. *Methods Enzymol*. **185**: 440-470.

Barlos K., Gatos D., Kallitsis J., Papaphotiu G., Sotiriu P., Yao WQ., Schäfer W. (1989). Darstellung geschützter peptid-fragmente unter einsatz substituierter triphenylmethyl-harze. *Tetrahedron Lett*. **30**: 3943-3946.

Bayas M. V., Leung A., Evans E., Leckband D. (2005). Lifetime Measurements Reveal Kinetic Differences between Homophilic Cadherin Bonds. *Biophys. J.* **90**: 1385-1395.

Beadle G. W., Tatum E. L. (1941). Genetic Control of Biochemical Reactions in Neurospora. *Proc. Natl. Acad. Sci.* **27**: 499-506.

Berg D. T., Burck P. J., Berg D. H., Grinnell B. W. (1993). Kringle glycosylation in a modified human tissue plasminogen activator improves functional properties. *Blood* **81**: 1312-1322.

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.

Biassoni R., Cantoni C., Falco M., Pende D., Millo R., Moretta L., Bottino C., Moretta A. (2000). Human natural killer cell activating receptors. *Mol. Immunol*. **37**: 1015-1024.

Boyington J. C., Riaz A. N., Patamawenu A., Coligan J. E., Brooks A. G., Sun P. D. (1999). Structure of CD94 reveals a novel C-type lectin fold: implications for the NK cell-associated CD94/NKG2 receptors. *Immunity* **10**: 75-82.

Brewer, C. F. (2001). Lectin cross-linking interactions with multivalent carbohydrates. *Adv. Exp. Med. Biol*. **491**: 17-25.

Brocke C., Kunz H. (2004). Synthetic Tumor-Associated Glycopeptide Antigens from the Tandem Repeat Sequence of the Epithelial Mucin MUC4. *Synthesis* **4**: 525-542.

Cantz M., Ulrich-Bott B. (1990). Disorders of glycoprotein degradation. *J. Inherit. Metab. Dis*. **13**: 523–537.

Carlyle J. R., Jamieson A. M., Gasser S., Cligan C., Arase H., Raulet D. H. (2004). Missing self-recognition of Ocil/Clr-b by inhibitory NKR-P1 natural killer cell receptors. *Proc. Natl. Acad. Sci.* **101**: 3527-3532.

Cebrián M., Yagüe E., Rincón M., López-Botet M., de Landázuri M. O., Sánchez-Madrid F. (1988). Triggering of T cell proliferation through AIM, an activation inducer molecule expressed on activated human lymphocytes. *J. Exp. Med*. **168**: 1621-1637.

Davis B. G. (2002). Synthesis of glycoproteins. *Chem. Rev*. **102**: 579-602.

Delorme E., Lorenzini T., Giffin J., Martin F., Jacobsen F., Boone T., Elliott S. (1992). Role of glycosylation on the secretion and biological activity of erythropoietin. *Biochemistry* **31**: 9871-9876.

Dimasi N., Biassoni R. (2005). Structural and functional aspects of the Ly49 natural killer cell receptors. *Immunol. Cell. Biol.* **83**: 1-8.

Dodd R. B., Drickamer K. (2001). Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology* **11**: 71-79.

Dordal M. S., Wang F. F., Goldwasser E. (1985). The role of carbohydrate in erythropoietin action. *Endocrinology.* **116**: 2293-2299.

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

Drickamer K. (1993). Ca^{2+} dependent carbohydrate-recognition domains in animal proteins. *Curr. Opin. Struct. Biol*. **3**: 393-400.

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

Dudkin V. Y., Orlova M., Geng X. D., Mandal M., Olson, W. C., Danishefsky S. J. (2004). Toward fully synthetic carbohydrate-based HIV antigen design: on the critical role of bivalency. *J. Am. Chem. Soc*. **126**: 9560-9562.

Dwek R. A. (1996). Toward Understanding the Function of Sugars. *Chem. Rev.* **96**: 683-720.

Dziadek S., Brocke C., Kunz H. (2004). Biomimetic synthesis of the tumor-associated (2,3)-sialyl-T antigen and its incorporation into glycopeptide antigens from the mucins MUC1 and MUC4. *Chem. Eur. J*. **10**: 4150-4162.

Ecker M., Mrsa V., Hagen I., Deutzmann R., Strahl S., Tanner W. (2003). O-mannosylation precedes and potentially controls the N-glycosylation of a yeast cell wall glycoprotein. *EMBO Rep*. **4**: 628-632.

Egrie J. C., Dwyer E., Browne J. K., Hitz A., Lykos M. A. (2003). Darbepoetin alfa has a longer circulating half-life and greater in vivo potency than recombinant human erythropoietin. *Exp. Hematol*. **31**: 290-299.

Ezekowitz R. A., Stahl P. D. (1988). The structure and function of vertebrate mannose lectin-like proteins. *J. Cell. Sci. Suppl*. **9**: 121-133.

Floersheimer A., Riniker B. (1991). *Proc. Eur. Pept. Symp*.**131**: 21.

Fujita M., Shoda S., Haneda K., Inazu T., Takegawa K., Yamamoto K. (2001). A novel disaccharide substrate having 1,2-oxazoline moiety for detection of transglycosylating activity of endoglycosidases. *Biochim. Biophys. Acta* **1528**: 9-14.

Gabius H. J., Siebert H. C., Andre S., Jimenez-Barbero J., Rudiger H. (2004). Chemical biology of the sugar code. *ChemBioChem* **5**: 740-764.

Gamblin D. P.; Scanlan E. M.; Davis B. G. (2009). Glycoprotein Synthesis: An Update. *Chem. Rev*. **109**: 131–163.

Gemmill T. R., Trimble R. B. (1999). Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta* **1426**: 227-237.

Gerardy-Schahn, R., Oelmann S., Bakker H. (2001). Nucleotide sugar transporters: Biological and functional aspects. *Biochimie* **83**: 775-782.

Goldstein I. J., Hughes R. C., Monsigny M., Osawa T., Sharon N. (1980). What should be called a lectin. *Nature* **285**: 66-74.

Goldwasser E., Kung C. K., Eliason J. (1974). On the mechanism of erythropoietin-induced differentiation. The role of sialic acid in erythropoietin action. *J. Biol. Chem*. **249**: 4202-4206.

Goto S., Taniguchi M., Muraoka M., Toyoda H., Sado Y., Kawakita M., Hayashi S. (2001). UDP-sugar transporter implicated in glycosylation and processing of Notch. *Nature Cell Biol.* **3**: 816-822.

Hamann J., Fiebig H., Strauss M. (1993). Expression cloning of the early activation antigen CD69, a type II integral membrane protein with a C-type lectin domain. *J. Immunol*. **150**: 4920-4927.

Haselbeck, A., Tanner W. (1982). Dolichyl phosphate-mediated mannosyl transfer through liposomal membranes. *Proc. Natl. Acad. Sci*. **79**: 1520-1524.

Hara T., Jung L. K., Bjorndahl J. M., Fu S. M. (1986). Human T cell activation. III. Rapid induction of a phosphorylated 28 kD/32 kD disulfide-linked early activation antigen (EA 1) by 12-o-tetradecanoyl phorbol-13-acetate, mitogens, and antigens. *J. Exp. Med*. **164**: 1988-2005.

He Y., Hinklin R. J., Chang J. Y., Kiessling L. L. (2004). Stereoselective N-glycosylation by Staudinger ligation. *Org. Lett*. **6**: 4479-4482.

Helenius A. (1994). How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell* **5**: 253-265.

Helenius J., Ng D., Marolda T. C. L., Walter P., Valvano M., Aebi A. M. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415**: 447-450.

Henrissat B., Coutinho P.M. (1999). Carbohydrate-Active Enzymes server. www. cazy.org
Henrissat B., Davies G. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* **3**: 853–859.

Hewitt M. C., Snyder D. A., Seeberger P. H. (2002). Rapid synthesis of a glycosylphosphatidylinositol-based malaria vaccine using automated solid-phase oligosaccharide synthesis. *J. Am. Chem. Soc*. **124**: 13434-13436.

Higuchi M., Oh-eda M., Kuboniwa H., Tomonoh K., Shimonaka Y., Ochi N. (1992). Role of sugar chains in the expression of the biological activity of human erythropoietin. *J. Biol. Chem*. **267**: 7703-7709.

Hirschberg C. B., Robbins P. W., Abeijon C. (1998). Transporters nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu. Rev. Biochem.* **67**: 49-69.

Holm L., Kjellen P., Holmdahl R., Kihlberg J. (2005). Identification of the minimal glycopeptide core recognized by T cells in a model for rheumatoid arthritis. *Bioorg. Med. Chem.* **13**: 473-482.

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

Imperiali B., O'Connor S. E. (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol*. **3**: 643-649.

Kelleher D. J., Kreibich G., Gilmore R. (1992). Oligosaccharyltransferase activity is associated with a protein complex composed of ribophorins I and II and a 48 kd protein. *Cell* **69:** 55-65.

Kern G., Kern D., Jaenicke R., Seckler R. (1993). Kinetics of folding and association of differently glycosylated variants of invertase from Saccharomyces cerevisiae. *Protein Science* **2**: 1862-1868.

Knauer R., Lehle L. (1994). The N-oligosaccharyltransferase complex from yeast. *FEBS Lett.* **344**: 83-86.

Kornfeld R., Kornfeld S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**: 631-664.

Krist P., Herkommerová-Rajnochova E., Rauvolfova J., Semenuk T., Vavruskova P., Pavlícek J., Bezouska K., Petrus L., Kren V. (2001). Toward an optimal oligosaccharide ligand for rat natural killer cell activation receptor NKR-P1. *Biochem. Biophys. Res. Commun.* 287: 11-20.

Kubota N., Orita T., Hattori K., Oh-eda M., Ochi N., Yamazaki T. (1990). Structural characterization of natural and recombinant human granulocyte colony-stimulating factors. *J. Biochem.* **107**: 486-492.

Kukuruzinska M. A., Robbinson P. V. (1987). Protein glykosylation in yeast – transcript heterogeneity of the Alg7 gene. *Porc. Natl. Acad. Sci.* **84**: 2145-2149.

Lanier L. L. (2005). NK cell recognition. *Annu. Rev. Immunol.* **23**: 225-274.

Lanier L. L., Corliss B. C., Wu J., Leong C., Phillips J. H. (1998). Immunoreceptor DAP12 bearing a tyrosine-based activation motif is involved in activating NK cells. *Nature* **391**: 703-707.

Lazetic S., Chang C., Houchins J. P., Lanier L. L., Phillips J. H. (1996). Human natural killer cell receptors involved in MHC class I recognition are disulfide-linked heterodimers of CD94 and NKG2 subunits. *J. Immunol*. **157**: 4741-4745.

Leibson P. J. (1997). Signal transduction during natural killer cell activation: inside the mind of a killer. *Immunity* **6**: 655-661.

Li Y., Hofman M., Wang Q., Teng L., Chlewicki L. K., Pircher H., Mariuzza R. A. (2009). Structure of natural killer cell receptor KLRG1 bound to E-cadherin reveals basis for MHC-independent issing self recognition. *Immunity* **31**: 98-105.

Liu H., Wang L., Brock A., Wong C. H., Schultz P. G. (2003). A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc*. **125**: 1702-1703.

Macmillan D., Bertozzi C. R. (2004). Modular assembly of glycoproteins: towards the synthesis of GlyCAM-1 by using expressed protein ligation. *Angew. Chem. Int. Ed.* **43**: 1355-1359.

Mandal M., Dudkin V. Y., Geng X. D., Danishefsky S. (2004). In pursuit of carbohydrate-based HIV vaccines, part 1: The total synthesis of hybrid-type gp120 fragments. *Angew. Chem. Int. Ed.* **43**: 2557-2561.

Markussen J., Diers I., Engesgaard A., Hansen M. T., Hougaard P., Langkjaer L., Norris K., Ribel U., Soerensen A. R. (1987). Soluble, prolonged-acting insulin derivatives. II. Degree of protraction and crystallizability of insulins substituted in positions A17, B8, B13, B27 and B30. *Protein Eng*. **1**: 215-223.

Markussen J., Diers I., Hougaard P., Langkjaer L., Norris K., Snel L., Soerensen A. R., Soerensen E., Voigt H. O. (1988). Soluble, prolonged-acting insulin derivatives. III. Degree of protraction, crystallizability and chemical stability of insulins substituted in positions A21, B13, B23, B27 and B30. *Protein Eng*. **2**: 157-166.

Matsuda M., Nishimura S. I., Nakajima F., Nishimura T. (2001). Heterobifunctional ligands: practical chemoenzymatic synthesis of a cell adhesive glycopeptide that interacts with both selectins and integrins. *J. Med. Chem.* **44**: 715-724.

Mergler M., Tanner R., Gosteli J., Grogg P. (1988). Peptide synthesis by a combination of solid-phase and solution methods I: A new very acid-labile anchor group for the solid phase synthesis of fully protected fragments. *Tetrahedron Lett.* **29**: 4005-4008.

Measaeli N., Nakamura K., Zvaritch E., Dickie P., Dziak E., Krause K. H., Opas M., MacLennan D. H., Michalak M. (1999). Calreticulin is essential for cardiac development. *J. Cell. Biol*. **144**: 857-868.

Molinari M., Helenius A. (1999). Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* **402**: 90-93.

Montel A. H., Bochan M. R., Hobbs J. A., Lynch D. H., Brahmi Z. (1995). Fas involvement in cytotoxicity mediated by human NK cells. *Cell Immunol*. **166**: 236-246.

Naumoff D. G. (2006). Development of a hierarchical classification of the TIM-barrel type glycoside hydrolases. *Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure.* **1**: 294-298.

Ogata K., Nishikaw H., Ohsugi M. (1969). A yeast capable of utilizing methanol. *Agr. Biol. Chem.* **33**: 1519-1520.

O'Shea J. J., Weissman A. M., Kennedy I. C .S., Ortaldo J. R. (1991). Engagement of the natural killer cell IgG Fc receptor results in tyrosine phosphorilation of the £ chain. *J. Immunol*. **88**: 350-354.

Paulsen H., Peters S., Bielfeldt T., Meldal M., Bock K. (1995). Synthesis of the glycosyl amino acids N alpha-Fmoc-Ser[Ac4-beta-D-Galp-(1-->3)-Ac2-alpha-D-GalN3p]-OPfp and N alpha-Fmoc-Thr[Ac4-beta-D-Galp-(1-->3)-Ac2-alpha-D-GalN3p]-OPfp and the application in the solid-phase peptide synthesis of multiply glycosylated mucin peptides with Tn and T antigenic structures. *Carbohydr. Res*. **268**: 17-34.

Perrusia B., (2000). Signaling for cytotoxicity. *Nature Immunol.* **1**: 372-374.

Rey M., Mrazek H., Pompach P., Novak P., Pelosi L., Brandolin G., Forest E., Havlicek V and Man P. (2010). Effective Removal of Nonionic Detergents in Protein Mass Spectrometry, Hydrogen/Deuterium Exchange, and Proteomics. *Anal. Chem*. **82**: 5107–5116.

Rink H. (1987). Solid-phase synthesis of protected peptide fragments using a trialkoxy-diphenyl-methylester resin. *Tetrahedron Lett* **28**: 3787-3797.

Ryan J. C., Seaman W. E. (1997). Divergent functions of lectin-like receptors on NK cells. *Immunol. Rev*. **155**: 79-89.

Sakuraba H., Matsuzawa F., Aikawa S., Doi H., Kotani M., Nakada H., Fukushige T., Kanzaki T. (2004). Structural and immunocytochemical studies on alpha-N-acetylgalactosaminidase deficiency (Schindler/Kanzaki disease). *J. Hum. Genet*. **49**: 1-8.

Sato M., Sadamoto R., Niikura K., Monde K., Kondo H., Nishimura S. I. (2004). Site-specific introduction of sialic acid into insulin. *Angew. Chem. Int. Ed.* **43**: 1516-1520.

Schofield L., Hewitt M. C., Evans K., Siomos M. A., Seeberger P. H. (2002). Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. *Nature* **418**: 785-789.

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.

Seitz O., Kunz H. (1995). A Novel Allylic Anchor for Solid-Phase Synthesis—Synthesis of Protected and Unprotected O-Glycosylated Mucin-Type Glycopeptides. *Angew. Chem*. **34**: 803-805.

Seitz O., Wong CH. (1997). Solution- and Solid-Phase Synthesis of N-Protected Glycopeptide Esters of the Benzyl Type as Substrates for Subtilisin-Catalyzed Glycopeptide Couplings. *J. Am. Chem. Soc* . **120**: 1979-1989.

Shailubhai K., Pratta M. A., Vijay I. K. (1987). Purification and characterization of glucosidase I involved in N-linked glycoprotein processing in bovine mammary gland. *Biochem. J*. **247**: 555-562.

Shailubhai K., Pukazhenthi B. S., Saxena E. S., Varma G. M., Vijay I. K. (1991). Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *J. Biol. Chem*. **266**: 16587-16593.

Shao N., Guo Z. (2005). Solution-phase synthesis with solid-state workup of an O-glycopeptide with a cluster of cancer-related T antigens. *Org. Lett*. **7**: 3589-3592.

Ma S., Hill E. K., Burk F. R., Caprioli M. R. (2003). Mass spectrometry identification of N- and O-glycosylation sites of full-length rat selenoprotein P and determination of selenide-sulfide and disulfide linkages in the shortes isoforms. *Biochemistry* **42**: 9703-9711.

Sieber P. (1987). An improved method for anchoring of 9-fluorenylmethoxycarbonyl-amino acids to 4-alkoxybenzyl alcohol resins. *Tetrahedron Lett.* **28**: 6147-6150.

Silberstein S., Gilmore R. (1996). Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.* **10**: 849-858.

Singh S., Ni J. H., Wang L. X. (2003). Chemoenzymatic synthesis of high-mannose type HIV-1 gp120 glycopeptides. *Bioorg. Med. Chem. Lett.* **13**: 327-330.

Smyth M. J., Hayakawa Y., Takeda K., Yagita H. (2002). New aspects of natural-killer-cell surveillance and therapy of cancer. *Nat. Rev. Cancer*. **2**: 850-861.

Tanaka H., Iwata Y., Takahashi D., Adachi M., Takahashi T. (2005). Efficient stereoselective synthesis of gamma-N-glycosyl asparagines by N-glycosylation of primary amide groups. *J. Am. Chem. Soc*. **127**: 1630-1631.

Testi R., D'Ambrosio D., De Maria R., Santoni A. (1994). The CD69 receptor: a multipurpose cell-surface trigger for hematopoietic cells. *Immunol. Today* **15**: 479-83.

Tormo J., Natarajana K., Margulies D. H., Mariuzza R. A. (1999). Crystal structure of a lectin-like natural killer cell receptor bound to its MHC class I ligand. *Nature.* **402**: 623-631.

Trapani J. A., Davis J., Sutton V. R., Smyth M. J. (2000). Proapoptotic functions of cytotoxic lymphocyte granule constituents in vitro and in vivo. *Curr. Opin. Immunol*. **12**: 323-329.

Umehara F., Matsumuro K., Kurono Y., Arimura K., Osame M., Kanzaki T. (2004). Neurologic manifestations of Kanzaki disease. *Neurology* **62**: 1604-1606.

Unverzagt C., Kunz H., Paulson JC. (1990). High-efficiency synthesis of sialyloligosaccharides and sialoglycopeptides. *J. Am. Chem. Soc*. **112**: 9308-9309.

Unverzagt C.; Kelm S.; Paulson J. C. (1994). Chemical and enzymatic synthesis of multivalent sialoglycopeptides. *Carbohydr. Res*. **251**: 285-301.

Varki A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**: 97-130.

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**: 6400-6407.

Verez-Bencomo V., Fernandez-Santana V., Hardy E., Toledo M. E., Rodriguez M. C., Heynngnezz L., Rodriguez A., Baly A., Herrera L., Izquierdo M., Villar A., Valdes Y., Cosme K., Deler M. L., Montane M., Garcia E., Ramos A., Aguilar A., Medina E., Torano G., Sosa I., Hernandez I., Martinez R., Muzachio A., Carmenates A., Costa L., Cardoso F., Campa C., Diaz M., Roy R. (2004). A synthetic conjugate polysaccharide vaccine against Haemophilus influenzae type b. *Science* **305**: 522-525.

Vocadlo D. J., Davies G. J., Laine R., Withers S. G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**: 835-838.

Volk T., Cohen O., Geiger B. (1987). Formation of heterotypic adherens-type junctions between L-CAM containing liver cells and A-CAM containing lens cells. *Cell* **50**: 987–994.

Wang A. M., Schindler D., Desnick R. (1990). Schindler disease: the molecular lesion in the alpha-N-acetylgalactosaminidase gene that causes an infantile neuroaxonal dystrophy. *J. Clin. Invest.* **86**: 1752–1756.

Wang S. S. (1973). p-alkoxybenzyl alcohol resin and p-alkoxybenzyloxycarbonylhydrazide resin for solid phase synthesis of protected peptide fragments. *J. Am. Chem. Soc*. **95**: 1328-1333.

Wang Z.G., Zhang X. F., Visser M., Live D., Zatorski A., Iserloh U., Lloyd K. O., Danishefsky S. J. (2001). Toward Fully Synthetic Homogeneous Glycoproteins: A High Mannose Core Containing Glycopeptide Carrying Full H-Type 2 Human Blood Group Specificity. *Angew. Chem. Int. Ed.* **40**: 1728-1732.

Wang F., Hirs F. C. H. (1977). Influence of the heterosaccharides in porcine pancreatic ribonuclease on the conformation and stability of the protein. *J. Biol. Chem*. **252**: 8358-8364.

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.

Weis P., Ashwell G. (1989). Ligand-induced modulation of the hepatic receptor for asialoglycoproteins. Evidence for the role of cell surface hyposialylation. *J. Biol. Chem*. **264**: 11572-11574.

Wildt S., Gerngross T. U. (2005). The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol*. **3**: 119-128.

Wong CH., Haynie SL., Whitesides GM. (1982). Enzyme-catalyzed synthesis of N-acetyllactosamine with in situ regeneration of uridine 5'-diphosphate glucose and uridine 5'-diphosphate galactose. *J. Org. Chem*. **47**: 5416-5418.

Wong S. Y. C. (1995). Advances in the use of Bacillus subtilis for the expression and secretion of heterologous proteins. *Curr. Opin. Struct. Biol*. **5**: 517-522.

Xu R., Hanson S. R., Zhang Z., Yang Y. Y., Schultz P. G., Wong C. H. (2004). Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J. Am. Chem. Soc*. **126**: 15654-15655.

Yan A., Lennarz W. J. (2005). Unraveling the mechanism of protein N-glycosylation. *J. Biol. Chem*. **280**: 3121-3124.

Yan S. C., Razzano P., Chao Y. B., Walls J. D., Berg D. T., McClure D. B., Grinnell B. W. (1990). Characterization and novel purification of recombinant human protein C from three mammalian cell lines. *Biotechnology* **8**: 655-661.

Yokoyama W. M. (1998). Natural killer cell receptors. *Curr. Opin. Immunol*. **10**: 298-305.
Yokoyama S. (2003). Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**: 39-43.

Zhang Z., Gildersleeve J., Yang Y. Y., Xu R., Loo J. A., Uryu S., Wong C. H., Schultz P. G. (2004). A new strategy for the synthesis of glycoproteins. *Science* **303**: 371-373.

Appendixes

Appendix 1
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.

Appendix 2
Mrazek H., Benada O., Man P., Vaněk O., Kren V., Bezouska K., Weignerova L. (2011). Facile production of A*spergillus niger* α-*N*-acetylgalactosaminidase in yeast. *Prot. Exp.Purif.* Under review.

Appendix 3
Mrazek H., Weignerova L., Bezouska K. (2010). Aktivní forma α-*N*-acetylgalaktosaminidázy z vláknité houby *Aspergillus niger* a její rekombinantní exprese. ***Czech Patent Application***

Appendix 4
Kavan D., Kubícková M., Bílý J., Vanek O., Hofbauerová K., Mrázek H., Rozbeský D., Bojarová P., Kren V., Zídek L., Sklenár 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.

Appendix 5
Kovalová A., Ledvina M., Saman D., Zyka D., Kubícková M., Zídek L., Sklenár V., Pompach P., Kavan D., Bílý J., Vanek O., Kubínková Z., Libigerová M., Ivanová L., Antolíková M., Mrázek H., Rozbeský D., Hofbauerová 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.

Appendix 1

Kulik N., Weignerova L., Filipi T., Pompach P., Novak P., Mrazek H., Slamova K., Bezouska K., Kren V., Ettrich R.

The α-galactosidase type A gene *aglA* from *Aspergillus niger* encodes a fully functional α-*N*-acetylgalactosaminidase

Glycobiology **20**: 1410-1419 (2010).

Glycobiology vol. 20 no. 11 pp. 1410–1419, 2010
doi: 10.1093/glycob/cwq105
Advance Access publication on July 4, 2010

The α-galactosidase type A gene *aglA* from *Aspergillus niger* encodes a fully functional α-*N*-acetylgalactosaminidase

Natallia Kulik[2,3], Lenka Weignerová[4,7], Tomáš Filipi[4], Petr Pompach[5,7], Petr Novák[5,7], Hynek Mrázek[6,7], Kristýna Slámová[4], Karel Bezouška[6,7], Vladimir Křen[1,4], and Rüdiger Ettrich[1,2,3]

[2]Center for Biocatalysis and Biotransformation, Institute of Systems Biology and Ecology, Academy of Sciences of the Czech Republic, Zámek 136, CZ 373 33 Nové Hrady, Czech Republic; [3]Institute of Physical Biology, University of South Bohemia, Zámek 136, CZ 373 33 Nové Hrady, Czech Republic; [4]Center for Biocatalysis and Biotransformation, and [5]Laboratory of Molecular Structure Characterization, and [6]Laboratory of Protein Architecture, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, CZ 142 20 Praha 4, Czech Republic; and [7]Department of Biochemistry, Faculty of Science, Charles University in Prague, Albertov 2030, CZ 128 40 Praha 2, Czech Republic

Received on March 1, 2010; revised on June 28, 2010; accepted on June 30, 2010

Two genes in the genome of *Aspergillus niger*, *aglA* and *aglB*, have been assigned to encode for α-D-galactosidases variant A and B. However, analyses of primary and 3D structures based on structural models of these two enzymes revealed significant differences in their active centers suggesting important differences in their specificity for the hydrolyzed carbohydrates. To test this unexpected finding, a large screening of libraries from 42 strains of filamentous fungi succeeded in identifying an enzyme from *A. niger* CCIM K2 that exhibited both α-galactosidase and α-*N*-acetylgalactosaminidase activities, with the latter activity predominating. The enzyme protein was sequenced, and its amino acid sequence could be unequivocally assigned to the enzyme encoded the *aglA* gene. Enzyme activity measurements and substrate docking clearly demonstrated the preference of the identified enzyme for α-*N*-acetyl-D-galactosaminide over α-D-galactoside. Thus, we provide evidence that the α-galactosidase type A gene *aglA* from *A. niger* in fact encodes a fully functional α-*N*-acetylgalactosaminidase using a retaining mechanism.

Keywords: α-*N*-acetylgalactosaminidase/*Aspergillus niger*/α-galactosidase/molecular modeling/substrate binding

[1]To whom correspondence should be addressed: Tel: +420386361297; Fax: +420386361279; ettrich@nh.usbe.cas.cz (R. Ettrich), kren@biomed.cas.cz (V. Kren).

Introduction

α-Galactosidases (α-D-galactopyranoside galactohydrolase, EC 3.2.1.22) occur widely in microorganisms, plants, and animals and have considerable potential in practical applications (Weignerová et al. 2009). The enzymes encoded by the genes *aglA* and *aglB* from *Aspergillus niger* (Pel et al. 2007) belong to glycohydrolase (GH) family 27, which mainly comprises α-galactosidases, α-*N*-acetylgalactosaminidases, and related enzymes. Fungal α-galactosidases are typically enzymes with a length of 380–430 amino acids and a highly conserved active site (Hart et al. 2000; Ly et al. 2000).

Unlike α-galactosidases, the availability of α-*N*-acetylgalactosaminidases is somewhat lower. These enzymes have so far mostly been isolated and structurally characterized from animal sources (Garman et al. 2002) and bacteria (Hsieh et al. 2000, 2003). The first human crystal structure was reported only recently (Clark and Garman 2009). With respect to the catalytic mechanism of α-*N*-acetylgalactosaminidases, a double-displacement mechanism using two carboxylate groups as a catalytic pair is expected with the anomeric configuration of the leaving monosaccharide unit unchanged (Weignerová et al. 2008; Clark and Garman 2009).

A deficiency or mutations of α-*N*-acetylgalactosaminidase in humans lead to a lysozomal storage disorder causing Kanzaki disease resulting in neurodegenerative pathologies (Clark and Garman 2009). α-*N*-Acetylgalactosaminidase has been used to convert blood group epitope A to the universal blood group H(0) (Liu et al. 2007; Olsson and Clausen 2008). Recently, a promising candidate from fungi was reported, the α-*N*-acetylgalactosaminidase from *A. niger* CCIM K2 (Weignerová et al. 2008). Although the *aglA* gene is assigned to encode a α-galactosidase, this study attempts to prove that the experimentally reported α-*N*-acetylgalactosaminidase is in fact identical with the encoded enzyme using biochemical, genetic, and structural characterization in combination with computational modeling.

Results

The recent sequencing of the entire genome of *A. niger* (Pel et al. 2007) opened the possibility of a targeted search for genes encoding potential α-galactosidases. A BLAST search for α-galactosidase primary sequences within the *A. niger* genome in the non-redundant protein database found five distinct protein-coding genes. Apart from genes *aglA* and *aglB*, there are three sequences with the sequence identity more than 33%. However, these three sequences are not yet

Fig. 1. Summary of sequencing data for enzyme acting as α-*N*-acetylgalactosaminidase. (**A**) Separation of prepared enzymes differing in their indicated N-terminal sequences by 2D electrophoresis, pI ranged from 4.7 (left) to 4.9 (right), and molecular mass ranged from 50 kDa (top) to 40 kDa (bottom), indicating minor heterogeneity in enzyme preparation confirmed by N-terminal sequencing. (**B**) Upper lane indicates amino acid sequence of *aglA* gene from *Aspergillus niger* (Pel et al. 2007) with signal peptide sequence bold and underlined and individual putative *N*-glycosylation sites shown in italics and underlined. 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 obtained by in gel digestion using trypsin or Asp-N proteases (underlined). Experimentally confirmed sites of *N*-glycosylation are shown in bold. Final sequence coverage using all techniques was approximately 67%, i.e., 363 of the 514 amino acids of the mature form of the enzyme were covered.

characterized and, therefore, are assigned to hypothetical proteins with unknown function (gene IDs 4984860, 4978099, and 4987036). The lengths of the hypothetical open reading frames are similar to *aglB* and vary from 391 to 431 amino acids.

Sequence analyses revealed that the enzymes coded by *aglA* and *aglB* genes differ in their size and in the amino acids of their active site. The sequence identity of the two enzymes estimated by BLAST only reaches 28%. Whereas *aglB* has an average sequence identity of more than 70% with known α-galactosidases from the same GH 27 family, the enzyme encoded by *aglA* is much closer to another cluster of enzymes having 65–78% identity with various putative α-galactosidases from fungi, as well as 64% identity with the α-*N*-acetylgalactosaminidase from *Acremonium* sp. No. 413. These relationships opened up the possibility that this gene in fact encodes an α-*N*-acetylgalactosaminidase, e.g., an enzyme characterized as an exoglycosidase (α-GalNAc-ase, EC 3.2.1.49, GH family 27) specific for the hydrolysis of terminal α-linked *N*-acetylgalactosamine in various sugar chains.

The issue has been addressed in a large screening study aimed at obtaining a good producer of extracellular α-*N*-acetylgalactosaminidase activity. A library of filamentous fungi (42 strains) and a series of inducers and cultivation conditions were examined (Weignerová et al. 2008). We have observed the presence of at least four extracellular α-galactosidases in *A. niger* culture extracts. However, only a single enzyme demonstrated α-*N*-acetylgalactosaminidase activity, too. The purified protein from the best producer, *A. niger* CCIM K2, exhibited a dual enzyme activity and was able to hydrolyze both 2-nitrophenyl-2-acetamido-2-deoxy-α-D-galactopyranoside (2-NP-α-GalNAc) and 4-nitrophenyl-α-D-galactopyranoside (4-NP-α-Gal) as its substrate. The specific activity of the purified enzyme was 33.5 and 3.1 U/mg for the α-GalNAc and the α-Gal substrates, respectively, under the given experimental conditions. Thus, the ability to hydrolyze α-GalNAc substrate was more than tenfold higher compared to the hydrolysis of α-Gal substrate. Additional biochemical characteristics of the isolated enzyme were also provided. The native molecular weight was estimated to be approximately 440 kDa, and the experimentally obtained pI was close to 4.8. The K_M for 2-NP-α-GalNAc substrate was 0.73 mM, and the optimum of enzyme activity was achieved at pH 1.8 and 55°C. The enzyme belongs to retaining glycosidases as proved by nuclear magnetic resonance determination of the α/β pyranose forms of the monosaccharide formed during the hydrolysis of 2-NP-α-GalNAc (Weignerová et al. 2008).

The purity of the isolated α-*N*-acetylgalactosaminidase was verified by 2D electrophoresis using narrow pI strips. During this analysis, the enzyme resolved into two spots, differing only in the presence or absence of three N-terminal amino acids (Ser–Ile–Glu) as demonstrated by N-terminal sequencing (Figure 1A). Mass spectrometry techniques were used for additional sequencing that due to their sensitivity used less of the limited amount of protein material available. The results allowed us to conclude that the protein obtained is encoded by

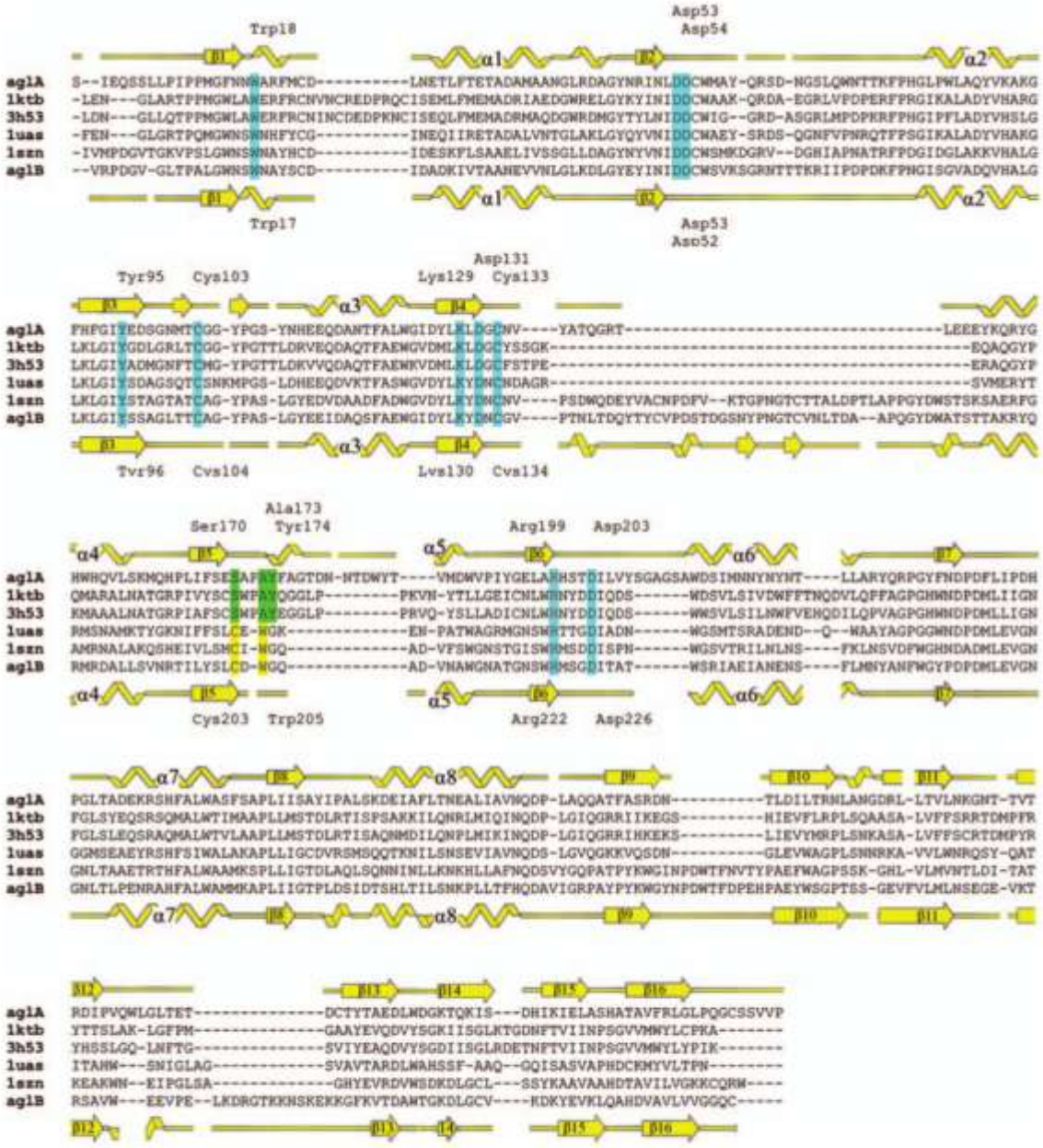


Fig. 2. Multiple sequence alignment of the α-*N*-acetylgalactosaminidase from chicken (pdb code: 1KTB), the α-*N*-acetylgalactosaminidase from humans (pdb code: 3H53), the α-galactosidase from rice (pdb code: 1UAS), the α-galactosidase *Trichoderma reesei* (pdb code: 1SZN) with enzymes encoded by *aglA* and *aglB* genes. Amino acids of α-*N*-acetylgalactosaminidases and α-galactosidases, which are responsible for substrate specificity, are highlighted in green and yellow respectively; amino acids common to both enzymes are cyan. Secondary structure was assigned with ProCheck (Laskowski et al. 1993).

the *aglA* gene in the *A. niger* genome. The overall sequence coverage was 67% and thus led to an unambiguous assignment (Figure 1B).

Alignment of the primary sequences of the two enzymes encoded by the *aglA* and *aglB* genes with the primary sequences of the available 3D structures determined by X-ray diffraction en-

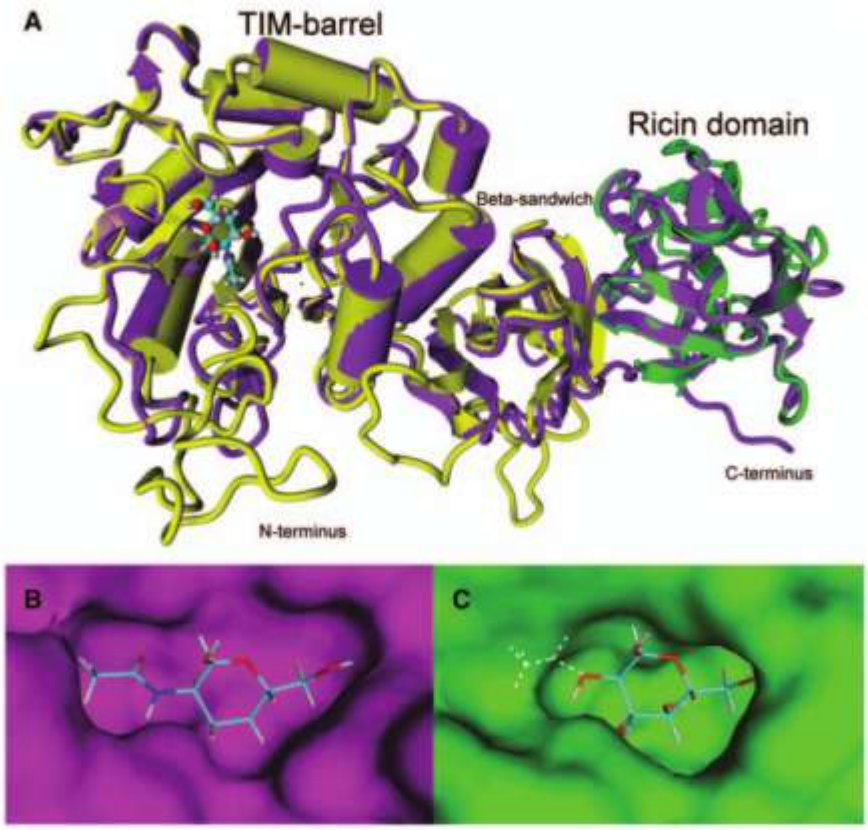


Fig. 3. Structure of the α-*N*-acetylgalactosaminidase from *Aspergillus niger*. (**A**) Overall fold shows a TIM-barrel with the active site at the N-terminus, a small domain of eight antiparallel β-strands packed in β-sandwich in the middle, and a ricin-like domain on the right. The generated model (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.

abled us to construct a model of each of these enzymes (Figures 2 and 3). As a result, the protein encoded by the *aglB* gene showed more than 50% sequence identity to the available solved α-galactosidase crystal structures (Golubev et al. 2004) and only 33% sequence identity to the available α-*N*-acetylgalactosaminidase crystal structure (enzyme from chicken, Garman et al. 2002). In contrast, the enzyme encoded by the *aglA* gene is characterized by 29% sequence identity to the available crystal structure of the α-galactosidase from *Trichoderma reesei* and 35% sequence identity to the α-*N*-acetylgalactosaminidase from chicken. Recently, the structure of human α-*N*-acetylgalactosaminidase was determined (Clark and Garman 2009). Our primary structure shows an equal sequence identity of 35% to this new human α-*N*-acetylgalactosaminidase and a similarity of 50%. The similarity is thus comparable to that between *aglA* and α-*N*-acetylgalactosaminidase from chicken described above. A phylogenetic analysis reveals a comparable evolutionary distance to chicken α-*N*-acetylgalactosaminidase and to human α-*N*-acetylgalactosaminidase for the *aglA* gene (0.764 and 0.768, respectively). α-*N*-Acetylgalactosaminidases from chicken and humans have a very close 3D structure with a root mean square deviation of only 0.54 Å for Cα over 387 aligned residues, with the active site not only fully conserved in the primary sequence (see alignment in Figure 2) but also in 3D. Significant differences between both crystal structures (chicken and human) and our structural model of *aglA* could only be found in the loop regions on the enzyme surface, where our model has three longer loops (Gly132-Glu145, Ala176-Tyr185, and Asp201-Ala210). Surprisingly, the three loops together cover a similar 3D space as the one large loop seen in the crystal structure of α-galactosidase from *T. reesei* (also present in our α-galactosidase model of gene *aglB*), indicating that occupation of this specific space could be a feature of all fungal α-galactosidases and α-*N*-acetylgalactosaminidases. In conclusion, since both structures are very similar over the entire length and identical within the active site, the inclusion of the

Table I. Changes in secondary structure as calculated by YASARA of the initial homology models for enzymes encoded by genes *aglA* and *aglB* during the molecular dynamics refinement

Model		Secondary structure content (%)			
		α-Helix	β-Sheet	Turn	Coil
aglA	Before refinement	29.2	27.0	11.8	32.0
	After refinement	28.8	26.5	10.5	34.2
aglB	Before refinement	26.3	23.7	16.0	34.0
	After refinement	26.1	24.4	16.6	32.9

recently released crystal structure of the human enzyme would not change the modeling results, nor would it affect the 3D arrangement in the active site. Therefore, at present we can state that the closest probable structure for the enzyme encoded by gene *aglB* is unambiguously α-galactosidase, whereas for the enzyme encoded by the *aglA* gene the primary sequence analysis shows a slightly higher probability that its 3D structure is closer to that of α-*N*-acetylgalactosaminidase (Figure 2).

A BLAST domain search identified two domains in the enzyme encoded by the *aglA* gene (NCBI reference sequence: XP_001390845.1): a melibiase domain at the N-terminal and a ricin-like domain at the C-terminal end (Figure 3). Comparison of the *aglA* gene with amino acid sequences from non-redundant sequence databases identified α-*N*-acetylgalactosaminidase from *Acremonium* sp. No. 413 (Ashida et al. 2000) as another example of glycosidase containing a ricin-like domain. Other similar sequences either do not contain a ricin-like domain or have a sequence identity of less than 40%. The enzyme encoded by the *aglB* gene (GenBank: CAK44445.1) contains only a melibiase domain.

The theoretical correctness of the generated models is an important issue in homology modeling. Analysis of the probability that the given primary sequence adopts the predicted fold using so-called *z* scores calculated by comparing the conformation energies with ProSA demonstrates the principal correctness of our two initial structural models with *z* scores of −7.35 (*aglA*) and −7.88 (*aglB*), while the local quality scores point to slightly problematic regions for *aglA* (residues 220–230 and 300–320) and

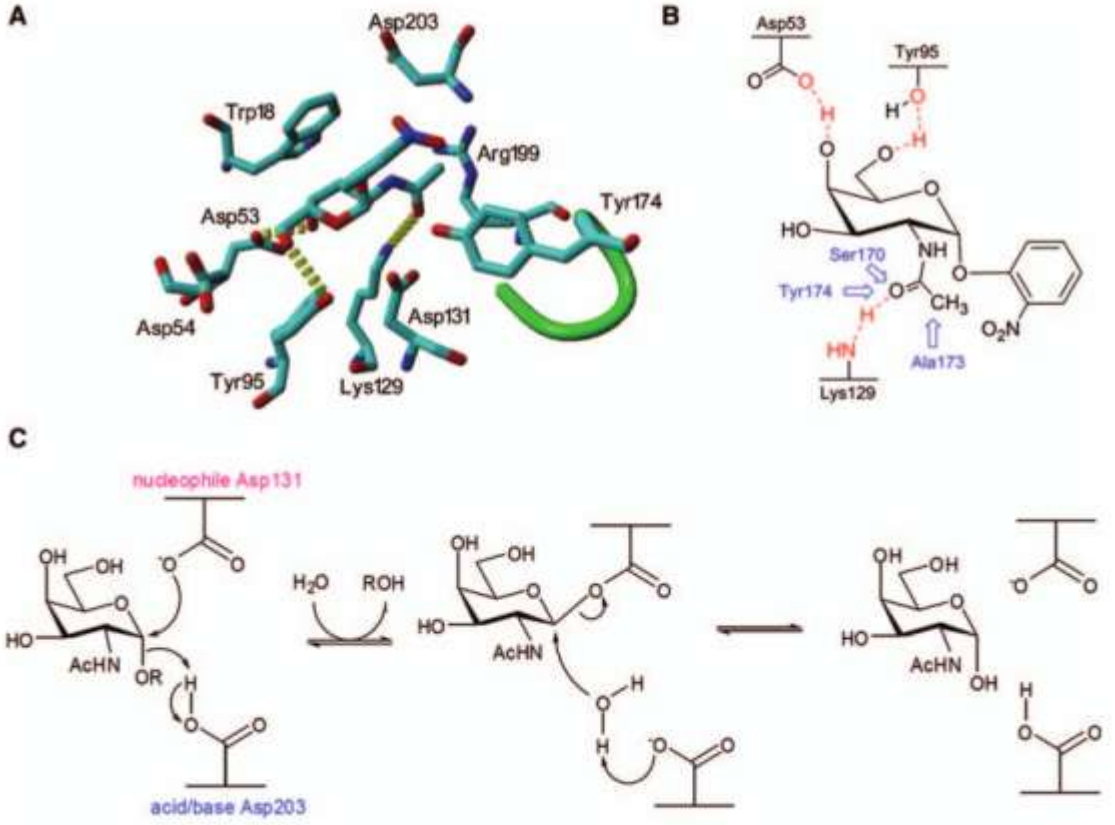


Fig. 4. (**A**) Active site amino acids identified for *aglA* enzyme with bound 2-NP-α-GalNAc after 10 ns of MD with so-called *N*-acetyl recognition loop (green), which extends the binding pocket of *aglA* enzyme so that it can accept the amino group at the C2-atom. (**B**) Scheme of hydrogen bonds, created by 2-NP-α-GalNAc with *aglA* enzyme. (**C**) Scheme of catalytic mechanism proposed for *aglA* enzyme, where Asp203 acts as acid/ base and Asp131 is responsible for nucleophilic attack at C1.

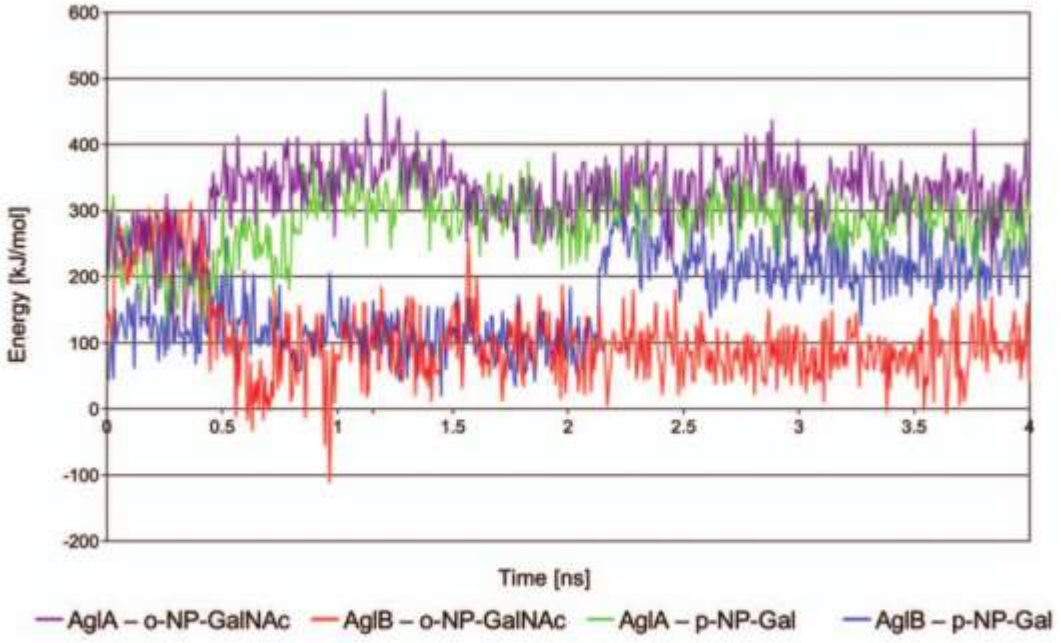


Fig. 5. Binding energy of substrate–enzyme complex during 4 ns of MD.

for *aglB* (residues 200–220 and 340–350). This was probably due to the low resolution in the corresponding template crystal structures, and, therefore, we must treat these regions as poorly resolved at this initial stage. Improvement was reached by a refinement using molecular dynamics simulation in explicit solvent with both structural models reaching equilibrium conformations after 2 ns of simulation. The root mean square deviation of Cα atoms reached a plateau with values the difference of which from those of the initial structure was less than 0.25 nm. During the refinement, we observed only minor changes (around 1%) in the secondary structure content of the model (Table I).

Model analysis discovered significant differences between the amino acid sequences in the active site of the enzymes encoded by the *aglA* and *aglB* genes. This result could explain the difference in substrate specificity between both enzymes (Figure 2). Reasonable structural models stable in molecular dynamics simulations in explicit solvent can be built only using α-galactosidase as a template for *aglB* and α-*N*-acetylgalactosaminidase as a template for *aglA*. Both models, *aglA* and *aglB*, have similar folding: the main domain is a TIM-barrel, containing the active site at the N-terminus of the β-barrel; this structure is followed by eight antiparallel β-strands folded into a β-sandwich. The α-*N*-acetylgalactosaminidase isolated from *A. niger* CCIM K2 has an additional C-terminal ricin-like domain, making this enzyme somewhat larger on sodium dodecyl sulfate polyacrylamide gel electrophoresis (57 vs 45 kDa). Its role is unknown, but its folding might indicate a potential interaction with monosaccharides, such as galactose or lactose (Figure 3A).

Automated docking of the corresponding substrates into the active sites of both enzymes followed by simulations of the resulting enzyme–substrate complexes enabled us to analyze the stability of the ligand–protein interaction in aqueous solution at room temperature. Although equilibration was already reached after 2 ns, the simulations were continued for 10 ns in total to demonstrate the stability of the whole complex. Amino acid residues in the active site of *aglA* within 0.3 nm of 2-NP-α-GalNAc at the end of the simulation are shown in Figure 4A and B. A docking attempt of 2-NP-α-GalNAc into the active site of *aglB* led to a fast repulsion of the 2-acetamido group due to Trp205 blocking the space needed to accommodate the 2-acetamido group properly and the side chain of Asp222 flipping in and pushing the *N*-acetyl-group completely out of the active side towards water. As a result, the distance between the oxygen at the aglycon of the ligand and the oxygen of the catalytic residue, which acts as an acid to hydrolyze the C1-O bond, increased during the first 1.5 ns of molecular dynamics (MD) to an average value of 0.67 nm, a distance too far for the catalytic residue to attack the C1-O bond of the ligand. The same distance measured for *aglB* with docked 4-NP-α-Gal after 10 ns of MD was 0.35 nm, allowing appropriate positioning of the substrate for eventual reaction. These results well demonstrate that the enzyme encoded by gene *aglB* is a pure α-galactosidase unable to perform any α-*N*-acetylgalactosaminidase enzymatic function. Figures 3B and C show the active sites of both enzymes with bound substrates in a surface representation.

The binding energies of the respective substrates during the MD simulations were calculated by Yasara for the first 4 ns of

Fig. 6. Multiple sequence alignment of galactosidase/galactosaminidase homologs in the genome of *Aspergillus niger*, shaded according to their sequence identity (intensity depends on the frequency at same position in other aligned sequences). The region containing the *N*-acetyl recognition loop in α-*N*-acetylgalactosaminidases is marked with a black rectangle. Other potential active site residues in *aglB* are marked with a dot.

MD (Figure 5). The complexes reached equilibration after 1.5 ns with an average binding energy for 2-NP-α-GalNAc docked into *aglA* of 324 kJ/mol and only slightly lower values for 4-NP-α-Gal, 287 kJ/mol. With *aglB*, the initial complexes have similar values for both substrates; however, after 2 ns, we can see a clear separation in favor of the complex with 4-NP-α-Gal, which reaches a higher average binding energy of 215 kJ/mol, in contrast to the complex with 2-NP-α-GalNAc with average values below 100 kJ/mol. When estimating the binding energy this way, positive values are used, and higher numbers indicate better binding. Positive numbers result from the calculation in which the sum of solvation and potential energies of the complex is subtracted from the sum of these energies of the single components. As this method according to our experience shifts the absolute values to higher numbers and thus overestimates the binding energies when compared with experimental values, we also give the binding energies calculated with Autodock, which should be closer to experimentally determined ΔG values: −24 kJ/mol for the *aglA*–2-NP-α-GalNAc complex and −18 kJ/mol for aglA–4-NP-α-Gal; −17 kJ/mol for *aglB*–4-NP-α-Gal. The *aglB*–2-NP-α-GalNAc complex was not ranked as a possible docked conformation by Autodock, indicating that 2-NP-α-GalNAc cannot be properly accommodated in the active site, and the bound position after the molecular dynamics simulation is not stable. Indeed, visual inspection confirms that the substrate is already partially released to the solvent at the end of the simulation. Thus, our computational modeling results corroborate the hypothesis that the enzyme encoded by gene *aglA* is an α-*N*-acetylgalactosaminidase that exhibits a dual enzyme activity and is able to hydrolyze both 2-NP-α-GalNAc and 4-NP-α-Gal, although at a different rate.

Discussion

The structural models constructed in this study were indispensable in allowing to explain the inability of α-galactosidase *aglB* from *A. niger* to accomodate the substrate containing a 2-acetamido group in the active site in a stable and proper position. Two reasons why the α-galactosidase *aglB* is unable to hydrolyze *o*NP-α-GalNAc have emerged: the lower binding affinity and the sterical hindrance connected with positioning of the C1-O bond close to the catalytic residue that makes it impossible for the hydrolytic reaction to take place. The computer modeling experiments enabled us to assign the protein possessing the dual activity to the enzyme encoded by gene

aglA. This is a completely new view of this gene since a possible α-*N*-acetylgalactosaminidase activity has never been mentioned in the literature before. Our experimental results show that this protein exhibits dual enzyme activity and is able to hydrolyze 2-NP-α-GalNAc and 4-NP-α-Gal with 10 times better cleavage of 2-NP-α-GalNAc than that of 2-NP-α-Gal. This contradicts with its original assignment as an α-D-galactopyranosidase belonging to GH family 27. Substrate docking into the structural model of this enzyme confirmed similar binding energies for both substrates with a clear preference for the α-*N*-acetyl-D-galactosaminide over the galactoside (Figure 3B and C), thus providing evidence that this gene from *A. niger* does not encode exclusively an α-galactosidase type A but that rather one combined with a fully functional α-*N*-acetylgalactosaminidase.

Figure 3B shows the binding pocket of *aglA* with bound α-*N*-acetylgalactosamine, in comparison to the binding pocket of *aglB* with bound α-galactose in Figure 3C. The enlarged binding pocket in *aglA*, which is able to accommodate an C2 acetamido group, is a result of the amino acid sequence Ser170, Ala171, Pro172, Ala173, and Tyr174, which forms a longer loop following β5 and connecting to α5 (Figure 2). Instead of Trp205, which blocks the part of the binding pocket accommodating the acetamido group in α-galactosidases, amino acid residues Ser170, Ala173, and Tyr174 (cf. Figure 4A and B) on this "*N*-acetyl recognition loop" create an open space that becomes part of the enlarged binding pocket of α-*N*-galactosaminindase. The presence of this loop explains the fact that the attempt to model *aglA* taking α-galactosidase as a template did not lead to a stable enzyme but induced distortions in the active site during molecular dynamics simulations. The multiple sequence alignment of the α-*N*-acetylgalactosaminindase/α-galactosidase homologs in *A. niger* (Figure 6) identified conserved residues at positions corresponding to the active site of *aglB* for all hypothetical proteins, with the exception of a tyrosine and cysteine in An14g01800 instead of Trp17 and Asp55 in *aglB*. *AglA* is the only homolog having the above described "*N*-acetyl recognition loop". Lacking this loop, all other found potential α-galactosidases in the *A. niger* genome could have α-galactosidase acitivity but definitely not α-*N*-galactosaminindase activity. This is in accordance with our experimental findings demonstrating that the other α-galatosidases isolated from *A. niger* did not show dual activity but are pure α-galactosidases.

The fact that the enzyme, which possesses a pocket capable of accommodating the *N*-acetyl group, shows dual activity and is able to degrade substrates having an OH group in the same position, excludes the possibility that the *N*-acetyl recognition loop would be directly involved in the reaction mechanism adopted by the glycoside hydrolase family 27. Instead, the *N*-acetyl recognition loop that is highly conserved in α-*N*-acetylgalactosaminidases functions to structurally recognize and accommodate the *N*-acetyl group, but does not have additional contacts or interactions with the substrate, thus, the enzyme can bind equally well to 4p-NP- α-Gal.

Therefore, we propose that the α-*N*-acetylgalactosaminidase/α-galactosidase encoded by *aglA* utilizes a classical double inversion reaction mechanism (Figure 4C) analogous to homologous α-galactosidases (Garman et al. 2002; Golubev et al. 2004) and that this mechanism is the same for both of its activities, e.g., α-*N*-acetylgalactosaminidase and α-galactosidase.

Materials and methods

Enzyme assay

Enzyme activity was tested as described previously (Weignerová et al. 2008). All data were measured in triplicate.

Purification procedure

The protein was purified as described previously (Weignerová et al. 2008). 2D electrophoresis was performed using a dry polyacrylamide gel strip, gradient pH 4.0–5.0 (GE Health Bio-Sciences AB, Uppsala, SE).

N-Terminal sequencing

N-Terminal sequencing was performed with the enzyme purified in a C-4 reversed-phase column (Vydac C-4, Dionex, Sunnyvale, CA, USA) or electrotransferred onto PVDF blots using a Procise 491 Protein Sequencer (Applied BioSystems-Life Technologies, Carlsbad, CA, USA) using the Pulsed Liquid program. For CNBr cleavage, the dried enzyme was incubated in 0.15 M CNBr (Aldrich, Buchs, CH) in 70% formic acid for 24 h in the dark, and the fragmented peptides were separated on a reversed-phase column (Vydac C-18) equilibrated in 0.1% trifluoroacetic acid (TFA, LiChrosolv, Merck, Darmstadt, DE) and eluted with acetonitrile gradient (0–70% acetonitrile).

Sample preparation for mass spectrometry

The spot of interest was excised from the 2D gel. After complete destaining, the protein was reduced with 50 mM Tris-carboxyethyl phosphine (Fluka, Buchs, CH) for 15 min at 80°C and alkylated using iodoacetamide (Sigma-Aldrich, Buchs, CH) for 40 min in the dark. The gel was then washed with deionized water, shrunk by dehydration in acetonitrile, and re-dissolved in deionized water again. The supernatant was removed and the gel partly dried in a SpeedVac concentrator (Savant, Holbrook, NY, USA). Gel pieces were then reconstituted in deglycosylation buffer pH 5.5 containing, 50 mM citrate buffer and 1000 units of Endo H (New England BioLabs, Ipswich, MA, USA). After overnight deglycosylation, the gel pieces were washed with deionized water, shrunk by dehydration in acetonitrile, and re-dissolved in deionized water again. Cleavage buffer containing 50 mM 4-ethylmorpholine acetate pH 8.2, 10% acetonitrile, and sequencing-grade trypsin (5 ng/μL, Promega, Madison, USA) was added to the gel pieces. After overnight digestion at 37°C, the resulting tryptic peptides were extracted using 30% acetonitrile in 1% acetic acid.

Mass spectrometry and data analysis

The peptide sample was dissolved in 200 μL of 0.1% TFA and then desalted in a peptide MacroTrap (Michrom Bioresources, Auburn, CA, USA), eluted with 80% MeCN/0.5% TFA. The purified peptides were dried up on SpeedVac and redissolved in 20 μL of 0.1% TFA/5% MeCN. Five microliters of sample was loaded into a reversed-phase Magic C_{18} column (Michrom Bioresources, Auburn, USA; column size, 150 × 0.2 mm I.D., particle size 5 μm; pore size 200 A) and separated with water/MeCN gradient using an Ultimate 1000 HPLC system (Dionex, Amsterdam, NL). The mobile phase consisted of the following solvents: 0.2% formic acid in water (A) and 0.16% formic acid in MeCN (B). Peptides were eluted under gradient conditions as

follows: 0–1 min, 1% B; 5 min, 15% B; 30 min, 40% B; 35 min, 95% B; 40 min, 95% B. Flow rate was 4 μL/min after splitting, and the column was kept at room temperature. The capillary column was directly connected to the mass analyzer. The mass spectrometric experiment was performed on a commercial APEX-Qe Fourier Transform Mass Spectrometry (FTMS) Lys/Arg instrument equipped with a 9.4-T superconducting magnet and Combi ESI/MALDI ion source (Bruker Daltonics, Billerica, MA, USA). Mass spectra were obtained by accumulating ions in the collision hexapole and running the quadrupole mass filter in non mass-selective (rf-only) mode so that a broad range of ions (300–2500 m/z) were passed to the FTMS analyzer cell. The accumulation time in the collision cell was set at 0.5 s, the cell was opened for 4000 μs, and 256 experiments were collected for one liquid chromatography run, where one experiment consisted of the accumulation of five spectra. The acquisition time was set to 512 k points at m/z 300 a.m.u. The instrument was externally calibrated using triply and doubly charged ions of angiotensin I and quadruply, quintuply, and sextuply charged ions of bovine insulin. This calibration typically results in a mass accuracy below 2 ppm.

The acquired spectra were apodized and processed using a sin function with one zero-fill and run through the data-reducing macro designed by Kruppa et al. (2003). A list of unique monoisotopic masses was generated for each sample. The output of the macro (the list of unique monoisotopic masses) was matched to a theoretical library of α-galactosidase tryptic peptides to within 2 ppm. The library was created using the program ASAP (Automated Spectrum Assignment Program) (Young et al. 2000; Kellersberger et al. 2004). The trypsin specificity was set to Lys/Arg, and the degree of incomplete digestion was up to 1. Also, the algorithm was allowed to include a single oxidation of methionin (+15.9949 a.m.u.) and modification of glycosylated asparagines by *N*-acetylglucosamine (+203.0794 a.m.u.).

Molecular modeling

Molecular models were generated by restrain-based homology modeling followed by a structural refinement to test their robustness using molecular dynamics simulations in explicit solvent. In the first step, homologues were identified by BLAST search and the top scoring extracted from the Protein Data Bank (Berman et al. 2000): 1KTB, 1UAS, 1R46, 1SZN—for the melibiase domain; 2AAI, 1RZO, 1V6X—for the ricin domain. Different combinations of templates were used for multiple sequence and structural alignments. The multiple structure-based sequence alignments were calculated with T-Coffee (http://www.tcoffee.org). Structural alignments of known 3D structures were performed with SHEBA in Yasara 7.5.14 (Jung and Lee 2000). 3D models comprising non-hydrogen atoms were built and examined using Modeller 9.1 package (Sali and Blundell 1993) and checked with ProSA (Wiederstein and Sippl 2007). The models were refined in YASARA (Krieger et al. 2004) by energy minimization followed by 2 ns of MD simulation in water (pH 7, TIP3P water model) with periodic boundary conditions. Substrates were built in YASARA, force field parameters were assigned using the AutoSMILES approach. The corresponding bond, angle, and torsion potential parameters are taken from the General AMBER force field. Initial ligand positions for docking experiments were determined by comparison with homologous crystal structures of α-galactosidase (Golubev et al. 2004) and α-*N*-acetylgalactosaminidase (Garman and Garboczi 2004), respectively, which were co-crystallized with the ligands. The exact position of ligands was set by Autodock 4.0 (Goodsell and Olson 1990) using local docking with the Lamarckian Genetic Algorithm, grid space 0.375. Substrate–enzyme complexes were refined by minimization in a YAMBER 2 force field (Krieger et al. 2004) and periodic boundary conditions. Sodium ions were iteratively placed at the coordinates with the lowest electrostatic potential until the cell was neutral. Molecular dynamics simulations were run in YASARA, using a multiple time step of 0.7 fs for intramolecular and 1.4 fs for intermolecular forces, a 0.78 nm cutoff for Lennard Jones forces and the direct space portion of the electrostatic forces, which were calculated using the Particle Mesh Ewald method (Essman et al. 1995) with grid spacing 0.1 nm, fourth-order B-splines, and a tolerance of 10^{-4} for the direct space sum. The simulation of their interaction was run in the following NPT ensemble: constant temperature (298 K), pressure and number of particles. The evaluation of ligand–enzyme complexes in time was analyzed on the basis of geometry and energy parameters.

Interaction energies were calculated using a similar method to that applied earlier for substrate–hexosaminidase complexes (Ettrich et al. 2007), considering the internal energy obtained with the specified force field, as well as the electrostatic energy and the entropy cost of fixing the ligand implicitly including Van der Waals solvation energy (Bultinck et al. 2003). The solvation energy was calculated using the boundary fast method implemented in YASARA. The boundary between solvent (dielectric constant 80) and solute (dielectric constant 1) was formed by the latter's molecular surface, constructed with a solvent probe radius of 1.4 Å and the following radii for the solute elements: polar hydrogens 0.32 Å, other hydrogens 1.017 Å, carbon 1.8 Å, oxygen 1.344 Å, nitrogen 1.14 Å, and sulfur 2.0 Å. The solute charges were assigned based on the YASARA 2 force field, using GAFF/AM1BCC for the ligands. An estimate of the entropy cost of exposing 0.01 nm^2 to solvent was calculated as follows: E_{entr} = S_{SAS}*solvation entropy, where S_{SAS} is the solvent accessible area, and the solvation entropy characterizes the entropy cost of exposing 0.01 nm^2 of surface and can vary. For protein structures, the solvation entropy is usually estimated to be approximately 0.65 $kJmol^{-1}A^{-2}$. It is almost impossible to calculate this entropy cost accurately, but this is fortunately not needed, since it mainly depends on characteristics that are constant during the simulation (ligand and protein size, side chains on the surface, etc.) and thus is a constant factor. Binding energies calculated in this way might therefore be shifted by an unknown amount that depends on the protein; however, their relative values are correct. The more positive interaction energy, the more favorable is the interaction in the context of the chosen force field. Ligand–enzyme complexes after 10 ns of MD were used for measuring binding energies by Autodock.

Acknowledgments

Support by the Academy of Sciences of the Czech Republic (AVOZ50200510, AVOZ60870520), the Ministry of Education of the Czech Republic (LC 06010, LC 7017,

MSM6007665808, MSM21620808), and by the Czech Science Foundation grants P207/10/0321 to L.W., P207/10/1934 to R.E, P207/10/1040 to P.N., and 303/09/0477 and 305/09/H008 to K.B. is acknowledged. Access to METACentrum supercomputing facilities was provided under the research intent MSM6383917201.

Conflict of interest statement

None declared.

Abbreviations

2-NP-α-GalNAc, 2-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside; 4-NP-α-Gal, 4-nitrophenyl α-D-galactopyranoside; *aglA*, α-galactosidase variant A in the *A. niger* genome; *aglB*, α-galactosidase variant B in the *A. niger* genome; GH, glycosylhydrolase; MD, molecular dynamics simulation; TFA, trifluoroacetic acid.

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.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res*. 28:235–242.

Bultinck P, De Winter H, Langenaeker W, Tollenare J. 2003. *Computational medicinal chemistry for drug discovery*. CRC Press.

Clark NE, Garman SC. 2009. The 1.9 Å structure of human α-*N*-acetylgalactosaminidase: the molecular basis of Schindler and Kanzaki diseases. *J Mol Biol*. 393:435–447.

Essman U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *J Chem Phys*. 103:8577–8593.

Ettrich R, Kopecký V, Hofbauerová K, Baumruk V, Novák P, Pompach P, Man P, Plíhal O, Kutý M, Kulik N, et al. 2007. Structure of the dimeric N-glycosylated form of fungal beta-*N*-acetylhexosaminidase revealed by computer modeling, vibrational spectroscopy, and biochemical studies. *BMC Struct Biol*. 17:32.

Garman SC, Garboczi DN. 2004. The molecular defect leading to Fabry disease: structure of human α-galactosidase. *J Mol Biol*. 337:319–335.

Garman SC, Hannick L, Zhu A, Garboczi DN. 2002. The 1.9 Å structure of α-*N*-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. *Structure*. 10:425–434.

Golubev AM, Nagem RAP, Brandão Neto JR, Neustroev KN, Eneyskaya EV, Kulminskaya AA, Shabalin KA, Savel'ev AN, Polikarpov I. 2004. Crystal structure of α-galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism. *J Mol Biol*. 339:413–422.

Goodsell DS, Olson AJ. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct Funct Genet*. 8:195–202.

Hart DO, He S, Chany CJ 2nd, Withers SG, Sims PF, Sinnott ML, Brumer H 3rd. 2000. Identification of Asp-130 as the catalytic nucleophile in the main alpha-galactosidase from *Phanerochaete chrysosporium*, a family 27 glycosyl hydrolase. *Biochemistry*. 39:9826–9836.

Hsieh HY, Calcutt MJ, Chapenan LF, Mitra MD, Smith S. 2003. Purification and characterization of a recombinant α-*N*-acetylgalactosaminidase from *Clostridium perfringens*. *Protein Expression Purif*. 32:309–316.

Hsieh HY, Mitra M, Wells DC, Smith DS. 2000. Purification and characterization of α-*N*-acetylgalactosaminidase from *Clostridium perfringens*. *Life*. 50:91–97.

Jung J, Lee B. 2000. Protein structure alignment using environmental profiles. *Protein Eng*. 13:535–543.

Kellersberger KA, Yu E, Kruppa GH, Young MM, Fabris D. 2004. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal Chem*. 76:2438–2445.

Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. 2004. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*. 57:678–683.

Kruppa GH, Schoeniger J, Young MM. 2003. A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom*. 17:155–162.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK – a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 26:283–291.

Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, Spence J, Nudelman E, Levary SB, White T, et al. 2007. Bacterial glycosidases for the production of universal red blood cells. *Nature Biotechnol*. 25:454–464.

Ly HD, Howard S, Shum K, He S, Zhu A, Withers SG. 2000. The synthesis, testing and use of 5-fluoro-α-D-galactosyl fluoride to trap an intermediate on green coffee bean α-galactosidase and identify the catalytic nucleophile. *Carbohydr Res*. 329:539–547.

Olsson ML, Clausen H. 2008. Modifying the red cell surface: towards an AB0-universal blood supply. *Br J Haematol*. 140:3–12.

Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, et al. 2007. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol*. 25:221–231.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 234:779–815.

Weignerová L, Filipi T, Manglová D, Křen V. 2008. Induction, purification and characterization of a α-*N*-acetylgalactosaminidase from *Aspergillus niger*. *Appl Microbiol Biotechnol*. 79:769–774.

Weignerová L, Simerská P, Křen V. 2009. α-Galactosidases and their application in biotransformation. *Biocat Biotrans*. 27:79–89.

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 35:407–410.

Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. 2000. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci USA*. 97:5802–5806.

Appendix 2

Mrazek H., Benada O., Man P., Vanek O., Kren V., Bezouska K., Weignerova L.

Facile production of A*spergillus niger* α-*N*-acetylgalactosaminidase in yeast.

Submitted for Protein expression and purification (under review).

Running title: *A. niger* α-*N*-acetylgalactosaminidase expressed in yeast (max. 54 characters and spaces)

Facile production of *Aspergillus niger* α-*N*-acetylgalactosaminidase in yeast

Hynek Mrázek[a,c], Oldřich Benada[d], Petr Man[d], Ondřej Vaněk[c], Vladimir Křen[b], Karel Bezouška[a,c] and Lenka Weignerová[b,c*]

[a]*Laboratory of Protein Architecture, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague, CZ*

[b]*Center for Biocatalysis and Biotransformation, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20, Prague, CZ*

[c]*Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 2030, 128 40 Prague, CZ*

[d]*Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague 4, CZ*

α-*N*-Acetylgalactosaminidase (α-GalNAc-ase; EC.3.2.1.49) is an exoglycosidase specific for the hydrolysis of terminal α-linked *N*-acetylgalactosamine in various sugar chains. The cDNA corresponding to the α-GalNAc-ase gene was cloned from *Aspergillus niger*, sequenced, and expressed in the yeast *Saccharomyces cerevisiae*. The α-GalNAc-ase gene contains an open reading frame which encodes a protein of 487 amino acid residues. The molecular mass of the mature protein deduced from the amino acid sequence of this reading frame is 54 kDa. The recombinant protein was purified to apparent homogeneity and biochemically characterized (p*I* 4.4, K_M 0.56 mmol/l for 2-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside, and optimum enzyme activity was achieved at pH 2.0-2.4 and 50-55 °C). Its molecular weight was determined by analytical ultracentrifuge measurement and dynamic light scattering. Our experiments confirmed that the recombinant α-GalNAc-ase exists as two distinct species (70 kDa and 130 kDa) compared to its native form, which is purely monomeric. *N*-Glycosylation was confirmed at six of the eight potential *N*-glycosylation sites in both wild type and recombinant α-GalNAc-ase.

*Corresponding author. Address: Laboratory of Biotransformation, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague 4, Czech Republic

Tel. +420 296 442766, Fax. +420 296 442569

α-*N*-Acetylgalactosaminidase is an exoglycosidase (α-GalNAc-ase, EC 3.2.1.49, GH family 27, Carbohydrate Active Enzymes database, http://www.cazy.org/) that selectively cleaves terminal α-linked *N*-acetylgalactosamine units in various sugar chains. Human α-GalNAc-ase plays an important role in the catabolism of glycoconjugates and lack of it may cause Schindler and Kanzaki (7) disease, resulting in neurodegenerative pathologies (9, 27, 28). α-GalNAc-ase purified from chicken (12), *Clostridium perfringens* (17), rice (11) and *Elizabethkingia meningosepticum* (18, 34) have been shown to efficiently cleave the terminal α-terminal GalNAc from blood group A erythrocytes, resulting in a blood group H(0) epitope structure (35).

Whereas α-galactosidases, which are widely distributed in microorganisms, plants, animals and humans, are available in various recombinant forms (30), the availability of α-GalNAc-ases is somewhat limited. These enzymes have been so far cloned from human (26, 27), chicken (35, 36), *Clostridium perfringens* (5) and *Acremonium sp.* (2). The first crystal structure of α-GalNAc-ase from human was reported recently (7).

Recently a large screening for α-*N*-acetylgalactosaminidase activity in filamentous fungi has been completed leading to the identification of the promising enzyme producer *Aspergillus niger* CCIM K2. The extracellular enzyme from this source was purified and fully characterized (31). This α-GalNAc-ase also catalysed the reverse hydrolysis to yield α-D-Gal*p*NAc-(1→6)-D-Gal*p*NAc, α-D-Gal*p*NAc-(1→3)-D-Gal*p*NAc and α-D-Gal*p*NAc-(1→6)-D-Gal*f*NAc and also the Tn antigen (Gal*p*NAc-α-*O*-Ser/Thr) and its *N*-(*tert*-butoxycarbonyl)-protected derivatives (32, 33).

Enzyme activity measurements and molecular modelling clearly demonstrated the dual specificity of *A. niger* α-GalNAc-ase, this enzyme was able to hydrolyze both *o*-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (*o*NP-α-GalNAc) and *p*-nitrophenyl α-D-galactopyranoside (*p*NP-α-Gal) (only 10 % of the activity with *o*NP-α-GalNAc) as its substrate. Evidence that the α-galactosidase type A gene *aglA* from *A. niger* in fact encodes a fully functional α-GalNAc-ase was presented recently (6, 14, 16).

In this report we describe the expression of α-GalNAc-ase from *A. niger* in a *S. cerevisiae* strain, its purification and characterization, and the comparison of the recombinant and wild enzyme. To our best knowledge, α-GalNAc-ase from non pathogenic source working in the acidic to neutral pH has not been described yet.

MATERIAL AND METHODS

Microorganisms. The native α-GalNAc-ase from *A. niger* CCIM K2 used in this work was prepared as described previously (30). The production strain is deposited in the Culture Collection of the Academy of Sciences of the Czech Republic, Institute of Microbiology, Prague, CZ. Bacterial strain *E.coli* BL21(DE3) (Strategene, USA) was cultivated on LB medium composed of 1% peptone, 0.5% yeast extract and 1% NaCl, pH 7.4 at 37˚C. *S.cerevisiae* W303-1A (MATa *ura3-1, leu2-3, 112 trp1, his3-11, 15 ade2-1*) was purchased from Invitrogen (USA). The expression strain, *S. cerevisiae*, was grown on YPD medium composed of 2% glucose, 2% peptone, and 1% yeast extract, pH 6.5 at 28˚C. For the selection of the transformants, SC plates were used composed of 2% glucose and 0.67% yeast nitrogen base without amino acids (Difco, USA) (1).

Isolation of poly(A) RNA and construction of a cDNA library. A submerged culture of *A. niger* was cultivated for 48 hours (30). Fresh mycelium (4 g, wet weight) was frozen with liquid nitrogen and homogenized. Total RNA was prepared using an RNeasy Plant Mini Kit (QIAGEN, DE). First-strand cDNA was synthesized directly from the total RNA by Moloney Murine Leukemia Virus reverse transcriptase (Invitrogen, USA), and the second strain was synthesized by DeepVent DNA polymerase (Invitrogen, USA). These cDNAs were purified using a DNA purification kit (Genomed, DE) (13).

Cloning the cDNA encoding α-*N*-acetylgalactosaminidase from *A. niger*. One set of primers was designed to amplify the α-GalNAc-ase gene. A sense primer (5´ATTGAATTCTTAGCCATCCCTCTCATAAAGACACGACTT-3´) containing the EcoRI site (underlined) and antisense primer (5´-TATTCTAGAATGGGTTTCAACAATTGGGCCCGC-3´) containing the XbaI restriction site (underlined) were designed on the basis of the known amino acid sequence (16). The fragments were amplified using Pfu DNA Polymerase (Invitrogen, USA) and a Thermal Cycler GeneE (Techne, UK). The PCR product (fragment 1462 bp) was amplified and ligated into the Bluescript II SK- vector (Invitrogen, USA) and transformed into *E. coli* BL21-DE3 (Strategene, USA). The resultant plasmid, Bluescript II SK/α-GalNAc-ase, was confirmed by restriction digestion and sequencing. The Bluescript II SK/α-GalNAc-ase plasmid was digested by *Eco*RI and *Xba*I and the gene product was ligated into pYES-2CT (Invitrogen, USA). The ligation reaction product was transformed into *S. cerevisiae* W303-1A (Invitrogen, USA) by electroporation (25μF, 22ns, 1.2kV, MicroPulserTM Electroporator, Bio-Rad, USA) and the transformants were selected on an SC plate supplemented with tryptofane, leucine, adenine and histidine, but without uracil (14).

Protein expression. Precultures were prepared by inoculating 15 ml of SC selective medium containing 2% glucose, tryptophan, leucine, adenine and histidine, without uracil with a single colony of S. *cerevisiae* W303-1A containing the pYES-2CT/α-GalNAc-ase construct. After overnight growth at 30 ˚C and 220 rpm, the pellets were collected by centrifugation, resuspended in 200 ml of SC selective medium as above but with the 2% glucose replaced by 2% galactose, and the culture was incubated at 30 ˚C and 220 rpm for 3 days. After centrifugation, the yeast pellet was resuspended in lysis buffer composed of 50 mM citrate phosphate (pH 3.5), 5% glycerol, 1 mM PMSF and 1mM dodecyl maltoside. An equal volume of acid-washed glass beads (r=0.25-0.5 mm, Pierce, DE) was added. The cells were disrupted by vortexing for 30 s, followed by 30 s on ice (repeated four times for complete cell lysis). The mixture was centrifuged and the supernatant transferred into a fresh microcentrifuge tube for the enzyme activity assay (30).

Enzyme assay. Enzymatic activity was assayed as described previously (30).

Purification of the recombinant enzyme. A BioSys HPLC System with a UV and conductivity detector (Beckman-Coulter, USA) was used for all chromatography steps (10). The starting material (cell lysate) was diluted in 60 ml of solvent **A** (1 M NaCl in 50 mM citrate-phosphate buffer, pH 3.2) and the suspension was centrifuged in a Beckman-Coulter J2-21 centrifuge using the JA-25.50 rotor at 18000 rpm for 30 min at 4 °C. The extract was injected into a Phenyl-Sepharose HR column (2.6 × 10.6 cm, 2 ml/min, Merck, DE) equilibrated in solvent **A**. The enzyme was eluted from the column using a linear gradient (0-1 M NaCl, 90 min, 2 ml/min). Fractions containing α-GalNAc-ase activity were pooled and transferred into 20 mM citrate-phosphate, pH 4.5 (**C**, using an Amicon Ultra centrifugal filter device, cut off 10 kDa, Millipore, USA) and then injected into an S-Sepharose FF column (1.6 × 12.5 cm, 1 ml/min, Merck, DE) equilibrated in **C**. The enzyme was eluted from the column using a linear gradient (0-1 M NaCl, 60 min, 1 ml/min). Fractions containing α-GalNAc-ase activity were pooled, concentrated and resolved in a Superdex 200 10/300 GL gel filtration column (1.0 × 30 cm, 0.4 ml/min, Amersham Bioscience, AT), equilibrated with 50 mM sodium acetate buffer (pH 3.6). Pooled fractions containing α-GalNAc-ase were injected into a Mono P 5/200 GL (Amersham Bioscience, AT) chromatofocusing column equilibrated with 25 mM piperazine buffer (pH = 5.5, 40 ml) and with a pregradient (3.0 ml PB 74, 8:1, pH 4.0, Amersham Bioscience, AT) (30). The enzyme fractions from this step were stored at 4˚C.

Sedimentation velocity and sedimentation equilibrium measurements. Sedimentation analysis was performed using a ProteomeLabXL-I analytical ultracentrifuge equipped with an An50Ti rotor (BeckmanCoulter, USA). The protein (0.4 mg/ml) was

dialyzed against 50 mM sodium citrate buffer (pH 3.5, used also as a reference and sample dilution buffer). The sedimentation velocity experiment was carried out at 40,000 rpm and 20 °C, absorbance scans were recorded at 280 nm in 5 min intervals with 30 µm spatial resolution. Buffer density and α-GalNAc-ase partial specific volume were estimated in SEDNTERP 1.09 (www.jphilo.mailway.com). Data were analyzed with SEDFIT 12.1 (25). The sedimentation equilibrium experiment was performed with α-GalNAc-ase concentration of 0.13 mg/ml at 10-12-14-16-18-20-22,000 rpm at 4 °C. Absorbance data were collected at 280 nm by averaging 20 scans with 10 µm spatial resolution after 30 h (first scan) or 18 h (consecutive scans) of achieving equilibrium and were globally analyzed with SEDPHAT 8.2 using a non-interacting discrete species model (24).

Dynamic light scattering (DLS). Dynamic light scattering measurement of α-GalNAc-ase (15 µl of the sample used for sedimentation analysis) was carried out at a 90° scattering angle and 532 nm at 20°C on a Laser-Spectroscatter 201 (RiNa GmbH, DE) for 60 s and the signal accumulated from 60 individual measurements was evaluated with the software supplied by the manufacturer (22).

Electron microscopy. Negatively stained samples of wild-type and cloned enzyme complexes were prepared in parallel as follows: drops of enzyme solution in 50 mM citrate-phosphate buffer (pH 3.5), diluted to approx. 50 µg/ml were applied onto glow discharge-activated (4) carbon coated grids. After adsorption for 30 s, the grids were negatively stained with unbuffered 2% (wt/vol) uranyl acetate or unbuffered 0.5% (wt/vol) uranyl formate in water, respectively. The samples were analyzed in a Philips CM100 electron microscope (FEI, formerly Philips PEO, NL) at 80 kV. Digital images were recorded using a MegaView II slow-scan camera (Olympus, formerly SiS GmbH, DE) at a primary magnification of 64,000, giving a pixel size of about 1 nm. All image processing was done using AnalySis 3.2 software.

Analysis of the *N*-glycosylation sites. Analysis of the *N*-glycosylation sites was done as described previously (23).

RESULTS

Cloning and sequencing of native α-GalNAc-ase. Native extracellular α-GalNAc-ase was purified from a cultivation medium of *A. niger* CCIM K2 by three column chromatography steps as described previously (30). The final preparation gave a single protein band of 76 kDa on SDS-PAGE. To determine the *N*-terminal amino acid sequence, the purified enzyme was subjected to SDS-PAGE, electroblotted to a PVDF membrane and the

76 kDa protein band was then analyzed by automatic Edman degradation, yielding the sequence MGFNNWAR. The *C*-terminal amino acid sequence was identified as QMKSCLYE using mass spectrometric peptide mapping. Based on the experimentally determined amino acid sequence, primers were designed and used for PCR amplification.

A PCR product 1461 bp in size was amplified from *A. niger* cDNA, obtained as described above. The α-GalNAc-ase gene contains an open reading frame with an ATG initiation codon.

Expression and purification of recombinant α-GalNAc-ase. Firstly, the α-GalNAc-ase gene was ligated into the pET-30a$^+$ vector and expressed in *E. coli* BL21(DE3) Gold strain. However, the enzyme was produced in inclusion bodies and no enzymatic activity was detectable. Next, the α-GalNAc-ase gene was inserted into pPICZα in *Pichia pastoris* X33. Over 100 colonies were tested for enzymatic activity, but no clone with α-*N*-acetylgalactosaminidase activity was found.

Finally, *S. cerevisiae* was chosen as the host organism. The expression plasmid pYES-2CT/α-GalNAc-ase was constructed and used to transform the *S. cerevisiae* W303-1A strain. Positive transformants were grown on SC selective medium. The cultivation conditions for enzyme production were then optimized and scaled up. A series of various volumes of the SC media (50, 100, 200, 500 ml) in various cultivation flasks (100, 200, 500, 1000, 2000 ml), various concentrations of the galactose as an inducer (1-5%), cultivation temperatures (25, 28, 30 °C) and production times (12, 24, 36, 48, 72, 96, 120, 148 hours) were tested. Notably, the volume ratio of medium/flask was shown to be a critical parameter. Increasing the volume of the cultivation medium in the cultivation flask caused a dramatic decrease in enzyme production. The highest production of intracellular α-GalNAc-ase (approximately 0.4 U/mg) was achieved in a 2 l cultivation flask with 200 ml of the cultivation medium at 30 °C, with 2% galactose and 72 hours of cultivation (Fig. 1A,B). No extracellular α-GalNAc-ase activity was observed (Fig. 1C). Changes of other cultivation parameters (except volume) also led to a nonsignificant decrease in intracellular enzyme production. On the other hand the addition of detergent (1mM dodecyl maltoside) to the lysis buffer caused a twofold increase in the α-GalNAc-ase activity (Fig. 1C).

As shown in Table 1, the enzyme purification was achieved with good recovery. The enzyme purification was carried out by four-step chromatography. The majority of the contaminating proteins were removed by hydrophobic chromatography (Phenyl-Sepharose HR). The subsequent purification steps were used to remove the other proteins with similar biochemical properties. The enzyme was purified with a final yield of 12.1%. The specific activity of the recombinant enzyme against *o*NP-α-GalNAc was determined to be 42.3 U/mg

and SDS-PAGE of the purified α-GalNAc-ase revealed one single protein band with an estimated molecular mass of 76 kDa. This value is higher than the molecular mass calculated from the amino acids sequence (53961.2 Da) because of extensive protein glycosylation.

TABLE 1. Purification of the recombinant α-*N*-acetylgalactosaminidase from *Saccharomyces cerevisiae* W303

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

Purity is related to the starting material

Enzymatic properties of recombinant α-GalNAc-ase. The enzymatic properties of α-GalNAc-ase were assayed with *o*NP-α-GalNAc as the substrate. The recombinant α-GalNAc-ase exhibited a K_M value of 0.56 mmol/l and p*I* 4.4 in 50 mM citrate-phosphate buffer (pH 3.5) at 37 ˚C. The purified enzyme has a pH optimum at 2.0-2.4 (at 37˚C) and temperature optimum of 50-55 ˚C (at pH 3.0). The recombinant enzyme was stable in 50 mM citrate-phosphate buffer within the pH range 2 - 4 and at 4 ˚C for several weeks without any significant loss of activity. A loss of activity (35%) was observed after 2 days of incubation at 37 ˚C. Its enzymatic properties are consistent with the wild extracellular α-GalNAc-ase isolated from *A. niger* CCIM K2 (30).

Biochemical properties of the wild type and recombinant enzymes. The molecular weight of both the wild-type α-GalNAc-ase isolated from *A. niger* and the recombinant one was investigated using gel filtration. The gel filtration of the recombinant α-GalNAc-ase showed two active forms of α-GalNAc-ase with estimated molecular masses of approximately 70 kDa and 130 kDa (Fig. 2 A,B). The wild α-GalNAc-ase occurred only as 70 kDa monomers (Fig. 2 C,D). Moreover, the purified proteins exhibited dual enzyme activities; e.g. α-GalNAc-ase and α-Gal-ase, similarly to the native enzyme.

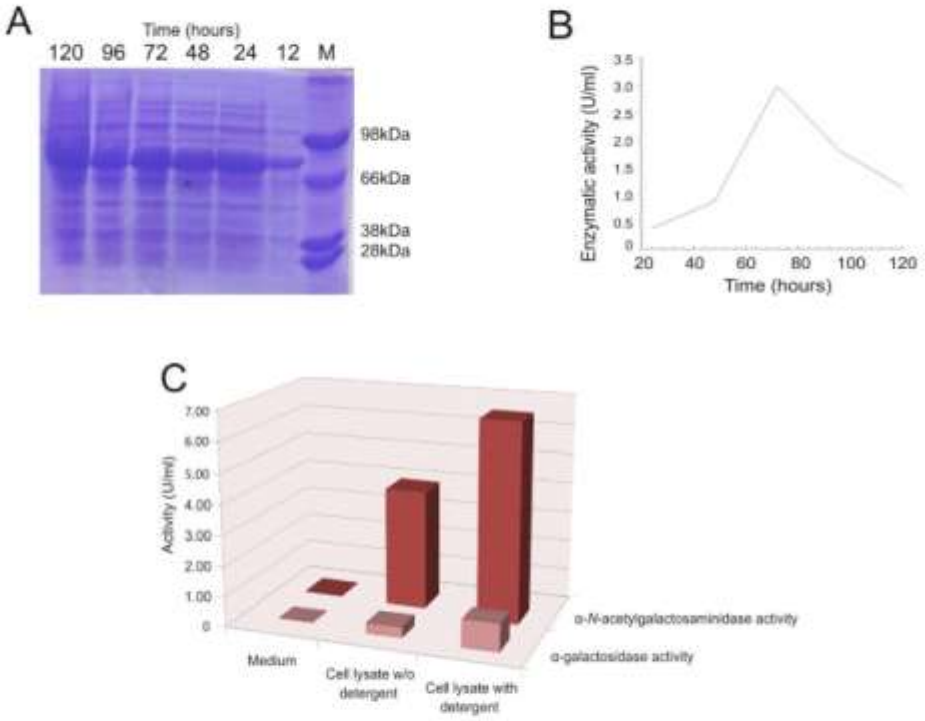


FIG. 1. Time profile of the recombinant intracellular α-GalNAc-ase production by *S.cerevisiae* W303: **A)** SDS-PAGE electrophoresis **B)** enzyme activity. Time optimization was carried out at 30 ˚C. Aliquots of the cell culture were harvested 12, 24, 48, 72, 96 and 120 hours after transfer to the SC medium with galactose as an inducer. The recombinant α-GalNAc-ase was identified as a band with an apparent molecular mass of approx. 76 kDa. **C)** α-*N*-acetylgalactosaminidase and α-galactosidase activity measured in cell lysates with and without dodecyl maltoside (detergent) in lysis buffer.

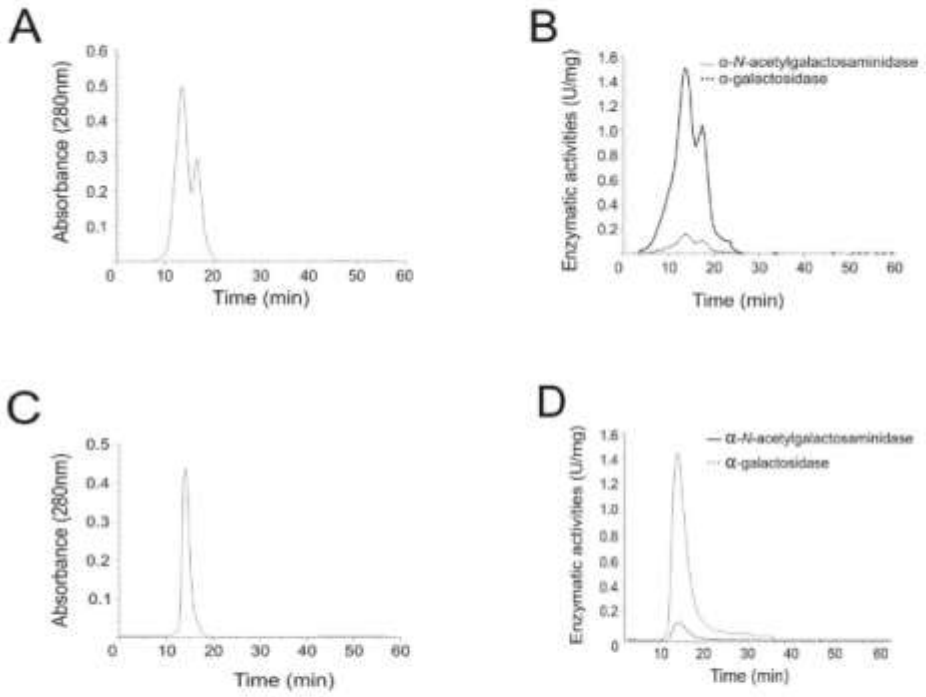


FIG 2. Elution profile of the wild and recombinant enzymes from gel filtration chromatography: **A:** protein profile of recombinant α-GalNAc-ase**, B:** enzyme activities of recombinant α-GalNAc-ase. **C:** protein profile of native α-GalNAc-ase**, D:** enzyme activities of native α-GalNAc-ase. The enzymes were injected into a Superdex 200 10/300 GL column (0.4 ml/min in 50 mM acetate buffer pH 3.6, 0.15 mM NaCl). Two peaks corresponded to the dimeric form (approximately 130 kDa) and the monomeric form (approximately 70 kDa), respectively. The purified proteins exhibited both α-GalNAc-ase and α-Gal-ase activity.

Analytical ultracentrifugation was emloyed to determine further structural parameters of both the native and recombinant α-GalNAc-ases. All experiments were performed at four various protein concentrations, at 4 °C, and the data were analyzed using the software SEDFIT. Native extracellular α-GalNAc-ase occurred only in a monomeric form compared to the recombinant α-GalNAc-ase, which existed as two distinct species with sedimentation coefficients of 7.65 S and 5.07 S (Fig. 3 A, B, D). These values corresponded well with the mass of the dimer and the monomer, respectively.

The sedimentation equilibrium experiment (Fig. 3 C), which enables direct calculation of molecular mass without the need for calibration or interaction with any matrix, yielded a molecular mass of 72 kDa and 127 kDa for the two ideal forms, correlating with the expected mass of recombinant α-GalNAc-ase monomer and dimer. This data corresponded well to the results obtained from the gel filtration (Fig. 2). Moreover, both the monomeric and the dimeric enzyme forms displayed both α-GalNAc-ase and α-Gal-ase activity in the same ratio.

Finally, the occurrence of the monomer and the dimer species and their ratio remained constant under various protein concentrations. Both sedimentation velocity and equilibrium experiments were performed with the protein at 2/3, 1/3 and 1/6 of the initial concentration and as shown in Fig. 3 C. Thus, there is no evidence for equilibrium between the two protein forms. Additionally, the size of the enzymatic complex was investigated by DLS analysis to be around 10 nm for the dimeric form.

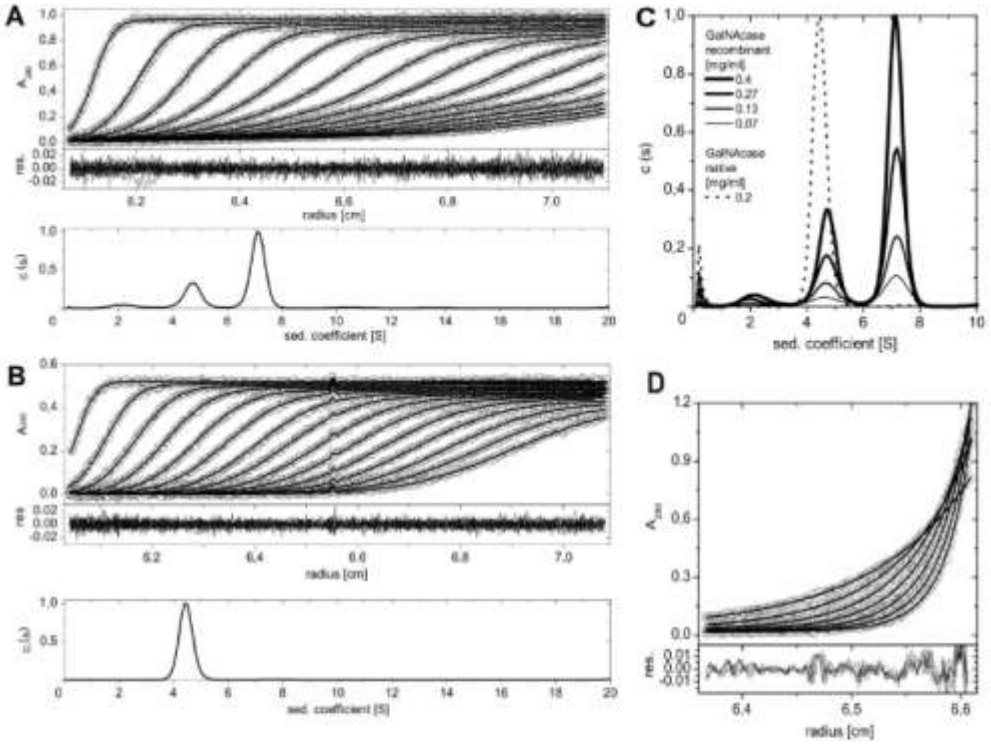


FIG. 3. Sedimentation velocity experiments. The recombinant α-GalNAc-ase **A**, and native α-GalNAc-ases **B**, were analyzed in an analytical ultracentrifuge using sedimentation velocity. Fitted data (upper panel) with residual plots (middle panel) showing the accuracy of the fit are shown together with the calculated continuous size distribution c(s) of the sedimenting species (lower panel) at various α-GalNAc-ase loading concentrations. **C:** The sedimentation velocity experiments showed that there is no apparent equilibrium between the

two α-GalNA-ase forms, since the decrease in both forms corresponds solely to the decrease in enzyme loading concentration. **D:** Equilibrium sedimentation distribution of recombinant α-GalNA-ase. Upper panel shows absorbance data with fitted curves (non-interacting discrete species model, lines), lower panel shows residuals derived from fitted data.

Electron microscopy. A negative staining technique was used for characterizing the wild-type and recombinant α-GalNAc-ases. Uranyl acetate or uranyl fomate was used for the staining. For image recording, a MegaViewII slow-scan camera was used at the primary microscope magnification of 64,000×. The pixel size at this magnification is approx. 1 nm, giving a Nyquist limit of around 2 nm. The wild-type α-GalNAc-ase sample showed a distribution of small particles with a mean diameter of 6 - 7 nm and shapes ranging from almost isomeric to elongated (Fig. 4 A, C). The sample of recombinant α-GalNAc-ase showed particles with a broader size distribution, with a mean diameter of 7 - 10 nm (Fig. 4 B, D). These data are in good agreement with the sedimentation velocity experiments (Fig. 3) and confirm that recombinant α-GalNAc-ase was present in the sample in two forms. We tried to find representative projections of the particles in the recombinant α-GalNAc-ase sample using EMAN 1.9, a suite of scientific image processing tools (19). Although our digital data sets were recorded with low resolution due to hardware limitations (approx. 2 nm) and the sizes of the wild and recombinant α-GalNAc-ases are small, we were able to resolve typical projections resembling the monomeric and dimeric form of recombinant α-GalNAc-ase (Fig 4. E).

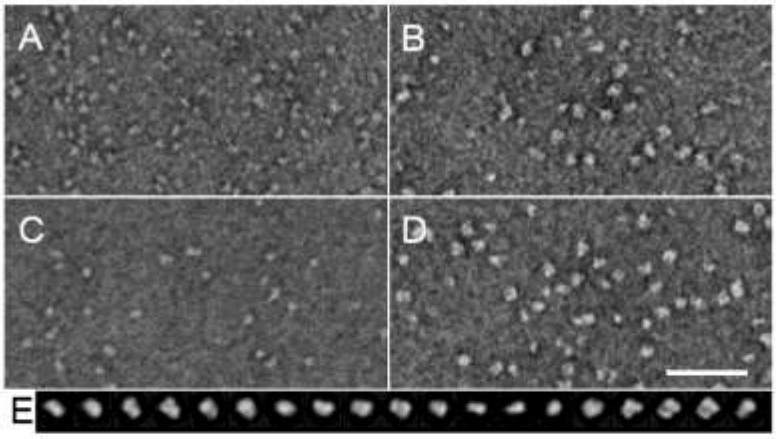


FIG 4. Electron microscopy: Negative staining of wild α-GalNAc-ase from *A. niger,* panels **A** and **C**; cloned enzyme from *S. cerevisiae,* panels **B** and **D**. The samples in panel **A** and **B** were negatively stained with uranyl acetate and the samples in panels **C** and **D** with

uranyl formate. The cloned enzyme appears mainly in dimers (B, D) in contrast to the wild-type (A, C), which is present as monomers. The magnification is the same for panels A through D and the scale bar represents 50 nm. In panel E representative class-averages of individual cloned enzyme particles from uranyl formate negatively stained sample created using EMAN software (19) are shown. It can be seen that the cloned enzyme sample contained predominately the dimeric form, but the monomeric form was also present.

***N*-Glycosylation sites.** *N*-Glycosidic linkage in proteins is mediated *via* the amide nitrogen of asparagine in the tripeptide consensus sequence Asn-Xxx-Ser/Thr, where Xxx is any amino acid except for proline. However, the presence of this sequence is not in itself sufficient to ensure glycosylation (21). There are eight such potential *N*-glycosylation sites in both the wild-type and recombinant α-GalNAc-ases, located at Asn 14, 52, 58, 88, 168, 320, 401 and 456.

α-GalNAc-ase was treated with Endo Hf and PNGase F in each individual digest and data were analyzed by mass spectrometry using the software mMass (37). 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-GalNAc-ase was treated with PNGase F (16, 21) 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 5b. Six asparagines of the eight potential *N*-glycosylation sites in the α-GalNAc-ases located at residues 14, 52, 58, 88, 320 and 456 were found to be glycosylated (Fig. 5 and 6).

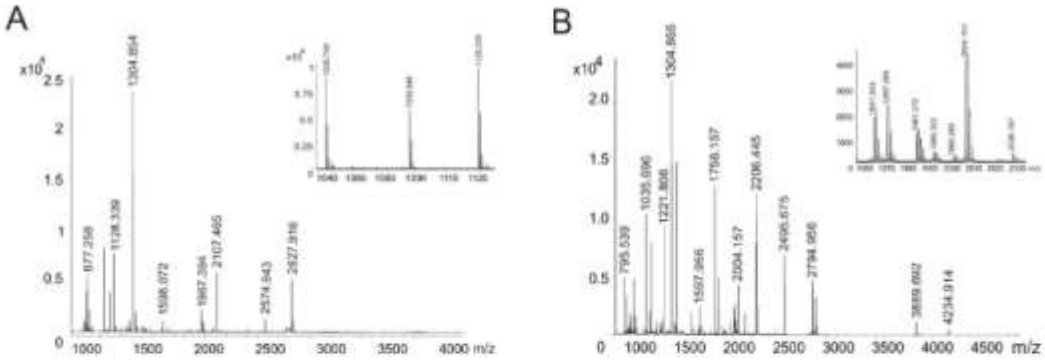


FIG 5. Determination of *N*-glycosylation sites. (**A**) α-GalNAc-ase was digested with Asp-N and treated with PNGase F, which cleaves off asparagine-linked carbohydrates and

converts the asparagine residue to aspartic acid, 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 1Da mass increase from the conversion of asparagines to aspartic acid by PNGase F. **(B)** α-GalNAc-ase was digested with Asp-N and deglycosylated with Endo Hf, which cleaves the bond between two GlcNAc units attached to asparagines (22). The final peptide has a mass increase of 203, 079Da. The peptide signal at *m/z* 2004.157 corresponded to peptide residue 310-325 (calculated *m/z* for $[M+H]^+$ 1801.076) with a 203.079Da mass increase for GlcNAc. The same methodology was used to identify the other *N*-glycosylation sites.

```
        10         20         30         40         50         60
MGFNNWARFM CDLNETLFTE TADAMAANGL RDAGYNRINL DDCWMAYQRS DNGSLRWNTT

        70         80         90        100        110        120
EFPHGLPWLA QYVKAKGFHF GIYEDSGNMT CGGYPGSYNH EEQDANTFAL WGIDYLKLDG

       130        140        150        160        170        180
CNVYATQGRT LEEEYKQRYG HWHQVLSKMQ HPLIFSESAP AYFAGTDNNT DWYTVMNWVP

       190        200        210        220        230        240
IYGELARHST DILVYSGAGS AWDSIMNNYN YNTLLARYQR PGYFNDPDFL IPDHPGLTAD

       250        260        270        280        290        300
EKRSHFALWA SFSAPLIISA YIPALSKDEI AFLTNEALIA VNQDPLAQQA TFASRDNTLD

       310        320        330        340        350        360
ILTRNLANGD RLXTVLNKGN TTVTRDIPVQ WLGLTETDCT YTAEDLWDGK TQKISDHIKI

       370        380        390        400        410        420
ELASHATAVF RLGLPQGCSS VVPTGLVFNT ASGNCLTAAS NSSVAFQSCN GETSQIWQVT

       430        440        450        460        470        480
LSGVIRPVSQ TTQCLAADGN SVKLQACDST DSDGQNWTYA VTGNLKNAKT DGCLTEGSVQ

MKSCLYE
```

FIG 6. α-GalNAc-ases have eight *N*-glycosylation sites. Six from eight glycosylation sites were confirmed (bold and underlined). Glycosylation sites without glycosylation are showed in italics and underlined.

DISCUSSION

The *A. niger* α-GalNAc-ase was successfully cloned, sequenced and expressed in the yeast *S. cerevisae*. This is the first report of α-GalNAc-ase gene from *Aspergillus*. The α-GalNAc-ase gene contains an open reading frame, which encodes a protein of 487 amino acid residues and displays 64 % identity with the only fungal α-GalNAc-ase described so far (2). Both *Aspergillus* α-GalNAc-ase and *Acremonium* α-GalNAc-ase belong to glycoside

hydrolase family 27 and the phylogenetic analysis revealed that they did not belong to the further α-GalNAc-ases described so far (all from vertebrates) but to another cluster, including yeast α-Gal-ases. The relationship between α-GalNAc-ases and α-Gal-ases was described previously (2, 16).

A recombinant and wild α-GalNAc-ase were purified, characterized and compared. Although vector containing α-signal sequence was used for the expression only intracellular α-GalNAc-ase activity was detected.

The recombinant α-GalNAc-ase retained its dual enzyme activity - this enzyme was able to hydrolyze both *o*NP-α-GalNAc and *p*NP-α-Gal as its substrate. Moreover, the pH profile of the recombinant α-GalNAc-ase was changed (Fig. 7), the recombinant enzyme was more active in neutral pH compare to wild one. This fact predertermines this enzyme for the red blood cell group A erytrocyte transformation in to the group of H (0), beeing a universal donor (18).

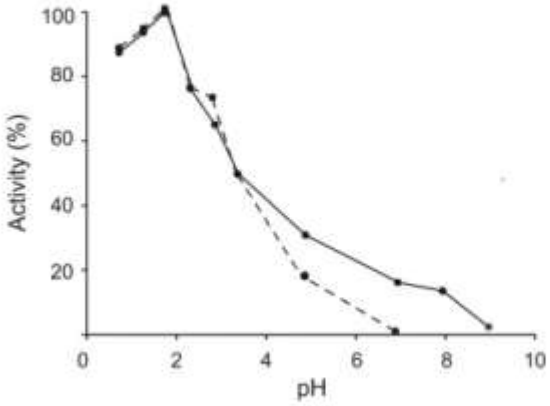


FIG 7. Effect of pH on the activity of wild α-GalNAc-ase (- -•- -) and recombinant α-GalNAc-ase (—•—).

Despite the fact that the recombinant and the wild enzyme have the same AA sequence and the same number and localisation of *N*-glycosylation sites, five different independent methods verified their different quaternary structure. Sedimentation velocity and sedimentation equilibrium measurements, dynamic light scattering, gel filtration and finally electron microscopy confirmed that native extracellular α-GalNAc-ase occurred purely in a monomeric form compared to recombinant α-GalNAc-ase, which existed as two distinct species – monomeric and dimeric forms simultaneously. Although the dimeric form was predominant, both forms were fully functional and separately showed the same enzymatic and biochemical properties (substrate specificity, pH optimum, as well as temperature optimum

and stability). A possible explanation for their parallel presence can consist in different distribution and processing of the α-GalNAc-ase in various cell compartments. α-GalNAc-ase can occur in various membrane organelles such as ER, Golgi aparatus and lysosome. After the saturation of the organelle transport systems (protein translocons) by an excess of the recombinant enzyme, the enzyme could be released into the cytoplasm and the microenviroment conditions (pH, oxidation/reduction potential), which could cause different folding pathways leading to different 3D structures.

In conclusion, we prepared stable, active recombinant α-GalNAc-ase 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.

ACKNOWLEDGMENTS

Financial support was provided by Grant Agency of Charles University, grant no. 19309/2009, by Ministry of Education of Czech Republic (MSM21620808 and 1M0505), by Czech Science Foundation (P207/10/0321, 305/09/H008, 303/09/0477), and by the research concept of the Institute of Microbiology AV0Z50200510.

REFERENCES

1. **Abelson, N. J., and I. M. Simon.** 2004. Guide to yeast genetic and molecular and cell biology. Methods in Enzymology. Part A, B. **194.**

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

3. **Bause, E., and G. Legler.** 1981. Structural requirements of *N*-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. **195**: 639-644.

4. **Benada, O., and V. Pokorny.** 1990. Modification of the polaron sputter-coater unit for glos discharge activation of carbon support films. J. Electron Microsc. Tech. **16**: 235-239.

5. **Calcutt, M., M. J. Hsieh, L. F. Chapman, and D. Smith.** 2002. Identification, molecular cloning and expression of an α-*N*-acetylgalactosaminidase gene from *Clostridium perfringens*. FEMS Microbiol. Lett. **214**: 77-80.

6. **Cheng, Q., H. Li, K. Merdek,and J. T. Park**. 2000. Molecular characterization of the β-*N*-acetylglucosaminidase of *Escherichia coli* and its role in cell wall recycling. J Bacteriol.**182**:4836–4840.

7. **Clark, N. E., and S. C. German**. 2009. The 1.9 Å structure of human α-*N*-acetylgalactosaminidase: the molecular basis of Schindler and Kanzaki diseases. J. Mol. Biol. **393**: 435-447.

8. **Dean, K. J., S. J. Sung, and C.C. Sweeley**. 1977. The identification of α-galactosidase B from human liver as an α-*N*-acetylgalactosaminidase. Biochem. Biophys. Res. Commun. **77**: 1411–1417.

9. **Desnick, R. J., and A. M Wang**. 1995. Schindler disease: α-*N*-acetylgalactosaminidase deficiency. In Scriver, C.R., Beaudet, A.L., Sly, W.S, and Valle, D. The Metabolic and Molecular Bases of Inherited Disease, **7th ed. McGraw-Hill, New York**, pp. 2509–2527.

10. **Ettrich, R., et al.** 2007. Structure of dimeric *N*-glycosylated form of fungal beta-*N*-acetylhexosaminidase revealed by computer modeling, vibrational spectroscopy, and biochemical studies, BMC Struct. Biol. **7**, 32.

11. **Fujimoto, Z., et al.** 2002. Crystallization and preliminary X-ray crystallographic studies of rice α-galactosidase. Acta Crystallogr. D Biol. Crystallogr. **D58**: 1374-1375.

12. **Hata, J., et al.** 1992. Purification and characterization of *N*-acetyl-α-D-galactosaminidase from *Gallus domestici*. Biochem. Int. **28**, 77-86

13. **Herrmann T., et al.** 1998. Molecular cloning, structural organization, sequence, chromosomal assignment and expression of the mouse α-*N*-acetylgalactosaminidase gene. Gene. **211**: 205–214.

14. **Krogh, T. N., et al.** 1997 Glycosylation analysis and protein structure determination of murine fetal antigen 1 (mFA1)--the circulating gene product of the delta-like protein (dlk), preadipocyte factor 1 (Pref-1) and stromal-cell-derived protein 1 (SCP-1) cDNAs. Eur. J. Biochem. **244**: 334-342.

15. **Krokhin O., et al.** 2004. Site-specific N-glycosylation analysis: matrix-assisted laser desoption/ionization quadrupole-quadrupole time-of-flight tandem mass spectral signatures for recognition and identification of glycopeptides. Rapid Commun. Mass Sp. **18**: 2020-2030.

16. **Kulik, N., et al.** 2010. The α-galactosidase type A gene *aglA* from *Aspergillus niger* encodes a fully functional a-*N*-acetylgalactosaminidase. Glycobiology. **20**: 1410-1419.

17. **Levy, G. N., and D Aminoff.** 1980. Purification and properties of α-*N*-acetylgalactosaminidase from *Clostridium perfringens*. J. Biol. Chem. **255**: 11373-11342.

18. **Liu, P. Q., et al.** 2007. Bacterial glycosidases for the production of universal red blood cells. Nat. Biotechnol. **25**: 454-464 .

19. **Ludtke, S. J., P. R. Baldwin, and W. Chiu.** 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol. **128:** 82-97.

20. **Patterson, S. D., and V. Katta.** 1994. Prompt fragmentation of disulfide-linked peptides during matrix-assisted laser desorption ionization mass spectrometry. Anal. Chem. **66**: 3727-3732.

21. **Ried, B.F., K. Tatsuta, and J. Thiem.** 2001. Glycoscience **I, II, III**. Springer-Heildelberg.

22. **Sartor, M.**, Dynamic light scatternig. 2002. University of California, San Diego

23. **Shunguang, M., K. E. Hill, R. F. Burk, and R. M. Caprioli**. 2003. Mass spectrometry identification of *N*- and *O*-glycosylation sites of full-length rat selenoprotein P and determination of selenide-sulfide and disulfide linkages in the shortes isoforms. Biochemistry **42**: 9703-9711.

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

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

26. **Tsuji S., et al.** 1989. Molecular cloning of a full-length cDNA for human α-*N*-acetylgalactosaminidase (α-galactosidase B). Biochem. Biophys. Res. Commun. **163**: 1498-1504.

27. **Wang, A. M., D. F. Bishop, and R. J. Desnick.** 1990. Lysosomal α-*N*-acetylgalactosaminidase deficiency, the enzymatic defect in angiokeratoma corporis diffusum with glycopeptiduria. J. Biol. Chem. **265**: 21859–21866.

28. **Wang, A. M., Y. A. Ioannou, K. M. Zeidner, and R. J. Desnick.** 1998. Murine α-*N*-acetylgalactosaminidase: isolation and expression of a full-length cDNA and genomic organization: further evidence of an α-galactosidase gene family. Mol. Genet. Metab. **65**: 165–173.

29. **Wang, A. M., D. F. Bishop, and R. J. Desnick.** 1990. Human α-*N*-acetylgalactosaminidase: molecular cloning, nucleotide sequence and expression of a full-length cDNA. J. Biol. Chem. **265**: 21859–21866.

30. **Weignerová, L., T. Filipi, D. Manglová, and V. Křen.** 2008. Induction, purification and characterization of α-*N*-acetylgalactosaminidase from *Aspergillus niger*. Appl. Microbiol. Biotechnol. **79**:769–774.

31. **Weignerová, L., P. Simerská, and V. Křen.** 2009. α-Galactosidases and their applications in biotransformations. Biocat. Biotrans. **27**: 79-89.

32. **Weignerová, L., et al.** 2010. Condensation reactions catalyzed by α-*N*-acetylgalactosaminidase from *Aspergillus niger* yielding α-*N*-acetylgalactosaminides. Biocat. Biotrans. **28**: 150-155.

33. **Weignerová, L., et al.**. 2003. Fungal β-*N*-acetylhexosaminidases with high β-*N*-acetylgalactosaminidase activity and their use for synthesis of β-GalNAc-containing oligosaccharides. Carbohydr Res. **338**:1003–1008.

34. **Yu, C. Y., et al.** 2008. Human RBCs blood group conversion from A to O using a novel α-*N*-acetylgalactosaminidase of high specific activity. Chinese Sci. Bull. **53**: 2008-2016.

35. **Zhu A., and J. Goldstein.** 1993. Cloning and characterization of cDNA encoding chicken liver α–*N*-acetylgalactosaminidase. Gene **137**: 309-314.

36. **Zhu, A., Z. K. Wang, and R. Beavis.** 1998. Structural studies of α-*N*-acetylgalactosaminidase: effect of glycosylation on the level of expression, secretion efficiency and enzyme activity. Arch. Biochem. Biophys. **352**: 1–8.

37. www.mmass.org

Appendix 3

Mrázek H., Weignerová L., Bezouška K.

Aktivní forma α-*N*-acetylgalaktosaminidázy z vláknité houby *Aspergillus niger* a její rekombinantní exprese.

Czech Patent Application.

P-0065-CZ

Aktivní forma α-*N*-acetylgalaktosaminidázy z vláknité houby *Aspergillus niger* a její rekombinantní exprese

Oblast techniky

Vynález se týká α-*N*-acetylgalaktosaminidázy neboli α-*N*-acetylgalaktosaminid-*N*-acetylgalaktosaminohydrolázy, enzymu s α-galaktosidázovou aktivitou, a přípravy aktivní formy tohoto enzymu v kvasinkovém expresním systému *Saccharomyces cerevisiae*

Dosavadní stav techniky

α-*N*-acetylgalaktosaminidáza (α-GalNAasa; EC 3.2.1.49) neboli α-*N*-acetylgalaktosaminid-*N*-acetylgalaktosaminohydroláza je enzym selektivně štěpící terminálně vázaný GalNAc, který je připojen O-α-D-glykosidovou vazbou na aminokyseliny Ser nebo Thr, či na oligosacharidový řetězec. Enzymová hydrolýza glykosidické vazby probíhá retenujícím způsobem, kdy je konfigurace anomerie vazby substrátu a produktu nezměněna. Uhlík C1 α-*N*-acetylgalaktosaminidu podílející se na glykosidické vazbě, podléhá dvěma následným nukleofilním atakům, při kterých se mění anomerie vazby. Nejprve atom kyslíku Asp140 atakuje elektrofilní uhlík C1 sacharidového skeletu, což vede k tvorbě kovalentního komplexu enzym-substrát. Následně je kovalentní komplex atakován deprotonovanou molekulou vody, což způsobí regeneraci enzymu a uvolnění produktu.

Podle enzymové nomenklatury IUB-MB (International Union of Biochemistry and Molecular Biology) je α-*N*-acetylgalaktosaminidáza řazena mezi hydrolázy-glykosidázy hydrolyzující *O*- a *S*-glykosidické vazby. Na základě příbuznosti genů a isoenzymů byla sestavena fylogenetická mapa tohoto enzymu. Tato mapa ukázala, že evoluční počátek genu pro α-*N*-acetylgalaktosaminidázu obratlovců a hub je odlišný. α-*N*-acetyl-galaktosaminidáza se pravděpodobně vyvinula z α-galaktosidázy u hub, které osídlily pevné substráty, jako adaptace na využití sloučenin obsahující terminálně vázaný *N*-acetylgalaktosamin.

α-*N*-acetylgalaktosaminidáza se vyskytuje ve více formách navzájem se lišících svou katalytickou aktivitou. Dimerní a tetramerní forma je enzymově aktivní na rozdíl od monomerní, která nevykazuje enzymovou aktivitu.

Lidská α-GalNAc-áza hraje důležitou roli při metabolismu glykokonjugátů. Nedostatek tohoto enzymu může způsobit Schindlerovu nebo Kanzakiho nemoc, která může vyústit až v nevratné neurodegenerativni postižení.

Bylo prokázáno, že α-GalNAc-áza purifikovaná z kuřecích jater (*J. Hata, et al, Purification and characterization of N-acetyl-α-D-galactosaminidase from Gallus domestici, Biochem. Int. 28 (1992) 77-86*), grampozitivní tyčinkovité anaerobní sporulující bakterie *Clostridium perfringens* (*G.N. Levy, and D Aminoff, Purification and properties of α-N-acetylgalactosaminidase from Clostridium perfringens, J. Biol. Chem. 255 (1980) 11373-11342*) a dalších bakteriálních kmenů (P.*Q. Liuet et al, Bacterial glycosidases for the production of universal red blood cells, Nat. Biotechnol. 25 (2007) 454-464*), je schopna účinně štěpit terminálně α-vázaný GalNAc z epitopů erytrocytů krevní skupiny A. Výsledkem tohoto štěpení jsou epitopické struktury erytrocytů krevní skupiny H(0) univerzální dárce. (*C.Y. Yu et al, Human RBCs blood group conversion from A to O using a novel α-N-acetylgalactosaminidase of high specific activity, Chinese Sci. Bull. 53 (2008) 2008-2016*). Nevýhodu všech dosud známých α-*N*-acetylgalaktosaminidáz, ať izolovaných či rekombinantně připravených (viz např. *H. Ashida, H. Tamaki, T. Fujimoto, K. Yamamoto, and H. Kumagai, Molecular cloning of cDNA encoding α-N-acetylgalactosaminidase from Acremonium sp. and its expression in yeast, Arch. Biochem. Biophys. 384 (2000) 305-310*)) je jejich nulová enzymatická aktivita v oblasti neutrálního pH. Z těchto důvodů nebylo možné provést experiment přeměny erytrocytů přímo na živých izolovaných krevních buňkách, které v oblasti nízkého pH podléhají erytrolýze. Všechny výše zmíněné experimenty se sacharidovými strukturami krevních skupin byly prováděny v nízkém pH výhradně se samotnými epitopy, které byly připraveny synteticky.

V Laboratoři architektury proteinů, MBÚ AV ČR, v.v.i a v Laboratoři biotransformací, MBÚ AV ČR, v.v.i bylo provedeno několik experimentů, které vedly k základním biochemickým poznatkům o α-*N*-acetylgalaktosaminidázách izolovaných z přírodních zdrojů a byl identifikován nejlepší producent tohoto enzymu - vláknitá houba *Aspergillus niger* CCIM K2. Extracelulární enzym z tohoto zdroje byl purifikován a plně charakterizován (*L. Weignerová, T. Filipi, D. Manglová, and V. Křen, Induction, purification and characterization of α-N-acetylgalactosaminidase from Aspergillus niger, Appl. Microbiol. Biotechnol. 79 (2008) 769–774.*).

α-GalNAc-áza je schopná také katalyzovat reverzní hydrolytické reakce, výsledkem těchto reakcí může být α-D-Gal*p*NAc-(1→6)-D-Gal*p*NAc, α-D-Gal*p*NAc-(1→3)-D-Gal*p*NAc, α-D-Gal*p*NAc-(1→6)-D-Gal*f*NAc a také Tn antigen (Gal*p*NAc-α-*O*-Ser/Thr) a jeho *N*-(*tert*-butoxykarbonyl) deriváty (*L. Weignerová, P. Simerská, and V. Křen, α-Galactosidases and their applications in biotransformations, Biocat. Biotrans. 27 (2009) 79-89*), (*L. Weignerova et al, Condensation reactions catalyzed by α-N-acetylgalactosaminidase from Aspergillus niger yielding α-N-acetylgalactosaminides. Biocat. Biotrans. 28 (2010) 150-155*).

Měření enzymové aktivity a molekulární modelování ukázalo duální enzymatickou aktivitu specifickou jen pro α-GalNAc-ázu izolovanou z *Aspergilllus niger*. Tento enzym je schopný hydrolyzovat jak *o*-nitrofenyl-2-acetamido-2-deoxy-α-D-galaktopyranosid (α-*N*-acetylgalaktosaminidázová aktivita), tak *p*-nitrofenyl-α-D-galaktopyranosid (α-galaktosidázová aktivita), přičemž α-galaktosidázová aktivita je ve srovnání s α-*N*-acetyl-galaktosaminidázovou přibližně 10 % (*N. Kulik et al, The α-galactosidase type A gene aglA from Aspergillus niger encodes a fully functional a-N-acetylgalactosaminidase, Glycobiology. 20 (2010) 1410-1419*).

Do současné doby byla v literatuře popsána izolace, purifikace a základní charakterizace několika α-*N*-acetylgalaktosaminidáz z přírodních zdrojů a pouze jedna rekombinantní příprava tohoto enzymu (viz výše). Avšak žádný z těchto enzymů nevykazoval α-*N*-acetylgalaktosaminidázovou aktivitu v neutrálním pH. Rekombinantně připravená α-*N*-acetylgalaktosaminidáza podle vynálezu překvapivě projevuje významnou α-*N*-acetylgalaktosaminidázovou aktivitu dokonce i při pH 8 (v 50mM HEPES pufru, pH=8,0). Tato aktivita je přibližně až 20% vzhledem k maximu, kterého je dosaženo přibližně při pH 2. Aktivita enzymu při pH v rozmezí 6 až 8 je dostatečná pro praktické využití enzymu.

Podstata vynálezu

První předmět vynálezu se týká rekombinantní aktivní formy α-*N*-acetylgalaktosaminidázy (α-GalNAc-asa; EC 3.2.1.49) neboli α-*N*-acetylgalaktosaminid-*N*-acetylgalaktosamino-hydrolázy z vláknité houby *Aspergillus niger*, jejíž aminokyselinová sekvence je následující (SEKV. ID. Č. 1):

```
MGFNNWARFMCDLNETLFTETADAMAANGLRDAGYNRINLDDCWMAYQRSDNGSLRWNTT  60
EFPHGLPWLAQYVKAKGFHFGIYEDSGNMTCGGYPGSYNHEEQDANTFALWGIDYLKLDG 120
CNVYATQGRTLEEEYKQRYGHWHQVLSKMQHPLIFSESAPAYFAGTDNNTDWYTVMNWVP 180
IYGELARHSTDILVYSGAGSAWDSIMNNYNYNTLLARYQRPGYFNDPDFLIPDHPGLTAD 240
EKRSHFALWASFSAPLIISAYIPALSKDEIAFLTNEALIAVNQDPLAQQATFASRDNTLD 300
ILTRNLANGDRLLTVLNKGNTTVTRDIPVQWLGLTETDCTYTAEDLWDGKTQKISDHIKI 360
ELASHATAVFRLGLPQGCSSVVPTGLVFNTASGNCLTAASNSSVAFQSCNGETSQIWQVT 420
LSGVIRPVSQTTQCLAADGNSVKLQACDSTDSDGQNWTYAVTGNLKNAKTDGCLTEGSVQ 480
MKSCLYE
```

Rekombinantně připravená forma tohoto enzymu překvapivě vykazuje v rozmezí pH 6 až 8 α-*N*-acetylgalaktosaminidázovou aktivitu na úrovni 15 až 30 % (při pH 7 přibližně 20 %) vzhledem k maximu, kterého je dosaženo přibližně při pH 2. Díky této významné aktivitě při neutrálním pH představuje enzym podle vynálezu výhodný nástroj pro použití při přeměně krevních buněk typu A na krevní buňky typu H(0) univerzální dárce. Rekombinantně připravený enzym vykazuje také α-galaktosidázovou aktivitu. Tento enzym může být využit při syntetických reakcí oligo/polysacharidových struktur.

Další předmět vynálezu se týká izolované molekuly nukleové kyseliny, která kóduje aktivní α-*N*-acetylgalaktosaminidázu s α-galaktosidázovou aktivitou. Tato izolovaná molekula může být DNA nebo RNA (mRNA), výhodně je to cDNA. Také nukleové kyseliny se sekvencí komplementární k výše uvedeným sekvencím spadají do rozsahu vynálezu. Nejvýhodněji se jedná o cDNA s následující sekvencí (SEKV. ID. Č. 2):

```
atgggtttcaacaattgggcccgcttcatgtgcgacctcaacgagaccctgtttaccgag       60
actgccgatgcgatggctgctaacggtctgcgggacgcaggctacaatcgcatcaatctg      120
gatgattgctggatggcttatcagcgatccgacaatggatccctacggtggaacacgact      180
gagttcccacacggcctgccttggctagctcaatatgtcaaagccaaagggtttcatttt      240
ggaatctatgaagattctggcaacatgacttgtggcggataccccggatcctacaaccac      300
gaggagcaggacgccaacacctttgctttatgggggattgactatctcaagctcgacggt      360
tgcaacgtctacgcaacacaaggtaggacactcgaggaggaatacaagcaacgctacgga      420
cattggcaccaagtcctcagcaagatgcagcacccactgatcttctccgagtcagccccg      480
gcatacttcgcgggcacagacaacaacacagactggtacaccgtgatgaactgggtcccg      540
atctacggggagctggcccgccattctaccgatatcctggtgtacagtggagcaggtagc      600
gcatgggacagcattatgaataactacaactacaacactcttcttgcgcgctaccagcga      660
ccggggtatttcaatgatcctgattttctgatcccggatcatcctggcctgacggcggat      720
gaaaagcgatcgcattttgcactgtgggcttctttctcggctccacttattatcagtgct      780
tatatacctgcactttcgaaggatgagattgccttcttgacgaacgaggcattgattgcg      840
gtgaatcaggatcccctggcccagcaggccacgtttgcgagccgcgataacacactggat      900
atattgacgcgtaatctggcaaacggcgacaggctgctgacggtgcttaataagggaaac      960
acaactgtaacgagggacattcccgtacaatggttgggtcttacagagactgactgtaca     1020
tacacggccgaggatctctgggatggcaagacccagaagatcagcgaccatataaagatt     1080
gaactagccagccatgcgacagcagtcttccggctcggtcttccgcagggttgttcctcg     1140
gtagtgccaacgggacttgtcttcaacacagcatcgggcaattgtctgaccgctgcctca     1200
aattcttcagtcgcatttcagtcctgcaatggagagacctctcagatctggcaggtgaca     1260
ctgtcaggagtcattcgtccagtatcgcagaccacacaatgcttggctgctgatggaaac     1320
tcagttaagctgcaagcatgtgacagcacggatagcgacggccagaactggacgtatgca     1380
gtcacgggaaatttaaagaatgcgaagacagatggttgcctgactgagggatcagtgcag     1440
atgaagtcgtgtctttatgaga
```

Ještě další předmět vynálezu se týká způsobu přípravy aktivní formy α-*N*-acetylgalaktosaminidázy, jak byla definována výše. Způsob přípravy aktivní formy α-*N*-acetylgalaktosaminidázy zahrnuje čtyři stupně: izolaci celkové RNA z vláknité houby *Aspergillus niger*, příprava fragmentů specifické DNA, klonování této DNA do vhodného klonovacího a expresního vektoru, transformaci plazmidové DNA do vhodných hostitelských buněk, exprese aktivního enzymu, izolace a purifikace tohoto enzymu.

1.stupeň: Izolace RNA a příprava DNA fragmentů kódujících aktivní formu enzymu α-*N*-acetylgalaktosaminidázy.

Celková mRNA se izoluje z vláknité houby *Aspergillus niger* a následně se užije jako templát pro syntézu cDNA pomocí reverzní transkriptázy. cDNA se poté použije jako templát pro PCR amplifikaci DNA kódující aktivní formu alfa-*N*-acetylgalaktosaminidázy. Pro amplifikaci se užijí specifické oligonukleotidové primery následujících sekvencí:

5´ – ttat tct aga atg ggt ttc aac aat tgg gcc cgc – 3´ (SEKV. ID. Č. 3)

5´ – att gaa ttc tta gcc atc cct ctc ata aag aca cga ctt – 3´ (SEKV. ID. Č. 4)

2.stupeň: Konstrukce expresního plasmidu a jeho transformace do kvasinkového expresního systému *Saccharomyces cerevisiae*.

Expresní plasmid je konstruován vložením DNA fragmentu vzniklého po PCR amplifikaci v předchozím stupni do vhodného expresního vektoru. Tento vektor se pak metodou transfekce (výhodně elektroporace, nebo jinou metodou odborníkům známou) vnese do kompetentních buněk expresního systému.

3.stupeň: Exprese aktivní formy α-*N*-acetylgalaktosaminidázy.

Z transformantů vzniklých ve stupni 2 se vybere nejlepší produkční klon, který se použije pro inokulum, které se následně použije pro velkoobjemovou produkci proteinu ve vhodném produkčním médiu za vhodných kultivačních podmínek. Exprese proteinu je indukována vhodným induktorem (v závislosti na typu promotoru v expresním vektoru a mechanismu regulace proteinové exprese).

4.stupeň: Izolace aktivní α-*N*-acetylgalaktosaminidázy z hostitelského expresního systému a purifikace tohoto aktivního enzymu.

Po velkoobjemové produkci je buněčná kultura odstředěna centrifugací a aktivní enzym je izolován pomocí lyzačního pufru s přídavkem detergentu. α-*N*-acetylgalaktosaminidáza se pak purifikuje, např. pomocí tří po sobě jdoucích chromatografií (Phenyl-Sepharosa, S-sepharosa a Superdex 200), nebo jiným vhodným způsobem odborníkům známým.

Další předmět vynálezu se týká rekombinantní α-*N*-acetylgalaktosaminidázy s duální aktivitou vykazující významnou α-*N*-acetylgalaktosaminidázovou aktivitu při neutrálním pH (resp. v rozmezí pH 6-8), která je připravitelná výše popsaným způsobem.

Dále se vynález týká použití dvojice oligonukleotidů sestávající ze sekvencí SEKV. ID. Č. 3 a SEKV. ID. Č. 4 pro přípravu nukleové kyseliny, výhodně cDNA, kódující rekombinantní α-*N*-acetylgalaktosaminidázu s duální aktivitou vykazující významnou α-*N*-acetylgalaktosaminidázovou aktivitu při neutrálním pH.

A nakonec také použití dvojice oligonukleotidů sestávající ze sekvencí SEKV. ID. Č. 3 a SEKV. ID. Č. 4 pro přípravu rekombinantní α-*N*-acetylgalaktosaminiázy s duální aktivitou vykazující významnou α-*N*-acetylgalaktosaminidázovou aktivitu při neutrálním pH.

Stručný popis obrázků

Obrázek 1: **A**. Elektroforetická analýza produkce α-*N*-acetylgalaktosaminidázy po jednotlivých časových úsecích. **B**. α-*N*-acetylgalaktosaminidázová aktivita měřena v produkčním médiu po jednotlivých časových úsecích

Obrázek 2: **A**. Eluční profil rekombinantní α-*N*-acetylgalaktosaminidázy z gelové filtrace

B. α-*N*-acetylgalaktosaminidázová a α-galaktosidázová aktivita měřena v jednotlivých eluovaných frakcích

Obrázek 3: Stanovení základních enzymatických vlastností rekombinantní α-*N*-acetylgalaktosaminidázy jako jsou hodnoty K_m a V_{max}

Obrázek 4: Stanovení pH optima a oblasti pH, kde rekombinantní a přirozeně izolované α-*N*-acetylgalaktosaminidáza vykazují enzymatickou aktivitu. α-*N*-acetylgalaktosaminidáza izolovaná z vláknité houby *Aspergillus niger* (--●--), rekombinantně připravená alfa-*N*-acetylgalaktosaminidáza podle vynálezu (—●—).

Příklady provedení vynálezu

Příklad 1

Produkce rekombinantní aktivní α-N-acetylgalaktosaminidázy v kvasinkovém expresním systému Saccharomyces cerevisiae

1.1 Příprava cDNA a expresního vektoru

Celková RNA byla izolována z vláknité houby *Aspergillus niger* pomocí RNeasy Plant Mini Kit (QIAGEN). Tato celková RNA byla použita jako templát pro syntézu cDNA s využitím reverzní transkritázy SuperScript III (Invitrogen). DNA fragmenty kódující aktivní formu alfa-*N*-acetylgalaktosaminidázy byly amplifikovány pomocí PCR amplifikace s využitím páru specifických oligonukleotidových primerů (SEKV. ID. Č. 3 a SEKV. ID. Č. 3) následujících sekvencí:

5´ – ttat tct aga atg ggt ttc aac aat tgg gcc cgc – 3´

5´ – att gaa ttc tta gcc atc cct ctc ata aag aca cga ctt – 3´

PCR amplifikace probíhala v reakční směsi o objemu 50µl o následujícím složení: 5µl 10x koncentrovaného ThermoPol reaction buffer (New England Biolabs); 1µl 10mM dNTPs; 2µl každého primeru; 1,5µl síranu hořečnatého; 1µl templátové cDNA připravené výše zmíněným postupem; 37,5µl sterilní deionizované vody; 0,5µl Deep Vent Polymerasy (2U/µl). S takto připravenou směsí bylo provedeno 30 cyklů PCR amplifikace, dle standardního protokolu s 30 sekundovým nasedáním (annealing) primerů na templátovou DNA při teplotě 50˚C a dvou minutovou polymerací při teplotě 72˚C.

Vzniklý PCR produkt byl separován pomocí agarosové elektroforézy, po které následovala purifikace amplikonu přímo z gelu pomocí komerčně dostupných souprav (Genomed). Po izolaci byl amplikon štěpen restrikčními endonukleázami XbaI a EcoRI (New England Biolabs) a takto připravený DNA fragment byl opět podroben separaci na agarosové elektroforéze, izolaci a purifikaci z agarozového gelu. Stejným způsobem byl připraven expresní vektor pYES-2CT.

Základní postupy molekulární biologie, které byly použity, jsou odborníkům známy a jsou popsány například v publikaci: *Sambrook J, Russel DW. Molecular cloning: A laboratory Manual. Cold Spring Harbor, Laboratory Press, New York, 2001, 3rd Edition.*

Následovala ligační reakce, která probíhala 16 hodin při laboratorní teplotě. Objem ligační směsi byl 20µl, reakční směs obsahovala cca 50-100ng linearizovaného expresního vektoru a stejné množství DNA fragmentu; 2µl 10x koncentrovaného ligačního pufru pro T4 DNA ligázu (Fermentas); 1µl T4 DNA ligázy (5WeissU/µl) (Fermentas). Celý objem ligační směsi byl následně použit pro transformaci kompetentních buněk *E. coli* BL21-DE3 (Stratagene) a byla provedena velkokapacitní izolace plasmidové DNA pomocí komerčně dostupných souprav (Genomed). Pro ověření kvality plasmidové DNA byla provedena automatická DNA sekvenace. Sekvence cDNA byla následující (SEKV. ID. Č. 2):

```
atgggtttcaacaattgggcccgcttcatgtgcgacctcaacgagaccctgtttaccgag       60
actgccgatgcgatggctgctaacggtctgcgggacgcaggctacaatcgcatcaatctg      120
gatgattgctggatggcttatcagcgatccgacaatggatccctacggtggaacacgact      180
gagttcccacacggcctgccttggctagctcaatatgtcaaagccaaagggtttcatttt      240
ggaatctatgaagattctggcaacatgacttgtggcggatacccggatcctacaaccac      300
gaggagcaggacgccaacacctttgctttatgggggattgactatctcaagctcgacggt      360
tgcaacgtctacgcaacacaaggtaggacactcgaggaggaatacaagcaacgctacgga      420
cattggcaccaagtcctcagcaagatgcagcacccactgatcttctccgagtcagccccg      480
gcatacttcgcgggcacagacaacaacacagactggtacaccgtgatgaactgggtcccg      540
atctacggggagctggcccgccattctaccgatatcctggtgtacagtggagcaggtagc      600
gcatgggacagcattatgaataactacaactacaacactcttcttgcgcgctaccagcga      660
ccggggtatttcaatgatcctgattttctgatcccggatcatcctggcctgacggcggat      720
gaaaagcgatcgcattttgcactgtgggcttctttctcggctccacttattatcagtgct      780
tatatacctgcactttcgaaggatgagattgccttcttgacgaacgaggcattgattgcg      840
gtgaatcaggatcccctggcccagcaggccacgtttgcgagccgcgataacacactggat      900
atattgacgcgtaatctggcaaacggcgacaggctgctgacggtgcttaataagggaaac      960
acaactgtaacgagggacattcccgtacaatggttgggtcttacagagactgactgtaca     1020
tacacggccgaggatctctgggatggcaagacccagaagatcagcgaccatataaagatt     1080
gaactagccagccatgcgacagcagtcttccggctcggtcttccgcagggttgttcctcg     1140
gtagtgccaacgggacttgtcttcaacacagcatcgggcaattgtctgaccgctgcctca     1200
aattcttcagtcgcatttcagtcctgcaatggagagacctctcagatctggcaggtgaca     1260
ctgtcaggagtcattcgtccagtatcgcagaccacacaatgcttggctgctgatggaaac     1320
tcagttaagctgcaagcatgtgacagcacggatagcgacggccagaactggacgtatgca     1380
gtcacgggaaatttaaagaatgcgaagacagatggttgcctgactgagggatcagtgcag     1440
atgaagtcgtgtctttatgaga
```

1.2 Příprava aktivního rekombinantního enzymu.

Expresní plasmid byl vnesen do kvasinkových buněk *Saccharomyce cerevisiae* W303 metodou elektroporace. Podmínky elektroporace byly 25μF; 22ns; 1.2kV; elektroporace byla provedena na přístroji MicroPulserTM Electroporator (Bio-Rad). Takto připravené transformované buňky byly rozsety na Petriho misky s SC selektivním médiem neobsahujícím uracil a s přídavkem adeninu a aminokyselin tryptofanu, leucinu a histidinu. Z takto připravených transformantů byl vybrán produkční klon, který je použit při inokulaci 15 ml SC selektivního média obsahujícím 2% glukosu, tryptofan, leucin, adenin a histidin, neobsahujícím však uracil který zde slouží jako selekční marker. Buněčná suspenze byla kultivována 12 hodin, při 30˚C při intenzivním třepání 240 ot./min.. Buněčné pelety byly odstředěny centrifugací 5000g/5min a resuspendovány ve 200ml SC media obsahujícím 2% galaktózu, tryptofan, leucin, adenin a histidin, neobsahujícím však uracil. Tato buněčná kultura byla kultivována 72 hodin při 30˚C, 240 ot./min..

1.3 Izolace α-N-acetylgalaktosaminidázy z kvasinkových buněk

Po ukončení inkubace byla biomasa odstředěna centrifugací a resuspendována v 10 ml lyzačního pufru obsahujícím 50mM citrát-fosfát (pH 3,5), 5% glycerol, 1 mM PMSF a1mM dodecyl maltosid. Dále byla ke směsi přidána skleněné kuličky o průměru r = 0,25-0,5 mm (Pierce), které slouží k mechanickému rozbití buněčných stěn kvasinek. Celá směs byla 1 minutu vortexována a jednu minutu ponechána na ledu. Tento proces byl opakován pětkrát. Po ukončení izolace byla směs odstředěna centrifugací, peleta byla odstraněna a supernatant obsahující aktivní α-*N*-acetylgalaktosaminidázu byl použit při purifikaci.

1.4 Purifikace rekombinantní α-N-acetylgalaktosaminidázy

Pro purifikaci byl použit HPLC systém BioSys (Beckman-Coulter) s UV detekcí. Buněčný lyzát vzniklý v kroku 1.3 byl naředěn 60 ml solventu A (1 M NaCl in 50 mM citrát-fosfát pufr, pH 3,5) a suspense byla centrifugována při 18000 ot./min., 30 min. při 4 °C. Buněčný

extrakt byl nastříknut na kolonu Phenyl-Sepharose HR (2,6 × 10,6 cm, Merck) ekvilibrovanou v solventu A. Enzym byl eluován použitím lineárního gradientu 0-1 M NaCl, 90 min při průtoku 2 ml/min. Frakce obsahující α-*N*-acetylgalaktosaminidázovou aktivitu byly spojeny a převedeny do pufru C (20 mM citrát-fosfát, pH 4,5). Převod byl proveden pomocí ultrafiltračních kolonek Amicon Ultra 10 kDa (Millipore). Pomocí stejných kolonek byl objem frakcí snížen na cca 2ml. Takto připravený roztok obsahující aktivní enzym byl nanesen na kolonu S-Sepharose FF (1,6 × 12,5 cm, Merck), ekvilibrovanou v pufru C. Enzym byl eluován lineárním gradientem 0-1 M NaCl, 60 min při průtoku 1ml/min. Získané frakce obsahující enzymovou aktivitu byly zakoncentrovány opět pomocí ultrafiltračních kolonek Amicon Ultra 10 kDa (Millipore) na objem cca 100µl. S takto připraveným vzorkem byla provedena gelová filtrace jako poslední krok purifikace α-*N*-acetylgalaktosaminidázy (obr. 1, tab. 1).

Byla použita kolona Superdex 200 10/300 GL (1,0 × 30 cm) (Amersham Bioscience) ekvilibrovaná v 50 mM Na-acetátovém pufru pH 3,6. Průtok mobilní fáze byl 0,4ml/min. V jednotlivých eluovaných frakcích byla měřena α-*N*-acetylgalaktosaminidázová a α-galaktosidázová aktivita (obr. 2). Frakce vykazující enzymovou aktivitu byly uchovány při 4°C.

Koncentrace proteinu byla stanovena metodou dle Bradfordové s použitím standartu BSA. Poté byl protein zakoncentrován na cca 1mg/ml s pomocí koncentračních kolonek Amicon Ultra 10kDa (Millipore).

Tabulka 1. *Bilanční tabulka purifikace rekombinantní α-N-acetylgalaktosaminidázy exprimované v kvasinkovém expresním systému Saccharomyces cerevisiae*

Stupeň	Protein (mg)	Aktivita (U)	Spec.aktivita (U mg^{-1})	Čistota (fold)	Výtěžnost (%)
Buněčný extrakt	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

1.5 Stanovení enzymové aktivity

α-*N*-acetylgalaktosaminidázová aktivita byla stanovena použitím 2 mM *o*-nitrofenyl-2-acetamido-2-deoxy-α-D-galaktopyranosidu (Sigma). Jednotka enzymové aktivity byla definována jako množství enzymu potřebné k uvolnění 1μmol *o*-nitrofenolu za jednu minutu v 50mM citrát-fosfátovém pufru při pH=3,5 a teplotě 35°C. Reakční směs byla inkubována 35°C, 10 minut, uvolněný *o*-nitrofenol byl stanoven spektroskopicky při vlnové délce 420 nm v alkalickém prostředí. Z reakční směsi bylo odebráno 50μl a přidáno do 1ml 0,1 M Na_2CO_3. K sestrojení kalibrační přímky bylo použito 0~2 mM *o*-nitrofenolu.

Bylo stanoveno pH optimum a oblasti pH, kde rekombinantní α-*N*-acetylgalaktosaminidáza podle vynálezu a přirozeně izolované α-*N*-acetylgalaktosaminidáza vykazují enzymatickou aktivitu. α-*N*-acetylgalaktosaminidáza izolovaná z vláknité houby *Aspergillus niger* vykazovala při pH 6 již velmi nízkou aktivitu (přibližně 10 % maxima při pH 2), při neutrálním pH, konkrétně při pH 7 byla aktivita již zcela neměřitelná (nulová). Naproti tomu rekombinantně připravená α-*N*-acetylgalaktosaminidáza podle vynálezu vykazovala v rozmezí pH 6 až 8 α-*N*-acetylgalaktosaminidázovou aktivitu na úrovni 15 až 30 % (při pH 7 přibližně 20 %) vzhledem k maximu, kterého je dosaženo přibližně při pH 2 (obr. 4).

1.6 Identifikace rekombinantního enzymu

Identifikace α-*N*-acetylgalaktosaminidázy byla provedena metodou peptidového mapování pomocí hmotnostní spektrometrie MALDI-TOF (Bruker). Vzniklé tryptické fragmenty byly porovnány z databází v programu Mascot (Matrix Science) a byla potvrzena identita enzymu, který má následující sekvenci aminokyselin (SEKV. ID. Č. 1):

```
MGFNNWARFMCDLNETLFTETADAMAANGLRDAGYNRINLDDCWMAYQRSDNGSLRWNTT  60
EFPHGLPWLAQYVKAKGFHFGIYEDSGNMTCGGYPGSYNHEEQDANTFALWGIDYLKLDG 120
CNVYATQGRTLEEEYKQRYGHWHQVLSKMQHPLIFSESAPAYFAGTDNNTDWYTVMNWVP 180
IYGELARHSTDILVYSGAGSAWDSIMNNYNYNTLLARYQRPGYFNDPDFLIPDHPGLTAD 240
EKRSHFALWASFSAPLIISAYIPALSKDEIAFLTNEALIAVNQDPLAQQATFASRDNTLD 300
ILTRNLANGDRLLTVLNKGNTTVTRDIPVQWLGLTETDCTYTAEDLWDGKTQKISDHIKI 360
ELASHATAVFRLGLPQGCSSVVPTGLVFNTASGNCLTAASNSSVAFQSCNGETSQIWQVT 420
LSGVIRPVSQTTQCLAADGNSVKLQACDSTDSDGQNWTYAVTGNLKNAKTDGCLTEGSVQ 480
MKSCLYE
```

Výtěžek z 1 litru buněčné suspenze se pohyboval přibližně okolo 1,5 mg aktivního enzymu se specifickou aktivitu 28 U/mg.

Příklad 2

Stanovení enzymové kinetiky reakce rekombinantní α-N-acetylgalaktosaminidázy

Měření enzymové kinetiky bylo prováděno diskontinuálně. Jako substrát byl použit roztok *o*NP-α-GalNAc o pěti různých koncentracích (0,5mM; 0,7mM; 1mM; 2mM; 5mM). K reakční směsi bylo vždy přidáno 5µl roztoku rekombinantní α-*N*-acetylgalaktosaminidázy o koncentraci 0,1mg/ml a 50µl roztoku substrátu. Takto připravená reakční směs byla inkubována při 35°C za stálého třepání. Probíhající enzymová reakce byla zastavena v šesti odlišných časech (0 min; 2 min; 4 min; 6 min; 8 min; 10 min) přidáním 150µl 1M Na_2CO_3 do reakční směsi. Naměřená data byla vyhodnocena s využitím programu Enzfitter. Byla určena hodnota Michaelisovy konstanty pro *o*NP-α-GalNAc a hodnota limitní rychlosti enzymové reakce (obr. 3).

Průmyslové využití

Rekombinantní enzymaticky aktivní α-*N*-acetylgalaktosaminidáza se uplatní při přeměně krevních skupin typu A na krevní skupinu typu H(0) univerzální dárce, při výrobě krevních derivátů. Dále se tento rekombinantně připravený enzym může použít při syntéze oligosacharidových či polysacharidových řetězců. Takové sacharidové struktury mohou být použity v mnoha průmyslových odvětvích (farmaceutický, potravinářský, chemický průmysl). Vzhledem k duální aktivitě α-*N*-acetylgalaktosaminidázy je možno tento enzym použít jako modelový protein vykazující α-galaktosidázovou aktivitu v oblasti nízkého pH.

SEZNAM SEKVENCÍ

<110> Univerzita Karlova v Praze, Přírodovědecká fakulta, Praha, CZ

<120> Aktivní forma α-*N*-acetylgalaktosaminidázy z vláknité houby *Aspergillus niger* a její rekombinantní exprese

<130> P-0065-CZ

<150>

<151>

<160> 4

<210> 1

<211> 487

<212> PRT

<213> *Aspergillus niger*

<400> 1

MGFNNWARFMCDLNETLFTETADAMAANGLRDAGYNRINLDDCWMAYQRSDNGSLRWNTT 60
EFPHGLPWLAQYVKAKGFHFGIYEDSGNMTCGGYPGSYNHEEQDANTFALWGIDYLKLDG 120
CNVYATQGRTLEEEYKQRYGHWHQVLSKMQHPLIFSESAPAYFAGTDNNTDWYTVMNWVP 180
IYGELARHSTDILVYSGAGSAWDSIMNNYNYNTLLARYQRPGYFNDPDFLIPDHPGLTAD 240
EKRSHFALWASFSAPLIISAYIPALSKDEIAFLTNEALIAVNQDPLAQQATFASRDNTLD 300
ILTRNLANGDRLLTVLNKGNTTVTRDIPVQWLGLTETDCTYTAEDLWDGKTQKISDHIKI 360
ELASHATAVFRLGLPQGCSSVVPTGLVFNTASGNCLTAASNSSVAFQSCNGETSQIWQVT 420
LSGVIRPVSQTTQCLAADGNSVKLQACDSTDSDGQNWTYAVTGNLKNAKTDGCLTEGSVQ 480
MKSCLYE

<210> 2

<211> 1462

<212> DNA

<213> *Aspergillus niger*

<400> 2

atgggtttcaacaattgggcccgcttcatgtgcgacctcaacgagaccctgtttaccgag 60
actgccgatgcgatggctgctaacggtctgcgggacgcaggctacaatcgcatcaatctg 120
gatgattgctggatggcttatcagcgatccgacaatggatccctacggtggaacacgact 180
gagttcccacacggcctgccttggctagctcaatatgtcaaagccaaagggtttcatttt 240
ggaatctatgaagattctggcaacatgacttgtggcggataccccggatcctacaaccac 300
gaggagcaggacgccaacacctttgctttatgggggattgactatctcaagctcgacggt 360

tgcaacgtctacgcaacacaaggtaggacactcgaggaggaatacaagcaacgctacgga 420
cattggcaccaagtcctcagcaagatgcagcacccactgatcttctccgagtcagccccg 480
gcatacttcgcgggcacagacaacaacacagactggtacaccgtgatgaactgggtcccg 540
atctacggggagctggcccgccattctaccgatatcctggtgtacagtggagcaggtagc 600
gcatgggacagcattatgaataactacaactacaacactcttcttgcgcgctaccagcga 660
ccggggtatttcaatgatcctgattttctgatcccggatcatcctggcctgacggcggat 720
gaaaagcgatcgcattttgcactgtgggcttctttctcggctccacttattatcagtgct 780
tatatacctgcactttcgaaggatgagattgccttcttgacgaacgaggcattgattgcg 840
gtgaatcaggatcccctggcccagcaggccacgtttgcgagccgcgataacacactggat 900
atattgacgcgtaatctggcaaacggcgacaggctgctgacggtgcttaataagggaaac 960
acaactgtaacgagggacattcccgtacaatggttgggtcttacagagactgactgtaca 1020
tacacggccgaggatctctgggatggcaagacccagaagatcagcgaccatataaagatt 1080
gaactagccagccatgcgacagcagtcttccggctcggtcttccgcagggttgttcctcg 1140
gtagtgccaacgggacttgtcttcaacacagcatcgggcaattgtctgaccgctgcctca 1200
aattcttcagtcgcatttcagtcctgcaatggagagacctctcagatctggcaggtgaca 1260
ctgtcaggagtcattcgtccagtatcgcagaccacacaatgcttggctgctgatggaaac 1320
tcagttaagctgcaagcatgtgacagcacggatagcgacggccagaactggacgtatgca 1380
gtcacgggaaatttaaagaatgcgaagacagatggttgcctgactgagggatcagtgcag 1440
atgaagtcgtgtctttatgaga

<210> 3
<211> 33
<212> DNA
<213>
<220> Znaky
<223> Umělá sekvence
<400> 3
ttattctagaatgggtttcaacaattgggcccgc 33

<210> 4
<211> 39
<212> DNA
<213>
<220> Znaky
<223> Umělá sekvence
<400> 4
attgaattcttagccatccctctcataaagacacgactt 39

PATENTOVÉ NÁROKY

1. Rekombinantní α-*N*-acetylgalaktosaminidáza s duální α-galaktosidázovou aktivitou z *Aspergillus niger* mající aminokyselinovou sekvenci uvedenou jako SEKV. ID. Č. 1 projevující α-*N*-acetylgalaktosaminidázovou aktivitu při neutrálním pH.

2. Rekombinantní α-*N*-acetylgalaktosaminidáza podle nároku 1 projevující při pH 6 až 8 α-*N*-acetylgalaktosaminidázovou aktivitu v úrovni 30 až 15 % vzhledem k maximální aktivitě při pH 2.

3. Rekombinantní α-*N*-acetylgalaktosaminidáza podle nároku 1 nebo 2 projevující při pH 7 α-*N*-acetylgalaktosaminidázovou aktivitu v úrovni 20 % vzhledem k maximální aktivitě při pH 2.

4. Izolovaná molekula nukleové kyseliny kódující aminokyselinovou sekvenci α-*N*-acetylgalaktosaminidázy podle kteréhokoliv z nároků 1 až 3.

5. Izolovaná molekula nukleové kyseliny podle nároku 4, kterou je DNA mající nukleotidovou sekvenci uvedenou jako SEKV. ID. Č. 2.

6. Způsob přípravy rekombinantní α-*N*-acetylgalaktosaminiázy podle kteréhokoliv z nároků 1 až 3 **vyznačující se tím,** že zahrnuje kroky, kdy se z *Aspergillus niger* izoluje celková RNA, tato RNA se přepíše do cDNA pomocí reverzní transkriptázy, s touto cDNA se provede PCR s dvojicí oligonukleotidových primerů majících SEKV. ID. Č. 3 a SEKV. ID. Č. 4, v PCR získaný fragment DNA se klonuje do vektoru a vnese do hostitelských buněk expresního systému *Saccharomyces cerevisiae*.

7. Rekombinantní α-*N*-acetylgalaktosaminidáza podle kteréhokoliv z nároků 1 až 3 připravitelná způsobem podle nároku 6.

8. Použití dvojice oligonukleotidů sestávající ze sekvencí SEKV. ID. Č. 3 a SEKV. ID. Č. 4 pro přípravu molekuly nukleové kyseliny podle nároku 5.

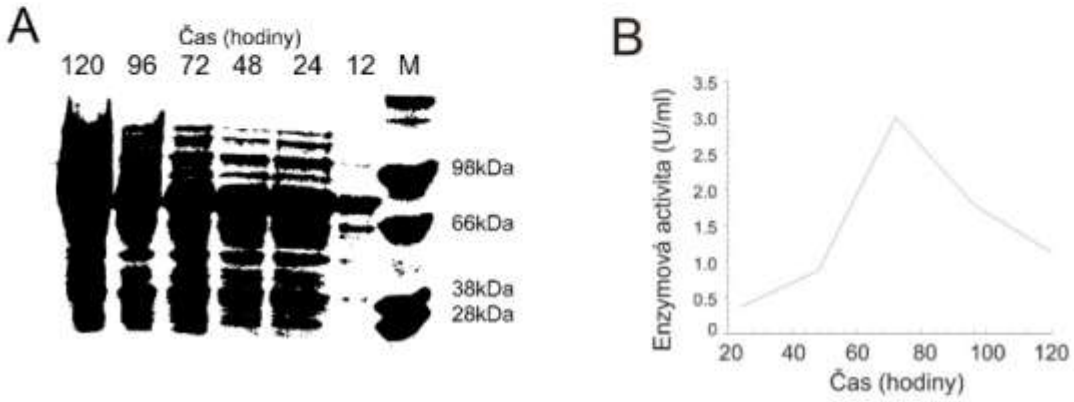
9. Použití dvojice oligonukleotidů sestávající ze sekvencí SEKV. ID. Č. 3 a SEKV. ID. Č. 4 pro přípravu rekombinantní α-*N*-acetylgalaktosaminiázy podle kteréhokoliv z nároků 1 až 3.

Anotace

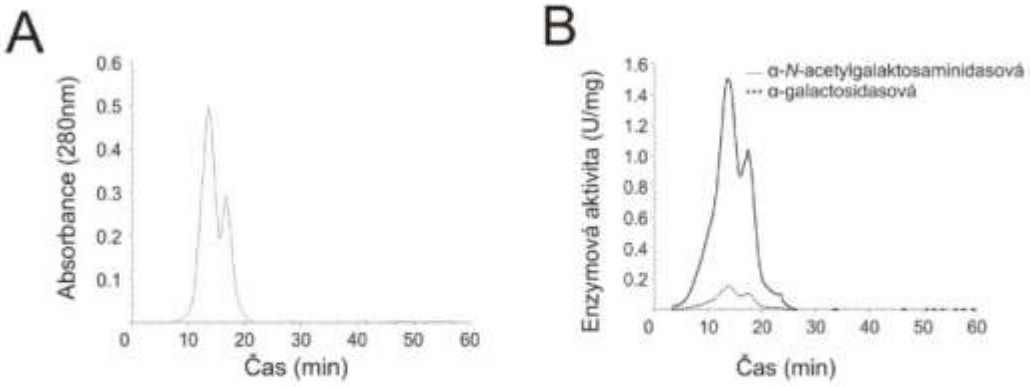
Název vynálezu: Aktivní forma α-N-acetylgalaktosaminidázy z vláknité houby *Aspergillus niger* a její rekombinantní exprese

Předkládaný vynález se týká α-*N*-acetylgalaktosaminidázy neboli α-*N*-acetylgalaktosaminid-*N*-acetylgalaktosaminohydrolázy, enzymu s duální aktivitou, a přípravy aktivní formy tohoto enzymu v kvasinkovém expresním systému *Saccharomyces cerevisiae*. α-*N*-acetylgalaktosaminidáza (α-GalNAcasa; EC 3.2.1.49) je enzym selektivně štěpící terminálně vázaný GalNAc, který je připojen O- α-D-glykosidovou vazbou na aminokyseliny Ser nebo Thr, či na oligosacharidový/polysacharidový řetězec. Tento enzym byl rekombinantně připraven v kvasinkovém expresním systému *Saccharomyces cerevisiae*. Byl biochemicky plně charakterizován a bylo zjištěno, že je aktivní. Rekombinantně připravená α-*N*-acetylgalaktosaminidáza podle vynálezu vykazuje 20% aktivitu v oblasti neutrálního pH (ve srovnání s maximem aktivity při pH 2). Této vlastnosti lze s výhodou využít při přípravě krevních buněk typu H(0) z krevních buněk typu **A**. Enzym podle vynálezu je také vhodný pro přípravu krevních derivátů změnou epitopů na povrchu krevních buněk. Dále lze z výhodou tento enzym použít při přípravě syntetických oligo/polysacharidových struktur, které mohou být použity například ve farmaceutickém či potravinářském průmyslu.

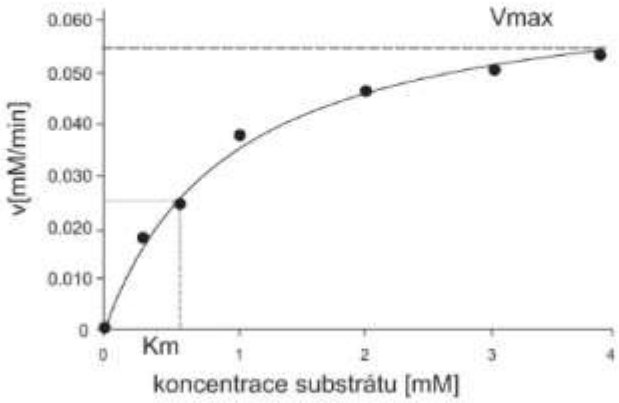
Obr. 1


A
Čas (hodiny)
120 96 72 48 24 12 M
98kDa
66kDa
38kDa
28kDa
B
Enzymová activita (U/ml)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
20 40 60 80 100 120
Čas (hodiny)

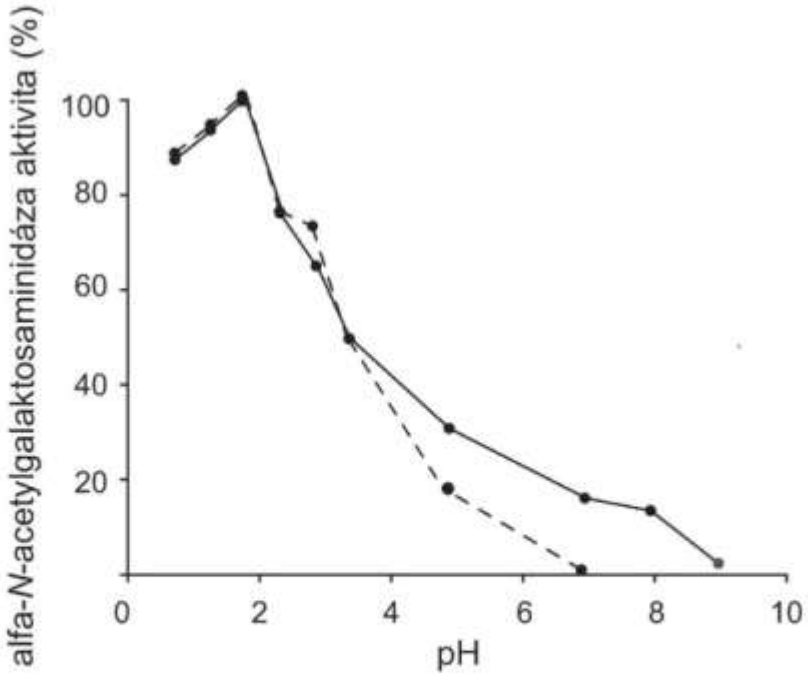
Obr. 2


A
Absorbance (280nm)
0.6
0.5
0.4
0.3
0.2
0.1
0 10 20 30 40 50 60
Čas (min)
B
Enzymová aktivita (U/mg)
α-N-acetylgalaktosaminidasová
α-galactosidasová
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0 10 20 30 40 50 60
Čas (min)

Obr. 3


Vmax
0.060
0.050
0.040
0.030
0.020
0.010
0
v[mM/min]
0
Km
1
2
3
4
koncentrace substrátu [mM]

Obr. 4


alfa-N-acetylgalaktosaminidáza aktivita (%)
100
80
60
40
20
0
2
4
6
8
10
pH

Appendix 4

Kavan D., Kubickova M., Bily J., Vanek O., Hofbauerova K., Mrazek H., Rozbesky D., Bojarova P., Kren V., Zidek L., Sklenar V., Bezouska K.

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 (2010).

4060 *Biochemistry* **2010,** *49,* 4060–4067
DOI: 10.1021/bi100181a

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Cooperation between Subunits Is Essential for High-Affinity Binding of *N*-Acetyl-D-hexosamines to Dimeric Soluble and Dimeric Cellular Forms of Human CD69[†]

Daniel Kavan,[‡,§,⊥] Monika Kubičková,[§,⊥] Jan Bílý,[‡] Ondřej Vaněk,[‡,∥] Kateřina Hofbauerová,[§] Hynek Mrázek,[‡,∥] Daniel Rozbeský,[‡,∥] Pavla Bojarová,[‡,∥] Vladimir Křen,[§] Lukáš Žídek,[§] Vladimir Sklenář,[§] and Karel Bezouška*[,‡,∥]

[‡]*Department of Biochemistry, Faculty of Science, Charles University, 12840 Prague, Czech Republic,* [§]*Institute of Microbiology v.v.i., Academy of Sciences of Czech Republic, 14220 Prague, Czech Republic, and* [∥]*National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic* [⊥]*These authors contributed equally to the experiments*

Received July 21, 2009; Revised Manuscript Received April 6, 2010

ABSTRACT: CD69 is an earliest lymphocyte activation antigen and a universal leukocyte triggering molecule expressed at sites of active immune response. The binding of GlcNAc to the dimeric human CD69 was followed by equilibrium dialysis, fluorescence titration, and NMR. Clear cooperation was observed in the high-affinity binding ($K_d = 4.0 \times 10^{-7}$ M) of the carbohydrate to two subunits of the dimeric CD69 (Hill coefficient 1.94). A control monosaccharide ManNAc was not bound by human CD69, and both monosaccharides had no effects on the structure of the receptor. However, a monomeric CD69 obtained by mutating Q93 and R134 at the dimer interface exhibited a much lower affinity for GlcNAc ($K_d = 1.3 \times 10^{-5}$ M) and no cooperativity (Hill coefficient 1.07). Perturbation of the dimer interface resulted in a severe impairment of the signaling ability of cellular CD69 when cross-linked with an antibody or with a bivalent high-affinity *N*-acetylhexosamine dimer-based ligand. The availability of stable preparations of soluble CD69 receptor with well-documented ligand binding properties will be beneficial for immunological experiments evaluating the role of this antigen in the complex environment of the immune system. Moreover, such preparations in combination with efficient ligand mimetics able to both activate CD69[+] lymphocytes and to block undesired hyperactivation caused by other cellular ligands will also become indispensable tools in explaining the exact role of the CD69 antigen in the interaction between the tumor cell and the effector natural killer lymphocyte.

CD69 is an early lymphocyte activation marker and a universal leukocyte triggering molecule expressed at sites of active immune response and chronic inflammation (*1*, *2*). Initial *in vitro* studies suggested that CD69 may function as an activating molecule in many leukocyte subsets including γ/δ T-cells and natural killer (NK)[‡] cells (*3*, *4*). The *CD69* gene is located within the NK gene complex on human chromosome 12. It codes a type II calcium-dependent membrane lectin, a member of one important family of abundant NK cell surface receptors (*5*, *6*) participating in the formation of the receptor "zipper" at the NK cell−tumor cell interface (*7*). Most receptors of NK cells that recognize target structures at the surface of tumor or virally infected cells mediate their activation or inhibitory effects through their sequentially diverse cytoplasmic domains (*8*). However, the cytoplasmic domain of CD69 is short and lacks the prominent function-associated peptide motifs. Previous studies have shown that the downstream activation processes initiated by CD69 engagement occur through Src-dependent activation of Syk, activation of phospholipase Cγ2 and Vav, and the subsequent transmission of signals through the Rac−ERK pathway (*9*, *10*). Alternatively, CD69 can also propagate activation signals through heterotrimeric G proteins and the subsequent intracellular signaling pathways coupled to these molecular switches (*11*, *12*). Recently, *in vivo* studies in CD69-deficient mice have added yet another dimension into the biology of this receptor revealing its non-redundant role in downregulation of the immune response through the production of the pleiotropic cytokine TGF-β and through interactions with regulatory T-cells (*2*). Using the same experimental model, it has been shown recently that CD69 forms a complex with sphingosine 1-phosphate receptor, negatively regulates its function, and thus inhibits lymphocyte egress from lymphoid organs downstream of interferon-α/β, known mediators of transient egress shut down (*13*). Furthermore, results of many immunological studies indicate that CD69 may be involved in pathogenesis of several diseases including rheumatoid arthritis, chronic inflammatory liver diseases, mild asthma, and acquired immunodeficiency syndrome (*14*).

The identification of the natural ligand for CD69 is a key critical step for further advancement of our knowledge on the biology of this receptor. The initial findings that CD69 binds to calcium and certain *N*-acetyl-D-hexosamines (*15*) could not be later reproduced using a somewhat different expression construct (*16*). Since then, these discrepancies have been at least partially explained by careful structural evaluations of the recombinant proteins used for binding studies, as well as by

[†]Supported by grants from Ministry of Education of Czech Republic (MSM_21620808, 1M0505, and AVOZ50200510 to K.B. and MSM0021622413 and LC0603 to V.S.) and from Czech Science Foundation (303/09/0477 and 305/09/H008 to K.B. and 203/09/P024 to P.B.) and by EU Project Spine 2 (Contract LSHG-CT-2006-02/220 to K.B.).

*To whom correspondence should be addressed at Charles University. Phone: +420-2-2195-1272. Fax: +420-2-2195-1283. E-mail: bezouska@biomed.cas.cz.

[‡]Abbreviations: DSS, disucinimidyl suberate; Gal, D-galactose; GlcNAc, *N*-acetyl-D-glucosamine; ManNAc, *N*-acetyl-D-mannosamine; MES buffer, 10 mM MES with 150 mM NaCl and 1 mM NaN_3; MES + C buffer, 10 mM MES, 150 mM NaCl, 1 mM $CaCl_2$, and 1 mM NaN_3; NK, natural killer.

 Published on Web 04/06/2010

establishing a direct link between the binding of calcium and carbohydrates (*17*). The proper folding of CD69 produced based on the recently suggested constructs encompassing G70−K199 of the entire receptor was established using detailed structural experiments based on both NMR (*18*) and protein crystallography (*19*). The most recent development of efficient structural mimetics of the high-affinity ligand for CD69 opened the way for manipulating with numerous activities of CD69 at the molecular and cellular level (*20*, *21*) and provided efficient compounds for further *in vivo* testing of their immunomodulating properties (*22*−*24*).

Here we report new findings indicating that the binding of *N*-acetyl-D-hexosamines 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 cross-linking 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.

EXPERIMENTAL PROCEDURES

Materials. All chemicals were analytical grade reagents of the best quality available commercially and were obtained from Sigma unless indicated otherwise. The preparation of dimeric *N*-acetylhexosamine disaccharide with the chemical composition (GalNAcβ1−4GlcNAcβ-NH-CS-NH-CH_2-)$_2$ by a combination of chemical and enzymatic steps has been described previously (*25*).

Preparation of Soluble Dimeric CD69. Preparation of soluble dimeric CD69 using protocol II has been described previously (*18*). For the preparation of a uniformly ^{15}N-labeled form of the receptor, a producing culture of *Escherichia coli* BL-21 Gold (Stratagene) harboring the expression plasmid was used, grown on a standard M9 minimal medium containing $^{15}NH_4Cl$.

Identification of Key Amino Acid Residues Disrupting the Receptor Dimer Interface. We examined the three-dimensional structure of the crystallized soluble dimeric CD69 deposited into the RCSB Protein Databank under accession code 3CCK (*18*). Glutamine Q93 appeared to be involved in two key hydrogen bonds between the amide group of its side chain and two adjacent acidic residues belonging to the opposite subunit, Asp88 and Glu87. With R134, the intertwining with the second subunit is even more profound, and the guanidyl group of this amino acid forms three hydrogen bonds with A136 and Y135 of the opposing subunit and is also involved in the stacking interaction with the phenyl ring of Y135.

Site-Directed Mutagenesis and Expression of the Mutated CD69 Proteins. Mutated forms of CD69 in which the Q93 and/or and R135 were mutated into A were produced using the CD69 expression plasmid in pRSETB (*18*). Mutations were introduced using the QuickChange site-directed mutagenesis kit (Stratagene) in combination with the following oligonucleotide pairs: CD69Q93F, 5′-GAGGACTGGGTTGGCTACGCGAG-GAAATGCTACTTTATT-3′, and CD69Q93R, 5′-AATAAA-GTAGCATTTCCTCGCGTAGCCAACCCAGTCCTC-3′, and CD69R134F, 5′-GACATGAACTTTCTAAAAGCATACGC-AGGTAGAGAGGAA-3′ and CD69R134R, 5′-TTCCTCTCT-ACCTGCGTATGCTTTTAGAAAGTTCATGTC. The introduced mutations were verified by DNA sequencing using an ABI Prism 3130 genetic analyzer (Applied Biosystems). The CD69Q93A/R134A double mutant was prepared sequentially, applying the R134A mutation process on the Q93A mutant. Mutated CD69 proteins were prepared using the same protocol (protocol II) used for the production of the wild-type protein (*18*). Moreover, the proper refolding of the protein was verified using NMR measurement with the homogenously ^{15}N-labeled proteins as described previously (*18*).

Gel Filtration. Gel filtration was performed using a Superdex 200 HR 10/30 column (GE Healthcare) connected to the protein purification system BioSys510 (Beckman Coulter) and equilibrated with MES buffer at room temperature. In order to examine the effect of monosaccharide binding to CD69 on the hydrodynamic volume of the CD69 protein, the protein samples were incubated overnight at 4 °C in the presence of 1 mM ManNAc or 1 mM GlcNAc and then injected onto the gel filtration column equilibrated in MES + C buffer containing 1 mM concentrations of the respective monosaccharides.

Analytical Ultracentrifugation. Sedimentation velocity and sedimentation equilibrium experiments were performed using a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter) equipped with an An50Ti rotor and dual absorbance and laser interference optics. Before the experiment, 0.5 mL samples of CD69 proteins diluted to 0.4 mg·mL^{-1} were dialyzed for 20 h against 2 L of MES + C buffer, with or without the monosaccharides, in concentration indicated in the text, and the dialysis buffer was used as a reference and sample dilution buffer. The sedimentation velocity experiment was conducted at 48000 rpm for dimeric CD69 at 20 °C. Data were analyzed with the program SEDFIT (*26*, *27*). Based on buffer composition and amino acid sequence using the program SEDNTERP (www.jphilo.mailway.com), buffer density and CD69 partial specific volume for CD69NG70 were estimated as 1.00309 g·mL^{-1} and 0.7183 mL·g^{-1}, respectively.

Protein Stability Experiments. CD69 proteins were diluted to 0.5 mg/mL, and UV spectra were taken in the 200−300 nm range in a Beckman DU-70 spectrophotometer (Beckman Coulter) equipped with a heated cuvette. The initial UV scan was taken at 25 °C, after which the temperature in the cuvette was increased in 5 °C increments. Experiments were performed routinely in MES + C buffer. Alternatively, protein stability was verified using differential scanning calorimetry and FTIR spectroscopy as has been previously described (*17*, *18*, *28*).

NMR Titrations. All NMR experiments were run at 300 K in a Bruker Avance 600 MHz spectrometer equipped with a cryogenic H/C/N TCI probehead. ^{1}H−^{15}N HSQC spectra of 0.3 mM ^{15}N-labeled wild-type CD69 protein CD69NG70 (*18*) were used as a routine check of protein folding and stability. The sample buffer consisted of 10 mM MES, pH 5.8, with 49 mM NaCl, 1 mM NaN_3, and 10% D_2O. During NMR titration, a 0.1 mM solution of the unlabeled wild-type CD69 protein (*7*) was titrated. In an initial experiment, aliquots of the GlcNAc ligand corresponding to 25%, 50%, 75%, 100%, 200%, and 500% of saturation were added, and signals of the free GlcNAc ligand were observed at 2.2 ppm in the 1D proton spectra and used for the estimation of the free ligand concentration. In a separate experiment aimed at estimating the binding constant, smaller ligand additions were used as the equivalence was appoached. The protein was titrated to 75% of the estimated number of binding sites, after which the amount of ligand was increased in increments of 5% of the estimated number of binding sites until the equivalence point was reached. All spectra were processed

using the software NMRPIPE (*29*). The dissociation constant K_d, defined as $K_d = (c_P - c_L + [L])[L]/(c_L - [L])$, was obtained by a nonlinear fitting of the [L] vs c_L titration curves (Figure 2A,B).

Equilibrium Dialysis. *N*-Acetyl-D-[1-^{3}H]glucosamine (specific activity 500 GBq/mmol) and *N*-acetyl-D-[1-^{3}H]mannosamine (specific activity 650 GBq/mmol) were prepared as described previously (*17*) or purchased from Amersham. To set up equilibrium dialysis experiments, a rotating apparatus with glass blocks containing separate sealable chambers with external access was used as described previously (*17*). Aliquots (200 μL) of 0.1 μM solutions of CD69 proteins in MES + C buffer were incubated with varying amounts of ligand at 27 °C (300 K) for 48 h. After equilibration, 100 μL aliquots were withdrawn from the control and from the protein-containing chambers. The results were calculated and plotted according to Scatchard as described previously (*17*).

Tryptophan Fluorescence Quenching. Tryptophan fluorescence quenching experiments were performed according to the described methodology (*30*) with minor modifications. In initial experiments, 100 nmol aliquots of CD69 protein were pipetted into multiple wells of a UV Star plate (Greiner, Germany) and mixed with 10-fold serial dilutions of the GlcNAc ligand. Incubation proceeded for 1 h at 27 °C (300 K) in the thermostated chamber of a Safire2 plate reader (Tecan, Austria), after which the fluorescence of tryptophan residues was measured in duplicate wells using the bottom fluorescence measurements and the following settings: λ_{ex} = 275 nm, λ_{em} = 350 nm, excitation and emission slits were set to 5 and 20, respectively, and the fluorescence gain was manually set to 66. After finding the lowest concentration of ligand that still caused the quenching of tryptophan fluorescence, detailed dilutions of the ligand by 10% saturation steps were performed, and the concentration of free and bound ligand was calculated as described previously (*17*, *30*).

Preparation of the Eukaryotic Expression Constructs, Transfection into Jurkat Cells, and Selection of the Transfectants. In order to mutate the dimerization cystein C68 (*15*) to A, site-directed mutagenesis was performed using the original expression plasmid (*15*) as described above using oligonucleotide primers CD69C68F, 5′-TCAGTGGGCCAATACAATGCTCCAGGCCAATACACATTC-3′, and CD69C68R, 5′-GAATGTGTATTGGCCTGGAGCATTGTATTGGCCCACTGA-3′, and the pCDA401 plasmid (*15*). Single mutation CD69C68A, double mutations CD69C68A/Q93A and CD69C68A/R134A, and the triple mutation CD69C68A/Q93A/R134A were prepared by applying the mutagenesis protocol onto expression plasmids for wild-type CD69 (*8*) and for the respective dimerization mutants described above. After the mutagenesis and DNA sequencing, DNA fragments coding the C-terminal extracellular segments of CD69 were linked with the DNA fragment coding the N-terminal part of the receptor (*31*) using linking PCR (*8*). The X construct corresponded to the religated pCR3 (mock) and was used as a control (*8*). The eukaryotic expression vectors were sequenced and transfected into a Jurkat T lymphoblastoid cell line maintained in RPMI1640 and supplemented with 10% fetal calf serum (*8*).

Precipitation of Cellular Forms of CD69 Using Antibodies and Dimeric N-Acetylhexosamines. Transfected Jurkat cells (1×10^6) were surface radioiodinated using lactoperoxidase (*31*), washed three times with medium, and then incubated with 1 mM concentrations of dimeric *N*-acetylhexosamine disaccharides for 1 h at room temperature. The incubation was followed by the addition of 100 mM DSS and by cross-linking of the receptors for another 1 h at 4 °C. Thereafter, cells were lysed, and CD69 receptor complexes were immunoprecipitated using G protein beads (GE Healthcare) coated with monoclonal antibodies against CD69, BL-KFB/B1 (*31*). Beads were washed extensively, boiled in sample buffer for SDS−PAGE, and analyzed using 15% SDS−polyacrylamide gels followed by autoradiography.

Cellular Activation Assays and Production of IL-2. Transfected Jurkat cells (10^6) were incubated with dimeric *N*-acetylhexosamine disaccharides as described in the preceding section or with saturating concentrations of monoclonal antibodies against CD69 for 5 min (cellular activation) or 12 h (IL-2 production) and used to determine the free cytoplasmic calcium (*11*) or IL-2 production (*8*).

RESULTS AND DISCUSSION

Evaluation of Calcium and Carbohydrate Binding Activity of Highly Stable CD69 Proteins. We and others have previously generated several constructs optimized for the preparation of highly stable soluble recombinant CD69 proteins suitable for ligand identification experiments (Supporting Information Table S1). Preliminary ligand binding experiments were performed to evaluate the ability of these constructs to bind calcium and monosaccharides shown to be important ligands for the receptor (*15*, *17*). With regard to the binding of calcium, there has been no difference in the ability to bind calcium between the covalent dimeric protein CD69CQ65 and noncovalent dimeric proteins CD69NG70 and CD69NV82 when compared to the monomeric protein CD69MS100: each of these proteins bound 1 mol of calcium/mol of CD69 subunit with K_d of approximately 58 μM (ref *17* and Supporting Information Figure S1). On the other hand, significant differences between these protein constructs were observed with regard to the binding of *N*-acetyl-D-hexosamines. While the IC_{50} values for the soluble monomeric CD69, CD69MS100, with regard to binding of the two active *N*-acetyl-D-hexosamines, D-GlcNAc and D-GalNAc, were each approximately 10^{-5} M, these values were about 10 times lower for the dimeric protein CD69NV82 and about 100 times lower for the other two highly stable dimeric proteins, CD69CQ65 and CD69NG70 (Supporting Information Figure S2). The latter protein has been selected for all of the subsequent binding experiments and will be referred to as soluble dimeric CD69. The homogeneity and monodispersity were routinely evaluated for each batch of the produced soluble dimeric CD69 using SDS electrophoresis under both reducing and nonreducing conditions and gel filtration on a Superdex 200 HR column (ref *18* and Figure 1). Moreover, the identity, quality, and proper refolding of each batch of the produced protein were also verified as described previously (*18*) using high-resolution ion cyclotron resonance mass spectrometry, one-dimensional proton NMR, thermal stability experiments, and tests of the biochemical stability (ref *18* and Table 1).

Cooperativity of GlcNAc Binding Proved by Direct Binding Experiments. The detailed binding studies with soluble dimeric CD69 were performed using D-GlcNAc as the high-affinity carbohydrate ligand, together with D-ManNAc and, in some experiments, D-Gal as negative controls. The initial evidence for the interaction of the soluble dimeric CD69 with GlcNAc was obtained by NMR titration. A 0.1 mM solution of the dimeric receptor was titrated up to equivalence assuming the existence of two high-affinity binding sites per receptor

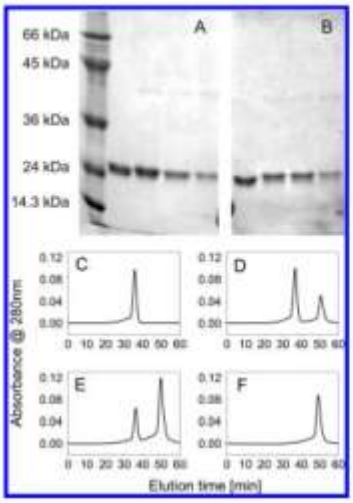


FIGURE 1: Analysis of wild-type and mutant CD69 proteins by SDS−PAGE and gel filtration. In (A) and (B), these proteins were analyzed under reducing and nonreducing conditions, respectively. Analyzed proteins, from left to right, were wild-type CD69, CD69Q93A mutant, CD69R134A mutant, and CD69QRDM. Marker proteins were BSA (65 kDa), ovalbumin (44 kDa), lactoglobulin (18 kDa), lysozyme (14 kDa), and aprotinin (6 kDa). In (C) to (F), these proteins were analyzed by gel filtration on a Superdex 200 HR column, and the four respective panels contain chromatograms for wild-type CD69, CD69Q93A mutant, CD69R134A mutant, and CD69QRDM double mutant.

Table 1: Summary of Stability Properties of Wild-Type Dimeric CD69 and CD69 Dimerization Mutants

protein	characteristics	T_d^a (°C)	T_d^b (°C)	T_d^c (°C)
CD6CD69WT	noncovalent dimers	65	67	65
CD69Q93A	dimer/monomer equilibrium	63	62	64
CD69R134A	dimer/monomer equilibrium	62	60	60
CD69QRDM	monomeric	60	62	61

[a]Determined from thermal UV denaturation measurements. [b]Determined from differential scanning calorimetry. [c]Determined from FTIR spectroscopy.

dimer (*17*). The results of this experiment (Figure 2B and Supporting Information Figure S3B) confirmed the specific binding of GlcNAc to the receptor (2 mol of GlcNAc bound to a receptor dimer) and provided an affinity estimation in the low micromolar range ($K_d = 4.0 \times 10^{-7}$ M). On the other hand, no interaction could be seen with the ManNAc negative control under the same experimental conditions (Figure 2A and Supporting Information Figure S3A). However, NMR did not enable the fraction of the bound ligand to be measured.

In order to confirm the results obtained by NMR titration, additional direct binding experiments were performed. When the bound and the unbound ligands had been separated by dialysis under equilibrium, two binding sites per receptor dimer were detected. Direct binding experiments enabled the degree of saturation at each particular ligand concentration to be calculated. The resulting saturation curve, showing the saturation (fraction bound normalized per receptor subunit) dependence on the ligand concentration (Figure 2C), clearly revealed a striking cooperativity in the highly specific ($K_d = 4.0 \times 10^{-7}$ M) binding

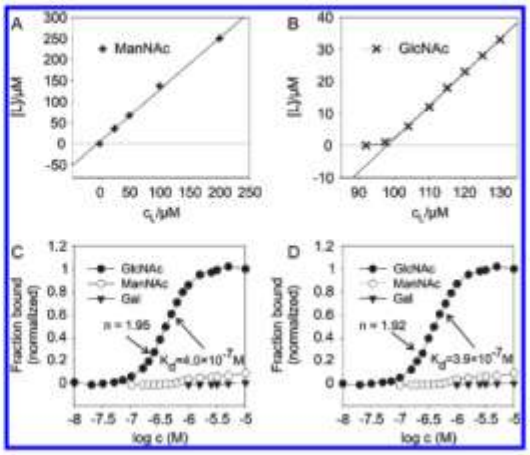


FIGURE 2: Measurements of direct interaction of soluble CD69NG70 with ManNAc and GlcNAc. (A, B) NMR titration of soluble CD69 with ManNAc and GlcNAc, respectively. (C, D) Concentration dependence of receptor saturation measured by equilibrium dialysis and tryptophan fluorescence quenching, respectively, using GlcNAc, ManNAc, and Gal as indicated.

of GlcNAc to the receptor with the Hill coefficient approaching the maximum theoretical value (theory 2.00, experiment 1.95; see Figure 2C). These results were also independently confirmed by the third binding assay, the fluorescent titration, which gave binding parameters essentially identical to those obtained by the equilibrium dialysis (Figure 2D). On the other hand, very little specific binding for both ManNAc and Gal control monosaccharides could be seen in both of the latter assays (Figure 2C,D).

Binding of N-Acetyl-D-hexosamines Did Not Result in Significant Conformational Change in CD69 Protein. To analyze the structural changes of soluble CD69 upon ligand binding, variations in the hydrodynamic properties of the receptor were investigated. The molecular size of the receptor, which had been saturated with an excess of GlcNAc, was studied by gel filtration on Superdex 200 HR and by analytical ultracentrifugation and compared with the size of the receptor preincubated in the ManNAc control. The elution time decrease from 36.5 to 31.7 min in the gel filtration would indicate a change in the molecular size of CD69 upon GlcNAc binding when compared to the presence of ManNAc (Supporting Information Figure S4). However, the detailed analysis of soluble dimeric CD69 in the absence of any ligand, in the presence of 1 mM ManNAc, and in the presence of 1 mM GlcNAc did not reveal any changes in hydrodynamic properties since the value of the experimentally determined sedimentation coefficient was identical (Supporting Information Figure S5).

Binding of N-Acetyl-D-glucosamine to the Stable Monomeric CD69 Follows a Single Site Model and Proceeds with Much Lower Affinity. In the next step, the interactions of GlcNAc with the monomeric subunit of CD69 were studied. Since it proved extremely difficult to prepare the monomeric form of the receptor by dissociation of the CD69 dimer (*18*), we used the available crystal structure of the CD69 dimer and analyzed the dimer interface for critical residues participating in the dimerization. Two such critical residues, namely Q93 and R134, both interacting with two residues of the other subunit, could be identified (Supporting Information Figure S6).

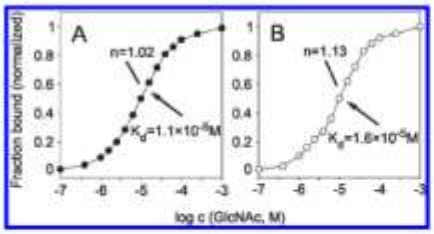


FIGURE 3: Binding of GlcNAc to the monomeric subunit of CD69. (A, B) Binding of GlcNAc to monomeric CD69 analyzed using equilibrium dialysis and fluorescence titration, respectively.

These amino acid residues were mutated to alanine, singly or in combination. All three produced mutated proteins were after their refolding and purification extensively verified using the methodology described previously (*18*) for the wild-type protein CD69NG70 using SDS electrophoresis and ion cyclotron resonance mass spectrometry of the entire protein, as well as nuclear magnetic resonance to check the proper folding of these proteins (Figure 1 and results not shown). The stability of the mutant soluble CD69 proteins was comparable to that of the wild-type protein, indicating that the introduced mutations did not result in any decrease of protein stability. Only the double mutant behaved as a monomeric protein (Figure 1), with stability comparable to that of the dimer receptor (Table 1). This protein was used to analyze the binding of GlcNAc to the monomeric protein. The results clearly showed that binding to the monomeric subunit of CD69 was much weaker and noncooperative (Hill coefficient of 1.07; Figure 3).

Q93/R134/C68 Triple Mutation Is Necessary to Disrupt the Dimerization of the Cellular Form of CD69. The soluble CD69 receptor used in the first part of this study utilized a previously described construct CD69NG70 (*18*) that refolded as a noncovalent soluble dimer from a polypeptide consisting of amino acids G70−K199 (Supporting Information Table S1). This construct contained an extended dimer interface involved in contacts between the ligand binding domains, as well as the neck regions. However, it did not contain the C68 residue that has been shown (*8*, *15*) to participate in the covalent dimerization of the natural form of CD69 found at the surface of leukocytes. It thus appeared interesting to look into the effects of mutations of the critical residues C68, Q93, and R134 at the CD69 dimer interface on the structure and on the well-documented signaling functions of the cellular form of CD69. In order to trigger the CD69-mediated activation of transfected Jurkat cells bearing both wild-type and mutated forms of CD69, an efficient ligand is required for receptor cross-linking. Alternatively, the receptor can be aggregated using specific antibodies. In the experiments presented here, we used both forms of activation using two specific cross-linking monoclonal antibodies against CD69, BL-KFB/B1 and BL-Ac/p26, as well as the *N*-acetylhexosamine disaccharide dimer (Figure 4B), which has been previously described (*25*) as the most efficient carbohydrate ligand at enabling precipitation (and thus cross-linking) of the soluble CD69. The structurally closely related ligand, *N*-acetyllactosamine dimer (Figure 4A), served as a suitable negative control in these experiments.

We transferred inserts coding the CD69 receptors into a eukaryotic expression vector for the transfection of the Jurkat T-lymphocyte leukemic cell line (refs *9* and *10* and Table 2). Those clones displaying identical surface expression of the wild-type and mutated receptors (as shown by flow cytometry) were

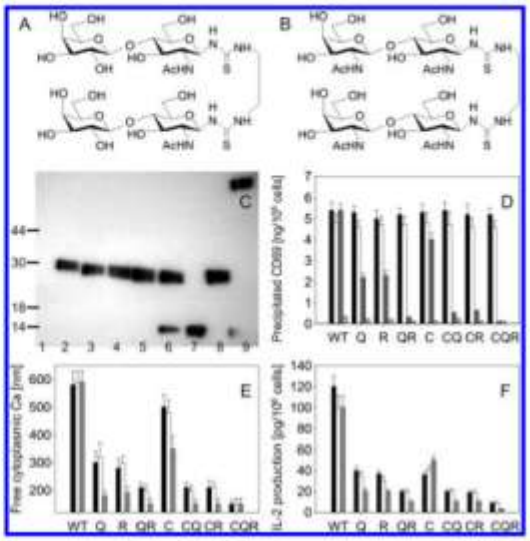


FIGURE 4: Analysis of cellular CD69 on transfected cells and evaluation of binding and signaling efficiency of receptor disturbed at the dimer interface. (A, B) Chemical formulas of *N*-acetyllactosamine and *N*-acetylhexosamine disaccharide dimer, respectively. (C) Surface expression of CD69 immunoprecipitated from control non-transfected cells (lane 1) or from cells containing wild-type CD69 (lane 2), CD69Q93A (lane 3), CD69R134A (lane 4), CD69C68A/Q93A (lane 5), C68A/R134A (lane 6), C68A/Q93A/R134A (lane 7), C68A/Q93A/R134A cross-linked with *N*-acetylhexosamine disaccharide dimer (*8*) (lane 8), and wild-type CD69 cross-linked with *N*-acetylhexosamine disaccharide dimer (lane 9). Receptors were surface-labeled using lactoperoxidase, dimerized using DSS cross-linking, immunoprecipitated, and analyzed by SDS electrophoresis under reducing conditions followed by autoradiography. (D) Precipitation of various forms of wild-type and mutated CD69 using *N*-acetyllactosamine or *N*-acetylhexosamine disaccharide dimer and antibodies against CD69. (E) Cellular activation of Jurkat leukemic T-cells transfected with wild-type or mutated CD69 with mutations at the dimer interface. (F) Production of IL-2 by Jurkat leukemic T-cells transfected with wild-type or mutated CD69 with mutations at the dimer interface. The designation of mutants in (D) to (F) is based on the mutated amino acids and their combinations: C, C68; Q, Q93; R, R134. The cellular CD69 in (D) to (F) was cross-linked using monoclonal antibody BL-KFB/B1 (*2*) (left columns), monoclonal antibody BL-Ac/p26 (*2*) (middle columns), or *N*-acetylhexosamine disaccharide dimer (right columns). The fourth, small column in (D) placed after the three above-mentioned columns indicates precipitation with *N*-acetyllactosamine control.

subcloned, frozen, and used in individual experiments (Figure 5). First, the molecular forms of the surface CD69 were analyzed in both native and ligand cross-linked receptors. In order to determine the precise molecular forms of the receptors in these experiments, native receptors were fixed by di(*N*-hydroxysuccinimido) suberate (DSS), shown previously to produce covalent dimers with soluble CD69 (I. Polakovičová, unpublished observation), before their extraction from the plasma membrane and molecular analysis. Cellular CD69 bearing single mutations in amino acids responsible for the covalent (C68) or noncovalent (Q93, R134) dimerization remained mostly dimeric, as did the double mutants C68A/Q93A, C68A/R134A (Figure 4C, lanes 2−6), and Q93/R134 (not shown). Only the triple mutant C68A/Q93A/R134A was completely monomeric (Figure 4C, lane 7). However, this monomeric protein could still be efficiently dimerized by the *N*-acetylhexosamine disaccharide dimer (lane 8), which caused

Table 2: Characterization of Individual Clones of Jurkat T-Cell Leukemia Transfected with CD69 Isoforms[a]

CD69 isoform	clone	general characteristic	CD69 expression by FACS[b]	used for expts
CD69WT	WT1	good growth	10	no
	WT2	**moderate growth**	**30**	**yes**
	WT3	moderate growth	80	no
CD69Q93A	Q1	good growth	20	no
	Q2	**good growth**	**30**	**yes**
	Q3	moderate growth	70	no
CD69R134A	R1	moderate growth	30	yes
	R2	**good growth**	**90**	**no**
CD69QRDM	RQ1	moderate growth	10	no
	RQ2	**good growth**	**30**	**yes**
	RQ3	good growth	70	no
CD69C68A	**C1**	**good growth**	**30**	**yes**
	C2	good growth	60	no
CD69CQDM	CQ1	good growth	10	no
	CQ2	**moderate growth**	**30**	**yes**
	CQ3	moderate growth	80	no
CD69CRDM	CR1	moderate growth	10	no
	CR2	**moderate growth**	**30**	**yes**
	CR3	slow growth	90	no
CD69CQRTM	CQR1	moderate growth	10	no
	CQR2	good growth	20	no
	CQR3	**moderate growth**	**30**	**yes**
	CQR4	slow growth	60	no

[a]Only the successful transformants with relative fluorescence intensity of CD69 expression between 10 and 100 have been included into this table. [b]Median of relative fluorescence intensity from FACS analyses at the top of the peak.

extensive cross-linking and formation of high molecular weight aggregates in wild-type cellular CD69 (lane 9).

Mutations in Q93 and R134 of Cellular CD69 Result in More Significant Impairment in Its Signaling than Mutation in C68. We further investigated the effect of mutations destabilizing the dimerization of cellular CD69 on the cross-linking of this receptor and its function in cellular activation. While all of the expressed cellular CD69 could be efficiently cross-linked by two monoclonal antibodies against this antigen, the cross-linking by the dimerized ligand was severely impaired in the mutants affecting receptor dimerization even if the dimerization process *per se* was not affected (cf. Figure 4C,D). Interestingly, mutations in the amino acids forming the noncovalent dimer interface had in several instances a more profound effect than the mutation in C68 responsible for covalent dimerization. On the other hand, no precipitation of the receptor occurred when the *N*-acetyllactosamine dimer control compound was used (Figure 4D).

Subsequently, the influence of the mutations affecting the dimerization of cellular CD69 on the ability of this receptor to activate the Jurkat cell line, as documented by the increase in the intracellular calcium levels, was analyzed. Strikingly, even if the cellular CD69 antigens bearing the above mutations could be fully cross-linked by binding to monoclonal antibodies, their ability to confer activation was severely affected (Figure 4E). For instance, single mutations in amino acids at the noncovalent dimer interface (i.e., Q93A or R134A mutations) lowered the efficiency of these receptors in cellular signaling by two-thirds. On the other hand, when using ligand cross-linking with the *N*-acetylhexosamine disaccharide dimer, a very low efficiency of cellular signaling was observed compared to the wild-type

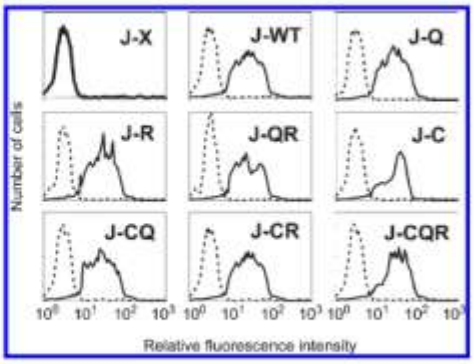


FIGURE 5: Analysis of surface expression of cellular CD69 in transfected Jurkat cell clones as determined by flow cytometry: J-X, insert-free (mock DNA) transfectant; J-WT, cells transfected with wild-type CD69; J-Q, cells transfected with CD69Q93A mutant; J-R, cells transfected with CD69R134A mutant; J-QR, cells transfected with CD69Q93A/R134A double mutant; J-C, cells transfected with CD69C68A mutant; J-CQ, cells transfected with CD69C68A/Q93A double mutant; J-CR, cells transfected with CD69C68A/R134A double mutant; J-CQR, cells transfected with CD69C68A/Q93A/R134A triple mutant. Reactivity with the monoclonal antibody against CD69 (BL-KFB/B1) is shown by the solid line; reactivity with the isotype-matched control antibody against NKR-P1 (HD14) is shown by the dotted line.

controls (Figure 4E). This may be explained by a combination of low efficiency of ligand binding and receptor cross-linking together with a direct effect on signaling efficiency. Similar results were obtained with the production of IL-2 as another measure of cellular activation (Figure 4F).

CONCLUSIONS

The results presented here indicate that the binding of carbohydrate ligands to both the soluble and cellular forms of CD69 is cooperative and is affected by the interaction of the two subunits of this receptor. This is a hitherto unreported observation, since the signaling through CD69 has been so far ascribed exclusively to its association with the signaling adapter proteins in the transmembrane region of the molecule (*8*). Thus, while the previous studies have defined the role of the individual polypeptide segments of CD69 in cellular expression, surface dimerization, and cellular signaling, the present results are unique in emphasizing the role of the intact dimer interface in the receptor molecule that was found to be much larger than the mere cysteine 68 thought previously (*8*, *15*) to be critical for receptor dimerization. Moreover, our findings suggest a novel mechanism for sensitive ligand recognition by CD69 and related immune receptors. Within the C-type lectin family, various alternative strategies have been employed to attain high-affinity binding. Classical hepatic lectins are oligomeric proteins in which the individual carbohydrate-recognition domains within the oligomer cooperate during the recognition of desialylated glycoproteins (*32*). Other members of this protein family employ alternative calcium-dependent processes to achieve this goal (*33*). Among the receptors of the immune system, the soluble mannose-binding protein again employs a multiple carbohydrate-recognition domain to bind specifically to surface polysaccharides of various pathogens and activate their effector functions such as opsonization or complement activation (*34*, *35*). Compared to these lectins, lymphocyte receptors forming group V of

the C-type lectin family are much smaller and mostly dimeric, and thus they seem to have developed alternative strategies for their recognition of specific ligands. These are based either on oligomerization of the receptor within the specialized plasma membrane microdomains of the immune cell (*36*) or on alternative strategies that may be used by some of these receptors (*37*). The molecular mechanism that we propose based on the results of the current work is unique in rapid propagation of the activation signal using the mechanism of rapid cooperative receptor cross-linking and oligomerization based on the positive cooperativity in binding of the multivalent ligands.

ACKNOWLEDGMENT

We thank Angela Risso for assistance and helpful discussions, Iva Polakovičková for help with DSS cross-linking experiments, and Michal Navrátil for help in the development of cellular receptor precipitation assays.

SUPPORTING INFORMATION AVAILABLE

A description of additional experimental procedures with supporting references, binding of calcium to the fourth generation soluble CD69 proteins (Figure S1), carbohydrate inhibition experiments for these proteins (Figure S2), primary data for NMR titrations with ManNAc and GlcNAc (Figure S3), elution profiles for gel filtration of soluble dimeric CD69 (CD69NG70) in the absence of ligand as well as in the presence of 1 mM ManNAc and 1 mM GlcNAc (Figure S4), sedimentation velocity experiments with CD69NG70 in the absence of ligand and in the presence of 1 mM ManNAc and 1 mM GlcNAc (Figure S5), identification of amino acids for the design of the monomeric CD69 mutant (Figure S6), and characterization of CD69 expression constructs (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Testi, R., D'Ambrosio, D., De Maria, R., and Santoni, A. (1994) The CD69 receptor: a multipurpose cell surface trigger for hematopoietic cells. *Immunol. Today 15*, 479–483.
2. Sancho, D., Gomez, M., and Sánchez-Madrid, F. (2005) CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol. 26*, 136–140.
3. Moretta, A., Poggi, A., Pende, D., Tripodi, G., Orengo, A. M., Pella, N., Augugliaro, R., Bortino, C., Ciccone, E., and Moretta, L. (1991) CD69-mediated pathway of lymphocyte activation: anti-CD69 monoclonal antibodies trigger the cytolytic activity of different lymphoid effector cells with the exception of cytolytic T lymphocytes expressing T cell receptor α/β. *J. Exp. Med. 174*, 1393–1398.
4. Borrego, F., Robertson, M. J., Ritz, J., Pena, J., and Solana, R. (1999) CD69 is a stimulatory receptor for natural killer cells and its cytotoxic effect is blocked by CD94 inhibitory receptor. *Immunology 97*, 159–165.
5. Lopez-Cabrera, M., Santis, A. G., Fernandez-Ruiz, E., Blacher, R., Esch, F., Sanchez-Mateos, P., and Sánchez-Madrid, F. (1993) Molecular cloning, expression, and chromosomal localization of the human earliest activation antigen AIM/CD69, a new member of C-type lectin superfamily of signal-transmitting receptors. *J. Exp. Med. 178*, 537–547.
6. Lanier, L. L. (2008) Up on the tightrope: natural killer cell activation and inhibition. *Nat. Immunol. 9*, 495–502.
7. Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. (2008) Functions of natural killer cells. *Nat. Immunol. 9*, 503–510.
8. Sancho, D., Santis, A. G., Alonso-Lebrero, J. L., Viedma, F., Tejedor, R., and Sánchez-Madrid, F. (2000) Functional analysis of ligand-binding and signal transduction domains of CD69 and CD23 C-type lectin leukocyte receptors. *J. Immunol. 165*, 3868–3875.
9. Pisegna, S., Zignoni, A., Pirozzi, G., Cinque, B., Cifoni, M. G., Morrone, S., Picolli, M., Frati, L., Palmieri, G., and Santoni, A. (2002) Src-dependent Syk activation controls CD69 mediated signaling and function of human NK cells. *J. Immunol. 169*, 68–74.
10. Zingoni, A., Palmieri, G., Morrone, S., Carretero, M., Lopez-Botet, M., Picolli, M., Frati, L., and Santoni, A. (2000) CD69-triggered ERK activation and functions are negatively regulated by CD94/NKG2A inhibitory receptors. *Eur. J. Immunol. 30*, 644–651.
11. Risso, A., Smilowich, D., Capra, M. C., Baldissarro, I., Yan, G., Bargilessi, A., and Cosulich, M. E. (1991) CD69 in resting and activated T lymphocytes. Its association with a GTP binding protein and biochemical requirements for its expression. *J. Immunol. 146*, 4105–4114.
12. Bikah, G., Pogue-Caley, R. R., and McHeyzer-Williams, M. G. (2000) Regulating T helper cell immunity through antigen responsiveness and calcium entry. *Nat. Immunol. 1*, 402–412.
13. Shiow, R. L., Rosen, D. B., Brdečková, N., Xu, Y., An, J., Langer, L. L., Cyster, J. G., and Matloubian, M. (2006) CD69 acts downstream of interferon-α/β to inhibit $S1P_1$ and lymphocyte egress from lymphoid organs. *Nature 440*, 540–544.
14. Marzio, R., Mauel, J., and Betz-Corradin, S. (1999) CD69 and regulation of the immune functions. *Immunopharmacol. Immunotoxicol. 21*, 565–582.
15. Bezouška, K., Nepovím, A., Horváth, O., Pospíšil, M., Hamann, J., and Feizi, T. (1995) CD69 antigen of human leukocyte is a calcium-dependent carbohydrate-binding protein. *Biochem. Biophys. Res. Commun. 208*, 68–74.
16. Childs, R. A., Galustian, C., Lawson, A. M., Doufán, G., Benwell, K., Frankel, G., and Feizi, T. (2000) Recombinant soluble human CD69 dimer produced in *Escherichia coli*: reevaluation of saccharide binding. *Biochem. Biophys. Res. Commun. 266*, 19–23.
17. Pavlíček, J., Sopko, B., Ettrich, R., Kopecký, V., Baumruk, V., Man, P., Havlíček, V., Vrbacký, M., Martinková, L., Křen, V., Pospíšil, M., and Bezouška, K. (2003) Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry 42*, 9295–9306.
18. Vaněk, O., Nálezková, M., Kavan, D., Borovičková, I., Pompach, P., Novák, P., Kumar, V., Vannucci, L., Hudeček, J., Hofbauerová, K., Kopecký, V., Brynda, J., Kolenko, P., Dohnálek, J., Kadeřávek, P., Chmelík, J., Gorčík, L., Žídek, L., Sklenář, V., and Bezouška, K. (2008) Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice. *FEBS J. 275*, 5589–5606.
19. Kolenko, P., Skálová, T., Vaněk, O., Štěpánková, A., Dušková, J., Hašek, J., Bezouška, K., and Dohnálek, J. (2009) The high-resolution structure of the extracellular domain of human CD69 using a novel polymer. *Acta Crystallogr., F65*, 1258–1260.
20. Pavlíček, J., Kavan, D., Pompach, P., Novák, P., Lukšan, O., and Bezouška, K. (2004) Lymphocyte activation receptors: new structural paradigma in group V of C-type animal lectins. *Biochem. Soc. Trans. 32*, 1124–1126.
21. Bezouška, K., Šnajdrová, R., Křenek, K., Vančurová, M., Kádek, A., Adámek, D., Lhoták, P., Kavan, D., Hofbauerová, K., Man, P., Bojarová, P., and Křen, V. (2010) Carboxylated calixarenes bind strongly to CD69 and protect $CD69^+$ killer cells from suicidal cell death induced by tumor cell surface ligands. *Bioorg. Med. Chem. 18*, 1434–1440.
22. Křenek, K., Kuldová, M., Hulíková, K., Stibor, I., Lhoták, P., Dudič, M., Budka, J., Pelantová, H., Bezouška, K., Fišerová, A., and Křen, V. (2007) *N*-acetyl-D-glucosamine substituted calix[4]arenes as stimulators of NK cell-mediated antitumor immune response. *Carbohydr. Res. 342*, 1781–1792.
23. Hulíková, K., Benson, V., Svoboda, J., Šima, P., and Fišerová, A. (2009) *N*-Acetyl-D-glucosamine-coated polyamidoamine dendrimer modulates antibody formation via natural killer cell activation. *Int. Immunopharmacol. 9*, 792–799.
24. Benson, V., Grohárová, V., Richter, J., and Fišerová, A. (2010) Glycosylation regulates NK cell-mediated effector function through PI3K pathway. *Int. Immunol. 22*, 167–177.
25. Bojarová, P., Křenek, K., Wetjen, K., Adamiak, K., Pelantová, H., Bezouška, K., Elling, L., and Křen, V. (2009) Synthesis of LacdiNAc-terminated glycoconjugates by mutant galactosyltransferase—A way to new glycodrugs and materials. *Glycobiology 19*, 509–517.
26. Schuck, P. (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modelling. *Biophys. J. 78*, 1606–1619.
27. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem. 320*, 104–124.
28. Dousseau, F., Therrien, M., and Pézolet, M. (1989) On the spectral subtraction of water from the FT-IR spectra of aqueous solution of proteins. *Appl. Spectrosc. 43*, 538–542.

29. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on an UNIX pipes. *J. Biomol. NMR 6*, 277–293.
30. Chipman, D. M., Grisaro, V., and Sharon, N. (1967) The binding of oligosaccharides containing *N*-acetylglucosamine and *N*-acetylmuramic acid to lysozyme. *J. Biol. Chem. 242*, 4388–4394.
31. Hamann, J., Fiebig, H., and Strauss, M. (1993) Expression cloning of the early activation antigen CD69, a type II integral membrane protein with a C-type lectin domain. *J. Immunol. 150*, 4920–4927.
32. Blomhoff, R., Tolleshaug, H., and Berg, T. (1982) Binding of calcium ions to the isolated asialoglycoprotein receptor. Implication for receptor function in suspended hepatocytes. *J. Biol. Chem. 257*, 7456–7459.
33. Kimura, T., Imai, Y., and Irimura, T. (1995) Calcium-dependent conformation of a mouse macrophage calcium-type lectin. Carbohydrate binding activity is stabilized by an antibody specific for a calcium-dependent epitope. *J. Biol. Chem. 270*, 16056–16062.
34. Weis, W. I., and Drickamer, K. (1994) Trimeric structure of a C-type mannose-binding protein. *Structure 15*, 1227–1240.
35. Wallis, R., and Drickamer, K. (1999) Molecular determinant sof oligomer formation and complement fixation in mannose-binding proteins. *J. Biol. Chem. 274*, 3580–3589.
36. Diefenbach, A., and Raulet, D. H. (2001) Strategies for target cell recognition by natural killer cells. *Immunol. Rev. 181*, 170–184.
37. Coombs, P. J., Harrison, R., Pemberton, S., Quintero-Martinez, A., Parry, S., Haslam, S. M., Dell, A., Tailor, M. E., and Drickamer, K. (2010) Identification of novel contributions to high-affinity glycoprotein-receptor interactions using engineered ligands. *J. Mol. Biol. 396*, 685–696.

Appendix 5

Kovalová A., Ledvina M., Saman D., Zyka D., Kubícková M., Zídek L., Sklenár V., Pompach P., Kavan D., Bílý J., Vanek O., Kubínková Z., Libigerová M., Ivanová L., Antolíková M., Mrázek H., Rozbeský D., Hofbauerová K., Kren V., Bezouska K.

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 (2010).

4050 J. Med. Chem. 2010, 53, 4050–4065
DOI: 10.1021/jm100055b

Journal of Medicinal Chemistry Article

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

Anna Kovalová,[†] Miroslav Ledvina,[†] David Šaman,[†] Daniel Zyka,[†] Monika Kubičková,[‡] Lukáš Žídek,[‡] Vladimír Sklenář,[‡] Petr Pompach,[§,∥] Daniel Kavan,[§,∥] Jan Bílý,[§] Ondřej Vaněk,[§,∥] Zuzana Kubínková,[§,∥] Martina Libigerová,[§,∥] Ljubina Ivanová,[§,∥] Mária Antoliková,[§,∥] Hynek Mrázek,[§,∥] Daniel Rozbeský,[§,∥] Kateřina Hofbauerová,[⊥] Vladimír Křen,[⊥] and Karel Bezouška*[,§,∥]

[†]*Institute of Organic Chemistry and Biochemistry, V.V.I., Academy of Sciences of the Czech Republic, Flemingovo Nám. 2, 16610 Prague, Czech Republic,* [‡]*National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 2, 61137 Brno, Czech Republic,* [§]*Department of Biochemistry, Faculty of Science, Charles University Prague, Hlavova 8, 12840 Prague, Czech Republic, and* [∥]*Institute of Microbiology, V.V.I., Academy of Sciences of the Czech Republic, Vídeňská 1083, 14220 Prague, Czech Republic*

Received January 14, 2010

On the basis of the highly branched ovomucoid-type undecasaccharide that had been shown previously to be an endogenous ligand for CD69 leukocyte receptor, a systematic investigation of smaller oligosaccharide mimetics was performed based on linear and branched *N*-acetyl-D-hexosamine homooligomers prepared synthetically using hitherto unexplored reaction schemes. The systematic structure−activity studies revealed the tetrasaccharide GlcNAcβ1−3(GlcNAcβ1−4)(GlcNAcβ1−6)GlcNAc (compound **52**) and its α-benzyl derivative **49** as the best ligand for CD69 with IC_{50} as high as 10^{-9} M. This compound thus approaches the affinity of the classical high-affinity neoglycoprotein ligand $GlcNAc_{23}BSA$. Compound **68**, GlcNAc tetrasaccharide **52** dimerized through a hydrophilic flexible linker, turned out to be effective in activating CD69$^+$ lymphocytes. It also proved efficient in enhancing natural killing in vitro, decreasing the growth of tumors in vivo, and activating the CD69$^+$ tumor infiltrating lymphocytes examined ex vivo. This compound is thus a candidate for carbohydrate-based immunomodulators with promising antitumor potential.

Introduction

In view of the recent increasing incidence of pathological states characterized by secondary immune deficiencies in the general population, considerable attention has been given to the investigation of effective ways to modulate the activities of the individual components of the immune system. Natural killer (NK[a]) cells, which form 5−10% of mononuclear cells in the blood, have many unique immunological activities and thus provide an obvious target for immune activation. NK cells play a crucial role in innate immunity, as they are characterized by fast and strong cytolytic response against tumor or virally infected cells. They also have the ability to release cytokines and chemokines mediating inflammatory responses and to influence hematopoiesis and the adaptive immune response.[1−3] Despite their role in immune defense mechanisms, major questions regarding their therapeutical potential remained unanswered.[4,5] The ability of NK cells to discriminate between normal and tumor or virally infected cells is now much better understood because of the identification of various surface NK receptors contributing to the process of NK cell activation or inactivation.[4−6] The activation and triggering of natural killing are under the control of a complex signaling machinery to which many receptors, components of the cell surface "receptor zipper", are known to contribute.[7] In this study, we focused our attention on a group of calcium-dependent animal lectins with binding affinity for carbohydrate structures. The complex saccharide structures are involved in many biologically important signal transduction processes, and thus, they play a key role in molecular recognition events contributing to cell−cell, cell−bacteria, and cell−virus interactions.[8−10] The lectin receptors are able to recognize oligosaccharide structures present on the surface of tumor cells and initiate their lysis by cells of the immune system.[11,12] We are interested in two NK cell lectin activation receptors, rat NKR-P1 and human CD69, unique for their ability to distinguish between closely related carbohydrate structures and to recognize the *N*-acetyl-D-hexosamines (HexNAc) in both gluco and galacto configurations.[13,14] Carbohydrates interact with these lectins over an extensive surface area, but the structure and position of the oligosaccharide binding sites are unique for each of the two receptors. Rat NKR-P1 has a binding groove that accommodates the linear oligosaccharides,[15] whereas sugar-binding sites in human CD69 are at three separate locations, and thus branched carbohydrates seem to be preferred.[16,17] Our previous systematic study of the activating lectin receptor NKR-P1

*To whom correspondence should be addressed. Phone: +420-2-4106-2383. Fax: +420-2-2195-1283. E-mail: bezouska@biomed.cas.cz.

[a]Abbreviations: 3N, GlcNAcβ1−3GlcNAc; 6N, GlcNAcβ1−6GlcNAc; A, GalNAc; BN, GlcNAcβ1−3(GlcNAcβ1−4)GlcNAc; CB, GlcNAcβ1−4GlcNAc; CT, GlcNAcβ1−4GlcNAcβ1−4GlcNAc; HexNAc, *N*-acetyl-D-hexosamine; ION, ionomycin; LAC, lactose dimer; mAb, monoclonal antibody; N, GlcNAc; NG, $GlcNAc_{23}BSA$ neoglycoprotein; NK, natural killer; OM, [GlcNAcβ1−2(GlcNAcβ1−4)(GlcNAcβ1−6)Manα1−6]-[GlcNAcβ1−2(GlcNAcβ1−4)Manα1]3[GlcNAcβ1−4]-Manβ1−4GlcNAcβ1−4GlcNAc.

 Published on Web 04/30/2010

Chart 1. Structures of the Synthesized HexNAc Based Oligosaccharides Used in the Study

demonstrated the binding hierarchy of HexNAc type oligosaccharides. The binding affinity increases in the group of monosaccharides from GlcNAc and GalNAc to ManNAc and in the order of chitooligomers ([-β-D-GlcNAc(1→4)]$_n$) with elongation of the oligosaccharide chain up to four sugar units. In linear oligosaccharides of the *N*-acetylglucosamine type, the β(1→4) glycosidic bond is preferred over β(1→6)- and β(1→3)-linked regioisomers of chitobiose ([-β-D-Glc*p*NAc(1→4)]$_2$) showing lower binding affinity.[18,19]

In the case of the human CD69 receptor, the physiological ligands are not yet known. Using recombinant CD69 protein, we have previously identified three separate binding sites for GlcNAc in the monomeric unit of this receptor.[20] Moreover, we have shown that the complex pentaantennary bisecting undecasaccharide from egg white glycoprotein ovomucoid with the structure [GlcNAcβ1−2(GlcNAcβ1−4)(GlcNAcβ1−6)-Manα1−6][GlcNAcβ1−2(GlcNAcβ1−4)Manα13][GlcNAcβ1−4]Manβ1−4GlcNAcβ1−4GlcNAc (OM, Chart 1), is one of the best ligands of natural origin identified so far.[17] The affinity of this natural oligosaccharide for CD69 is in the low nanomolar range, thus competing successfully with the artificial high affinity ligand, GlcNAcBSA neoglycoprotein (NG, Chart 1). This natural structure has thus provided an important paradigm for the development of smaller synthetic oligosaccharide mimetics that would be accessible for large scale synthesis and further potential use for in vivo experimental tumor therapies. The total chemical synthesis of complex-type N-linked oligosaccharides identical to natural ones has been achieved,[21] but this involves a number of chemical protection/deprotection steps and remains a time-consuming and costly task. Additional approaches have been developed to supply the glycobiology community with a sufficient amount of pure complex oligosaccharides, but these procedures have not yet been sufficiently adapted for robust and inexpensive synthesis.[9,18] Yet another possibility is represented by de novo chemical syntheses of branched homooligosaccharides, which might afford new oligosaccharides rarely occurring in nature in pure form and in sufficient amounts.[22,23] Recently, we synthesized the branched homotrisaccharide β-D-GlcNAc-(1→3)[β-D-GlcNAc-(1→4)]-D-GlcNAc and found that its binding affinity for NK cell lectin receptors competed successfully with oligosaccharides of much greater complexity.[19]

Here we aimed to identify oligosaccharide ligands useful for the modulation of activities of NK cells through their surface receptors, NKR-P1 and CD69.[11,17] We focused on assessing the binding affinity of NKR-P1 receptor for linear homooligosaccharides composed of β-linked HexNAc with several types of glycosidic bonds. In the case of CD69, we tried to select a small size oligosaccharide composed of highly branched HexNAc that would mimic the natural oligosaccharide from ovomucoid described above.[17] For this purpose we synthesized and biologically tested a group of model oligosaccharides: (a) linear homooligosaccharides of *N*-acetylgalactosamine type, i.e., [-β-D-GalNAc(1→4)]$_2$ (**21**), [-β-D-GalNAc(1→4)]$_3$ (**32**), and [-β-D-GalNAc(1→3)]$_2$ (**37**); (b)

branched homooligosaccharides of *N*-acetyl-D-glucosamine type, i.e., triantennary tetrasaccharide β-D-GlcNAc-(1→3)-[β-D-GlcNAc-(1→6)]-[β-D-GlcNAc-(1→4)]-D-GlcNAc (**52**) and biantennary trisaccharides β-D-GlcNAc-(1→3)-[β-D-GlcNAc-(1→6)]-D-GlcNAc (**53**) and β-D-GlcNAc-(1→4)-[β-D-GlcNAc-(1→6)]-D-GlcNAc (**54**); and (c) biantennary homooligosaccharide of *N*-acetyl-D-galactosamine type, β-D-GalNAc-(1→3)-[β-D-GalNAc(1→4)]-D-GalNAc (**61**). Dimerization of oligosaccharide **52**, which proved to be one of the best ligands for human CD69 using well established chemistry used in peptide and protein chemistry,[24] provided a series of compounds of which **68** proved to be an efficient activators of the effector cells of the immune system when tested both in vitro and in vivo. Thus, the ability of the dimerized oligosaccharide **68** to efficiently activate cells of the immune system parallels that found previously for the dimerized peptide ligand for CD69.[24]

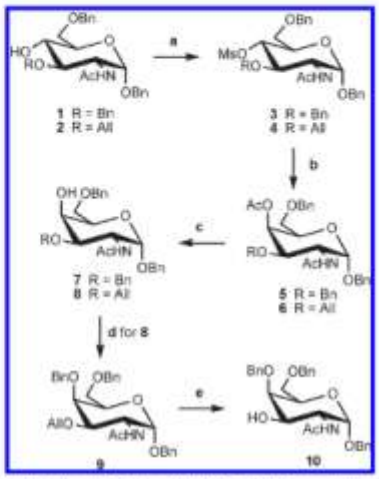
Results

We prepared three series of carbohydrate ligands for NK cell receptors NKR-P1 and CD69 (Chart 1). The first series consisted of linear GalNAc based oligosaccharides **37**, **21**, and **32** used together with commercially available (chitobiose, CB; chitotriose, CT) and previously described (3N, ref 19) oligosaccharides in the GlcNAc (N) format. The second series contained biantennary GalNAc based oligosaccharide **61** intended to provide an equivalent to the corresponding GlcNAc based oligosaccharide BN described previously.[19] The third, most important, series was prepared (based on the results of biological tests with two previous series) only in the GlcNAc (N) format and included the biantennary oligosaccharides **53** and **54** constituting (together with BN) the complete series of trisaccharide GlcNAc (N) isomers, as well as the fully branched triantennary tetrasaccharide **52**.

Chemistry. In the synthesis of the target oligosaccharides, we were often confronted with sterically hindered systems, especially in the case of highly branched structures and oligosaccharide structures consisting of *N*-acetylgalactosamine units. In general, the axially oriented C(4)−OH group on a galactopyranose skeleton is the least reactive secondary OH group.[25] To overcome this fact, the phthalimide glycosylation method was applied. This procedure is considered to be one of the most efficient 1,2-trans-stereoselective glycosylation processes for low-reactive secondary OH groups.[26] Nevertheless, we recently observed reduced stereoselectivity of this method when very low reactive aglycons were used.[27]

The parallel approach enabling simultaneous glycosylation of several hydroxyl groups of the glycosyl acceptor was employed for the preparation of branched oligosaccharides.[19] In addition, a silver perchlorate promoted glycosylation procedure was applied using glycosyl bromides as glycosyl donors in the presence of silver carbonate. This was selected because of relatively good efficiency and stereoselectivity in the case of low reactive, axially oriented OH(4) groups in comparison to other standard glycosylation methods.[27] Silver carbonate acts as a scavenger of perchloric acid and was used instead of a commonly used organic base, which can inhibit glycosylation reaction.[28,29] Appropriate glycosyl donors and glycosyl acceptors of galacto configuration were obtained via inversion of the configuration at the C(4) carbon of the corresponding synthons with gluco configuration by nucleophilic displacement of the mesyl group by sodium acetate.[27]

Scheme 1[a]



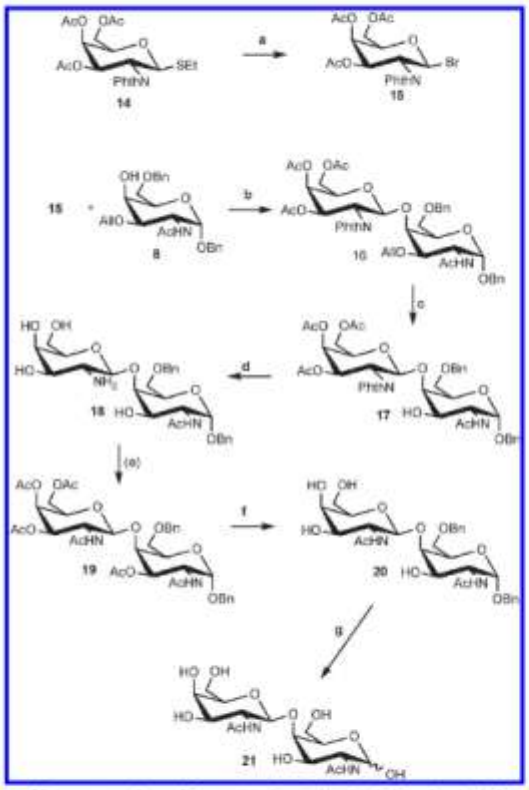
[a] Reagents and conditions: (a) MsCl, pyridine, room temp; (b) NaOAc, DMSO, 130 °C; (c) MeONa, MeOH, room temp; (d) BnBr, NaH, DMF, room temp; (e) $(Ph_3P)_3RhCl$, toluene−EtOH−H_2O, reflux, and then HCOOH, reflux.

Synthesis of Linear Oligosaccharides of the D-Galactosamine Type. In contrast to the large number of studies devoted to the synthesis of linear oligosaccharides with β(1→4)-linked 2-amino-2-deoxy-D-glucopyranose units, efficient and practical methods for the synthesis of analogous oligosaccharides with β(1→4)-linked 2-amino-2-deoxy-D-galactopyranose units remain scarce.[26] This unsatisfactory situation was apparently due to problems with formation of a β(1→4) glycosidic bond between two D-galactosamine units. Only few reports on disaccharide syntheses of this type have been published so far, and these used a glycosyl acceptor having the azido group as the masked amino function at the C(2) position and 2-deoxy-2-phthalimido-D-galactopyranosyl bromide as the glycosyl donor.[30,31]

Our primary goal in the synthesis of the target β(1→4)- and β(1→3)-linked linear oligosaccharides of the D-galactosamine type (**21**, **32**, and **37**) was to prepare appropriate glycosyl acceptors **7**, **8**, **10** (Scheme 1), and **13**. Compound **7** was obtained from benzyl 2-acetamido-3,6-di-*O*-benzyl-2-deoxy-4-*O*-methanesulfonyl-α-D-glucopyranoside[32] (**3**) upon treatment with anhydrous sodium acetate in dimethyl sulfoxide at 130 °C to give the galacto-derivative **5**, which afforded **7** by Zemplén deacetylation. The same synthetic approach was applied for the preparation of compound **8** starting from benzyl 2-acetamido-3-*O*-allyl-6-*O*-benzyl-2-deoxy-α-D-glucopyranoside[28] (**2**). O-Benzylation of **8** and subsequent deallylation of compound **9** so obtained by catalytic isomerization of the protecting allyl group to prop-1-enyl group using Wilkinson's catalyst $(Ph_3P)_3RhCl$, followed by acid hydrolysis, led to glycosyl acceptor **10**. Benzyl 3,6-di-*O*-benzyl-2-deoxy-2-phthalimido-β-D-galactopyranoside (**13**) was prepared from ethyl 4-*O*-acetyl-3,6-di-*O*-benzyl-2-deoxy-2-phthalimido-1-thio-β-D-galactopyranoside[27] (**11**) in two steps using a procedure described by Westerlind et al.[33]

The use of 3,6-di-*O*-benzylderivative **7** as a starting glycosyl acceptor in preparation of the target β(1→4)-linked oligosaccharides **21** and **32** was not successful. Glycosylation

Scheme 2[a]



[a] Reagents and conditions: (a) Br_2, CH_2Cl_2, 0 °C; (b) $AgClO_4$, Ag_2CO_3, CH_2Cl_2, −15 °C; (c) $(Ph_3P)_3RhCl$, toluene−EtOH−H_2O, reflux, and then HCOOH, reflux; (d) $BuNH_2$, MeOH, reflux; (e) Ac_2O, pyridine, room temp; (f) MeONa, MeOH, room temp; (g) H_2, Pd/C, AcOH−H_2O, room temp.

of **7** with 4-*O*-acetyl-3,6-di-*O*-benzyl-2-deoxy-2-phthalimido-D-galactopyranosyl bromide[27] (**22**) promoted by silver perchlorate in the presence of silver carbonate gave a complex mixture of reaction products. Any attempts to isolate the expected disaccharide benzyl 4-*O*-acetyl-3,6-di-*O*-benzyl-2-deoxy-2-phthalimido-β-D-galactopyranosyl-(1→4)-2-acetamido-3,6-di-*O*-benzyl-2-deoxy-α-D-galactopyranoside failed.

The disaccharide **21** was successfully prepared when the less sterically hindered 3-*O*-allyl derivative **8**, as glycosyl acceptor, and glycosyl bromide **15** protected with less bulky and more electronegative *O*-acetyl groups as glycosyl donor were used at the glycosylation step (Scheme 2). Their reaction under the same conditions afforded the required β(1→4)-linked disaccharide **16** in a yield of 37%. The glycosyl donor **15** described by Nilsson et al.[34] was prepared by an alternative way, i.e., from ethyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-1-thio-β-D-galactopyranoside[35] (**14**) in reaction with bromine.

Deallylation of the protected disaccharide **16** using the aforementioned procedure gave compound **17**. The alkali-labile protecting groups were removed by treatment with *n*-butylamine in boiling methanol, and the obtained amine **18** was peracetylated by reaction with acetic anhydride in pyridine to give **19**. Zemplén O-deacetylation of **19** yielded benzyl glycoside **20**. Hydrogenolysis of benzyl groups over Pd/C catalyst afforded the target unprotected β(1→4)-linked disaccharide **21**.

The problem of effectively synthesizing β(1→4)-linked galactosamine type oligosaccharides was solved by employing compound **13** as the starting glycosyl acceptor (Scheme 3). Compound **13**, bearing a phthalimido group at position C(2) instead of an acetamido group, was chosen to decrease the negative influence of intermolecular H-bonds between the OH group and the NH group of the amide function in the glycosylation process.[36−38] Glycosylation of **13** with glycosyl bromide **22** under the above-described conditions gave the required disaccharide **23** in a satisfactory yield of 48%. Furthermore, 4-*O*-acetyl-1,5-anhydro-3,6-di-*O*-benzyl-2-deoxy-2-phthalimido-D-lyxo-hex-1-enitol (**24**) and 4-*O*-acetyl-3,6-di-*O*-benzyl-2-deoxy-2-phthalimido-β-D-galactopyranosyl-4-*O*-acetyl-3,6-di-*O*-benzyl-2-deoxy-2-phthalimi-do-β-D-galactopyranoside (**25**) were isolated as the reaction side products, i.e., the products of elimination and cross-reaction of a glycosyl donor.[39,40] Formation of glycal and symmetrical disaccharide with the β head to head glycosidic bond similar to that found in β,β-trehalose from glycosyl donors of 2-deoxy-2-phthalimido-D-glucopyranose type has been described.[39,40]

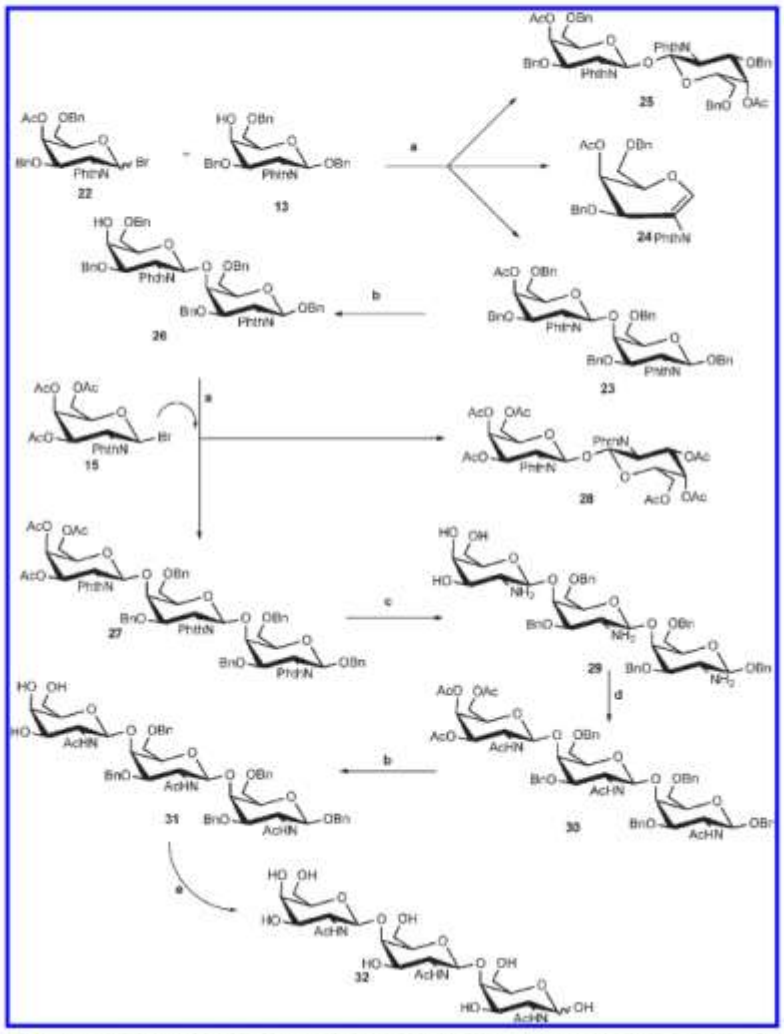
Zemplén deacetylation of disaccharide **23** followed by glycosylation of the obtained compound **26** with glycosyl donor **15** afforded trisaccharide **27** in an overall yield of 51%. Furthermore, the β,β-trehalose type product **28** was also isolated.[19] The alkali-labile protecting groups were removed by treatment with hydrazine hydrate in boiling ethanol, and the obtained crude amine **29** was acetylated by acetic anhydride in pyridine to give compound **30**. Zemplén deacetylation of **30** gave benzyl glycoside **31**, and its benzyl groups were subsequently hydrogenolysed over Pd/C catalyst to give the target unprotected trisaccharide **32**.

For the preparation of β(1→3)-linked galactosamine disaccharide **37** (Scheme 4) a synthetic approach analogous to the one used for the synthesis of linear oligosaccharides containing β(1→4)-linked *N*-acetyl-D-galactosamine was employed. Silver perchlorate in the presence of silver carbonate promoted glycosylation of glycosyl acceptor **10** with galactopyranosyl bromide **15** to provide β(1→3)-linked disaccharide **33** in a good yield of 70%. This was then deprotected to obtain the final disaccharide **37**.

Synthesis of Branched Oligosaccharides of the D-Glucosamine Type. Synthesis of the target triantennary tetrasaccharide **52** and biantennary trisaccharides **53** and **54** of D-glucosamine type was carried out starting from multiple glycosylation of the glycosyl acceptor benzyl 2-acetamido-2-deoxy-α-D-glucopyranoside (**39**) (Scheme 5). Glycosylation of **39** with 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl bromide[28] (**38**), promoted with silver perchlorate in the presence of silver carbonate, afforded the tetrasaccharide **40** (15%) and deletion trisaccharides **41** (8%) and **42** (17%). The obtained oligosaccharides were converted to unprotected oligosaccharides **52**, **53**, and **54** as described above.

Synthesis of Branched Trisaccharide of the D-Galactosamine Type. Similar to the synthesis of branched oligosaccharides containing D-glucosamine units, the approach based on multiple glycosylation was applied as well for D-galactosamine type oligosaccharides. The glycosyl acceptor benzyl 2-acetamido-6-*O*-benzyl-2-deoxy-α-D-galactopyranoside (**55**),

Scheme 3[a]



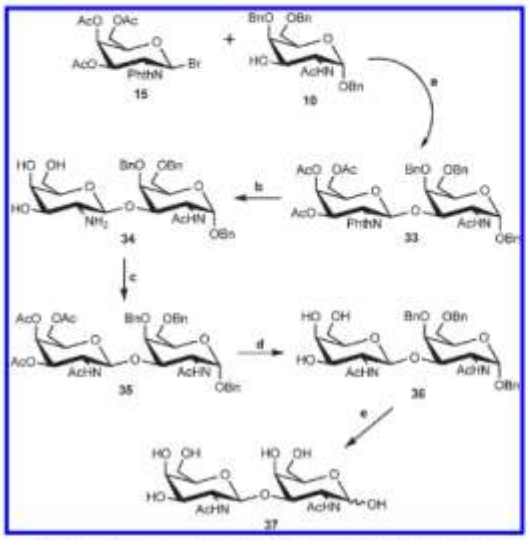
[a] Reagents and conditions: (a) $AgClO_4$, Ag_2CO_3, CH_2Cl_2, −15 °C; (b) MeONa, MeOH, room temp; (c) $N_2H_4.H_2O$, EtOH, reflux; (d) Ac_2O, pyridine, room temp; (e) H_2, Pd/C, AcOH - H_2O, room temp.

obtained by deallylation of compound **8**, was glycosylated by 3 molar excess of 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-*β*-D-galactopyranosyl bromide[34] (**15**), employing promotion by silver perchlorate in the presence of silver carbonate (Scheme 6). This reaction afforded not only branched trisaccharide **57** (9%) but also disaccharide **56** (13%) as an undesired side product resulting from the lower reactivity of the axial C(4)−OH group of the glycosyl acceptor. After deprotection as above we obtained the target biantennary disaccharide **61**.

Biological Evaluation. In order to evaluate the binding affinity of the synthesized oligosaccharides for the target NK cell receptors, we started with the recombinant rat NKR-P1A receptor shown previously[13,18] to have high affinity for GlcNAc monosaccharide and related compounds. Microtiter plates were coated with high affinity ligand $GlcNAc_{23}BSA$ neoglycoprotein, and individual compounds were tested as inhibitors of binding of the soluble radiolabeled receptor to these plates. Results shown in Figure 1 indicate that the synthesized compounds were mostly average or poor ligands compared to the GlcNAc control. Here, in the linear GlcNAc/GalNAc series, the $\beta1-4$ linkage is preferred to other ($\beta1-3$ or $\beta1-6$) linkages (Figure 1A). Branching of the oligosaccharides resulted in significant decrease in the inhibitory potencies independently of the series used (Figure 1B and Figure 1C).

More interesting results were obtained when the synthesized compounds were tested as inhibitors of binding of another lymphocyte receptor, human CD69, to its high affinity $GlcNAc_{23}BSA$ ligands.[14,20] This receptor has been

Scheme 4[a]



[a] Reagents and conditions: (a) $AgClO_4$, Ag_2CO_3, CH_2Cl_2, −15 °C; (b) $N_2H_4{\cdot}H_2O$, EtOH, reflux; (c) Ac_2O, pyridine, room temp; (d) MeONa, MeOH, room temp; (e) H_2, Pd/C, AcOH−H_2O, room temp.

shown previously[16,20] to contain multiple binding sites for GlcNAc in its molecule and thus to prefer branched *N*-acetyl-D-hexosamine sequences to linear ones.[17] Indeed, only minor differences have been found in the linear GlcNAc/GalNAc series compared to the GlcNAc monosaccharide control (Figure 1D). However, a notable hierarchical increase in inhibitory potencies has been found in the branched GlcNAc/GalNAc series (Figure 1E). This increase has been considerably more profound in the GlcNAc series compared to the GalNAc series (Figure 1E).

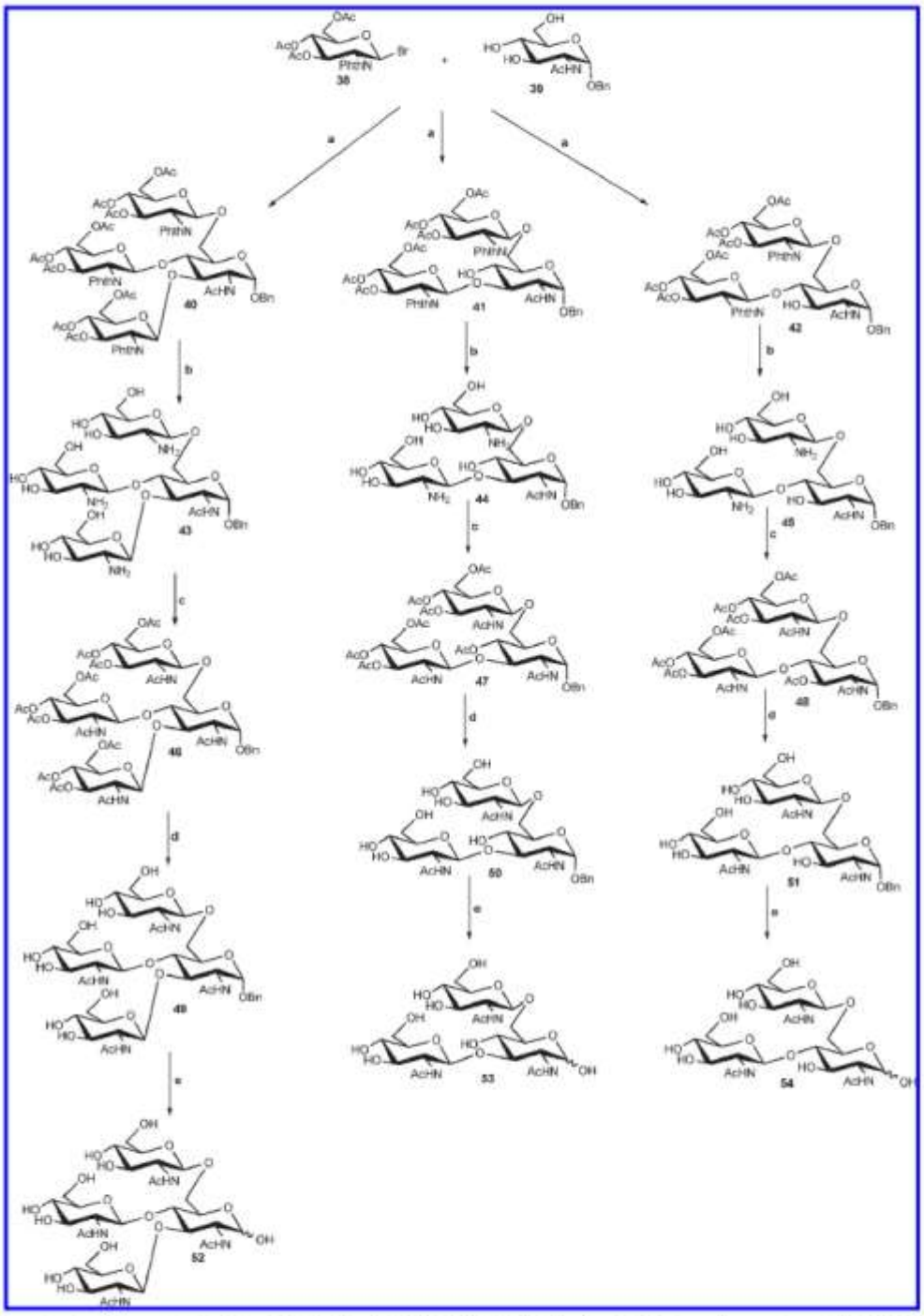
On the basis of these findings, detailed structure−activity studies were performed in the branched GlcNAc series. Here, an even more developed hierarchy of the gradually increasing affinities could be seen (Figure 1F). The trisaccharide BN described previously[19] has already attained 10 times better inhibitory potency compared to the GlcNAc control. The newly synthesized compounds **53** and **54** have reached 100 times and 1000 times greater inhibitory activities, respectively. The small, fully branched tetrasaccharide **52** was an even better inhibitor, achieving 10 000 times better inhibitory activity than the GlcNAc control. Its corresponding α-benzyl derivative was even 10 times more efficient, attaining the potency of the much more complex natural oligosaccharide OM, and it was a 100 000 times more potent inhibitor compared to the *N*-acetyl-D-glucosamine monosaccharide (Figure 1F). Thus, the tetrasaccharide **52** has been selected for further work as an efficient mimetic of a much larger ovomucoid undecasaccharide (OM, Chart 1) described previously as natural ligand,[17] as well as of the classical artificial high affinity ligand, GlcNAc neoglycoprotein (NG, Chart 1).[14]

Direct Binding Assay for 52 and 49. The high affinity interaction of **49** with the soluble recombinant lymphocyte receptor CD69 was further confirmed using several techniques allowing us to observe the binding directly. Compound **49** appeared to be particularly suitable for these studies, since its anomeric configuration is fixed in the α-position through the benzyl residue. The initial confirmation of high affinity binding of **49** to recombinant human CD69 was provided by NMR titration (Figure 2A). Evaluation of the binding curve revealed a complete disappearance of the acetate signals due to specific binding until the saturation point corresponding to one bound molecule of the oligosaccharide per receptor subunit, after which a linear increase of the free (unbound) ligand was observed. Evaluation of the binding curve also provided a quantitative estimation for the value of K_d in the nanomolar range, which is in complete agreement with the inhibition data. Two additional binding techniques were employed. Direct binding assay using ^{3}H-labeled tetrasaccharide **52** allowed us to establish the details of the binding parameters: there was one high affinity binding site per receptor subunit (two sites per receptor dimer) with $K_d = 3.2 \times 10^{-9}$ M (Figure 2B). Moreover, the binding to the dimeric receptor displayed the same cooperativity reported previously for the high affinity binding site for GlcNAc with the Hill coefficient approaching the maximal theoretical value for a two-subunit protein ($h_{theor.} = 2$, $h_{expt.} = 1.98$).[41] The third direct binding assay based on tryptophan fluorescence quenching was performed using both the dimeric CD69 receptor[42] and its monomeric form obtained by site-directed mutagenesis.[41] This arrangement allowed us to confirm the dependence of both affinity and ligand binding cooperativity on the dimeric arrangement of the receptor: the K_d value dropped from 3.4×10^{-9} to 1.12×10^{-7} M, and only binding according to the single-site model could be observed for the monomeric protein (Hill's coefficients were 1.94 and 1.05 for the dimeric and for the monomeric protein, respectively; Figure 2C and D).

Dimerized 52 Efficiently Precipitates Soluble CD69 Depending on Linker Chemistry. High affinity binding of **52** to CD69 shown by both inhibition and direct binding tests indicated that this oligosaccharide might represent a suitable minimum mimetic of the complex physiological ligand for the receptor. In order to activate CD69$^+$ immune cells through an engagement of this antigen, however, the ligand mimetic must be present in a multivalent (at least bivalent) form. To achieve efficient dimerization of the tetrasaccharide **52**, we have adopted the standard chemistry used for peptide and protein cross-linking[24] and elaborated it for efficient activation and coupling of oligosaccharides activated as β-glycosylamines (Scheme 7).[43] The entire procedure involves a conversion of the reducing oligosaccharide **52** into the corresponding β-glycosylamine **62**, followed by reaction with thiophosgene to produce the reactive isothiocyanate **63**. This compound is then coupled to a series of linear aliphatic diamines to produce the dimeric tetrasaccharides **64**−**67** having a linker of varying length defined by two to eight methylene groups. The synthesized and purified GlcNAc tetrasaccharide dimers were evaluated for their abilities to precipitate soluble CD69 protein in a process that resembles the cross-linking of the cellular form of the receptor. While the monomeric tetrasaccharide **52** was not active in this test, the dimeric compounds **64**−**67** proved to be positive (Figure 3A). The best activity was achieved for compound **65** bearing the butyl aliphatic linker.

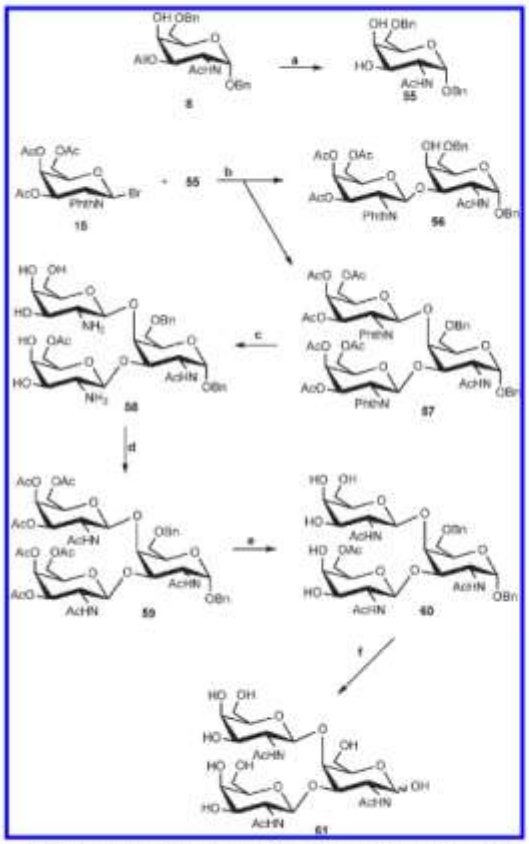
Recently published work has provided a clear indication that not only the length but also the chemical nature of a linker, such as its hydrophobicity and flexibility, may significantly influence the outcome of an interaction of a bivalent artificial ligand with a corresponding receptor.[44]

Scheme 5[a]



[a] Reagents and conditions: (a) $AgClO_4$, Ag_2CO_3, CH_2Cl_2, −15 °C; (b) $N_2H_4 \cdot H_2O$, EtOH, reflux; (c) Ac_2O, pyridine, room temp; (d) MeONa, MeOH, room temp; (e) H_2, Pd/C, AcOH−H_2O, room temp.

Scheme 6[a]



[a] Reagents and conditions: (a) $(Ph_3P)_3RhCl$, toluene−EtOH−H_2O, reflux, and then HCOOH, reflux; (b) $AgClO_4$, Ag_2CO_3, CH_2Cl_2, −15 °C; (c) $N_2H_4.H_2O$, EtOH, reflux; (d) Ac_2O, pyridine, room temp; (e) MeONa, MeOH, room temp; (f) H_2, Pd/C, AcOH−H_2O, room temp.

Therefore, we took the dimeric tetrasaccharide **65** as the lead compound and prepared three additional similar compounds, **68**, **69**, and **70**, differing in the hydrophobicity and flexibility of the linker based on the starting divalent amine (Scheme 8) used. Compounds **65**, **68**, **69**, and **70** thus all had the optimal length of the four-carbon linker, albeit in four different chemical variants: hydrophobic flexible, hydrophilic flexible, hydrophobic rigid, and hydrophilic rigid, respectively. These compounds, after synthesis and purification, were again tested in the receptor precipitation assay. The results show that optimal activity was achieved for compound **68** containing the hydrophilic flexible linker (Figure 3B). This activity was comparable to that of the previously described lactose-di-*N*-acetyl dimer.[45] The original lead compound **65** with the hydrophilic flexible linker was also quite active, but compounds **69** and **70** with the rigid linker were much less active.

Dimerized Tetrasaccharides 68 Activate Lymphocytes Expressing High Levels of CD69. We decided to verify our conclusions concerning the optimal immunostimulating compound further using $CD69^{high}$ lymphocytes containing the cellular form of the receptor in its natural dimeric and

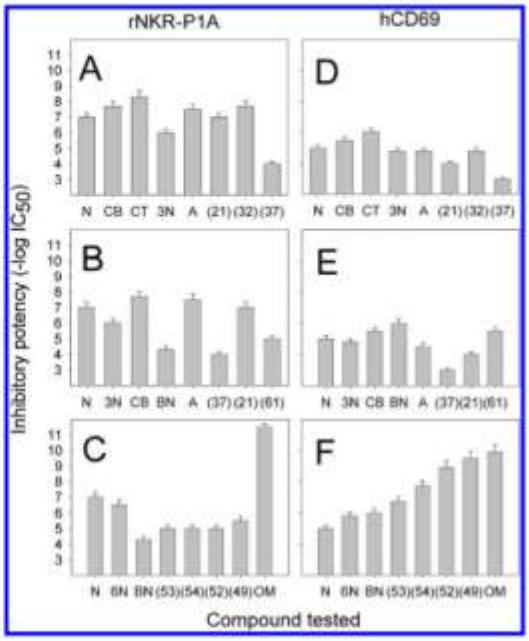


Figure 1. Biological testing of the synthesized HexNAc based oligosaccharides using inhibition assay. Indicated compounds were tested as the inhibitors of binding of the radiolabeled rNKR-P1A (left) or hCD69 (right) to the high affinity $GlcNAc_{23}BSA$ ligand. From the complete inhibition curves, IC_{50} values were calculated. Shown are the mean ± SD from at least three independent experiments.

fully glycosylated form.[46] Such verification is essential because the activities using the soluble and the cellular forms of the immune receptors may differ significantly.[17] We used the standard cellular activation assays based on the production of inositol phosphates and monitoring of intracellular calcium concentrations.[11] We isolated $CD69^{low}$ (<5% surface expression) and $CD69^{high}$ (>30% surface expression) lymphocytes from human peripheral blood mononuclear cells. Both monomeric tetrasaccharide **52** and four dimeric tetrasaccharides described above (compounds **65**, **68**, **69**, and **70**) were dissolved in PBS for testing as activators of these cellular populations and were compared to the effect of PBS only (negative control), or of specific monoclonal antibodies against CD69 as well as the *N*-acetyllactosamine disaccharide described previously[41,45] to be a potent activator of $CD69^+$ lymphocytes (positive controls). All tested compounds were added at 10-fold molar excess over the estimated amount of surface CD69, as well as over the measured K_d concentration. The inositol phosphate production results show that very little activation of $CD69^{low}$ lymphocytes occurred despite the presence of a small amount of $CD69^+$ cells in this population (Table 1). On the other hand, the dimerized (although not the monomeric) tetrasaccharides **65**−**70** all had notable effects on the production of inositol phosphates with the best compound (**68**) exerting effects comparable to those of the positive controls, monoclonal antibody, and *N*-acetyllactosamine disaccharide (Table 1). The effect of the tested compounds on an increase in intracellular calcium levels followed the effects on inositol phosphates production. In $CD69^{low}$ cellular population only the

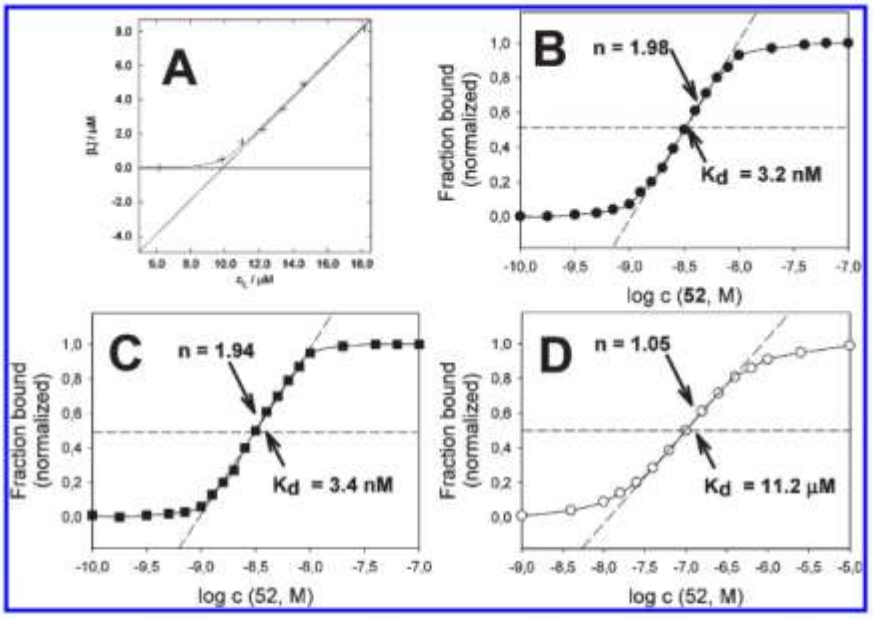
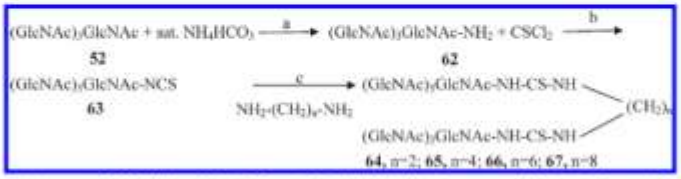


Figure 2. Direct binding of compound **52** to human CD69 receptor: (A) NMR titration of 10 μM soluble CD69 with **49**; (B) binding of ^{3}H-labeled **52** to CD69 followed by equilibrium dialysis; (C, D) binding of **52** to the dimeric and monomeric form of CD69 followed by tryptophan fluorescence quenching.

Scheme 7[a]



[a] Reagents and conditions: (a) sat. aq NH_4HCO_3, 30 °C, 7 days; (b) $CSCl_2$, $NaHCO_3$, acetone−H_2O; (c) 2.4 equiv of **63** per 1 equiv of diamine, CH_2Cl_2, room temp, 2 days.

addition of ionomycin resulted in a significant increase in intracellular calcium (Figure 4A). In $CD69^{high}$ population, all four dimeric tetrasaccharides were active, and the activity of compound **68** was comparable with that of monoclonal antibody positive control. Testing the compounds further at 10 times higher concentrations did not increase their activity (data not shown).

Dimerized Tetrasaccharide 68 Increased the Killing of Tumor Cell Lines in Vitro. To estimate the efficiency of the synthesized compounds in enhancing the antitumor potential of the immune system, we have first tested their effects in the standard short-term (4 h) cytotoxicity assays. Compound **68** and the lead compound **65** increased significantly the killing of human erythroleukemic cell line K562, a standard target cell line known to be sensitive for natural killing. The effect of compound **65** was comparable to that of the two monoclonal antibodies against CD69 used as positive controls, while compound **68** had even much higher effect enhancing the efficiency of natural killing about 4.5 times under the given experimental conditions in vitro (Figure 5A). Moreover, compound **68** was also active in the case of NK resistant tumor cell line RAJI (Figure 5B) in the situation where other compounds or monoclonal antibodies used as positive controls had little effect. Under these experimental conditions compound **68** enhanced natural killing about 3.5 times (Figure 5B).

Dimerized Tetrasaccharides 68 Suppresses the Growth of Experimental Tumors and Activates Tumor Infiltrating Lymphocytes. An initial assessment of the antitumor efficacy of the synthesized compounds was performed using the experimental model of mouse B16 melanoma using a low metastasis variant.[47] In this assay, compound **68** was most efficient decreasing the size of the tumors at day 26 and day 30 (Table 2). Compounds **65** and **70** also had some effects seen at day 30 after the injection of tumor cells. Interestingly, the two monoclonal antibodies against CD69 and the dimerized *N*-acetyllactosamine had very little effect in this assay (Table 2). In order to assess the possible effects of the dimerized tetrasaccharides on the immune system, a large number of immune system parameters must be monitored. However, one of the most important parameter is the activity of killer cells operating inside the tumors (especially tumor infiltrating lymphocytes and dendritic cells). In order to estimate this parameter, we performed cytotoxicity assay

using tumor (melanoma) infiltrating lymphocytes isolated from animals treated with the individual compounds and assayed after isolation of these cells ex vivo. Notably, compound **68** was the only compound among the tested set that remained effective in this assay using both B16 melanoma and NK resistant P815 mastocytoma (Figure 6A and Figure 6B, respectively). Compound **65** also had a statistically significant effect but only in the case of the B16 melanoma (Figure 6A). We have initially tested the effect of the injection of 1 μmol (approximately 2 mg) of the individual compounds, a dose that was much smaller compared to the tested antibodies that had to be injected in 10 mg amounts. While increasing the amounts of the tested compounds had very little benefit, injecting 0.1 μmol amounts had very similar effects, and 0.01 μmol amounts gave no effect.

Discussion

Although the physiological ligand for the widespread leukocyte activation marker and triggering receptor CD69 has not been identified, calcium and certain *N*-acetyl-D-hexosamines have been shown previously to be specific ligands for this receptor.[14,17] Direct binding assays and molecular modeling studies revealed the existence of three binding sites for GlcNAc in the CD69 molecule, one of which represents the high affinity binding site with an estimated value of K_d of 63 μM.[20] Later, by use of a new construct for recombinant

Table 1. Activation of Purified $CD69^{low}$ and $CD69^{high}$ NK Cells Using the Dimeric HexNAc Based Oligosaccharides

cell	control	Mab	LAC	**52**	**65**	**68**	**69**	**70**
$CD69^{low}$	60[a]	58[a]	72[a]	55[a]	70[a]	65[a]	56[a]	52[a]
	82[b]	77[b]	75[b]	65[b]	71[b]	78[b]	72[b]	75[b]
$CD69^{high}$	58[a]	1252[a]	1123[a]	58[a]	750[a]	1310[a]	520[a]	720[a]
	76[b]	1420[b]	1374[b]	60[b]	820[b]	1455[b]	610[b]	800[b]

[a] Amount of insP2 was measured 4 min after the addition of individual compounds. [b] Amount of insP3 (cpm) was measured 2 min after the addition of individual compounds.

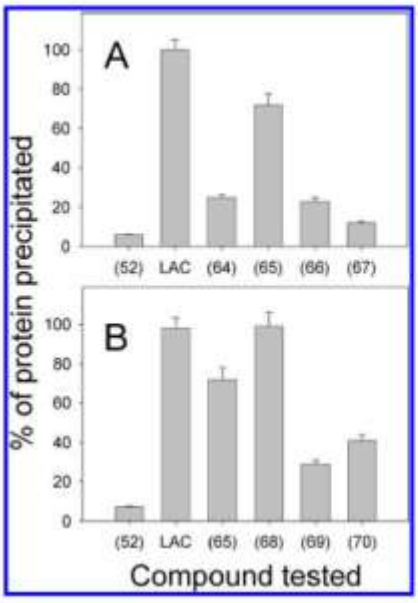


Figure 3. Precipitation of soluble recombinant CD69 by equimolar amounts of the tested ligand: (A) monomeric and dimerized **52** with linker of different length; (B) monomeric and dimerized **52** with linker having various chemical properties (hydrophobicity and rigidity). LAC designates the lactose-di-*N*-acetyl dimer described previously.[45]

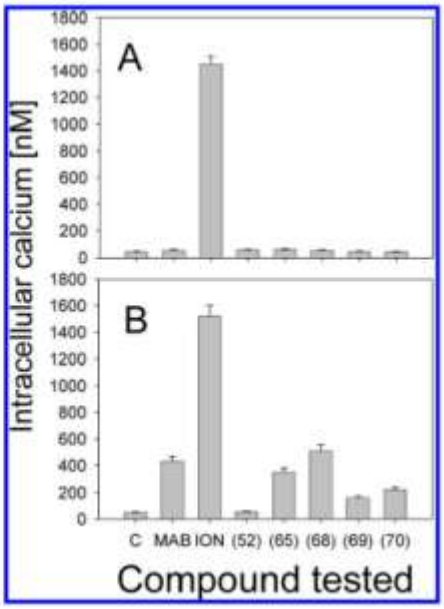
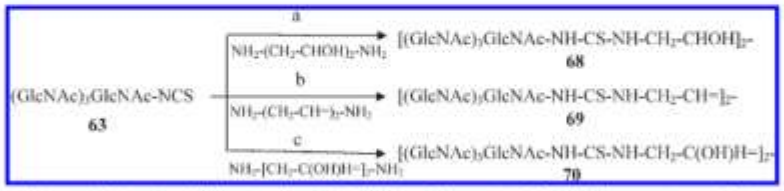


Figure 4. Monitoring of intracellular calcium 3 min after the addition of tested compounds was performed in $CD69^{low}$ (A) and $CD69^{high}$ (B) lymphocytes. ION is the ionomycin positive control. All changes in (B) were statistically significant at $p \leq 0.01$ compared to the PBS control (C). Values for the tested compounds and ionomycin were taken from the complete curves at 3 and 11 min, respectively.

Scheme 8[a]



[a] Reagents and conditions: (a) 1 equiv of 1,4-butanediamine-2,3-diol per 2.4 equiv of **63**, CH_2Cl_2, room temp, 2 days; (b) 1 equiv of 1,4-butanediamine-2-en per 2.4 equiv of **63**, CH_2Cl_2, room temp, 2 days; (c) 1 equiv of 1,4-butanediamine-2,3-diol-2-en, CH_2Cl_2, room temp, 2 days.

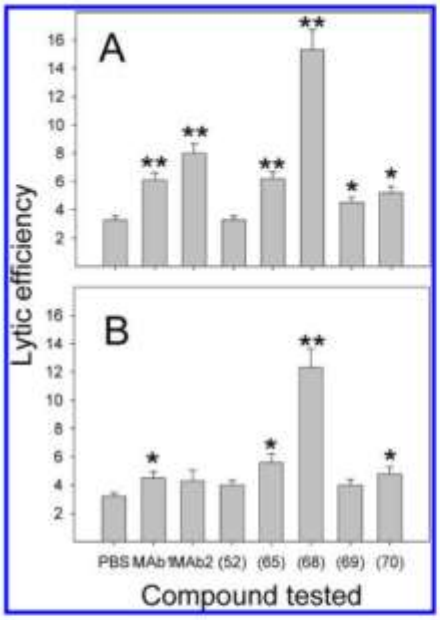


Figure 5. Natural killing assays in the presence of the tested compounds using sensitive human cell line K562 (A) and resistant human cell line RAJI (B). Statistical significance of changes compared to PBS control is marked by asterisks as described in Experimental Section.

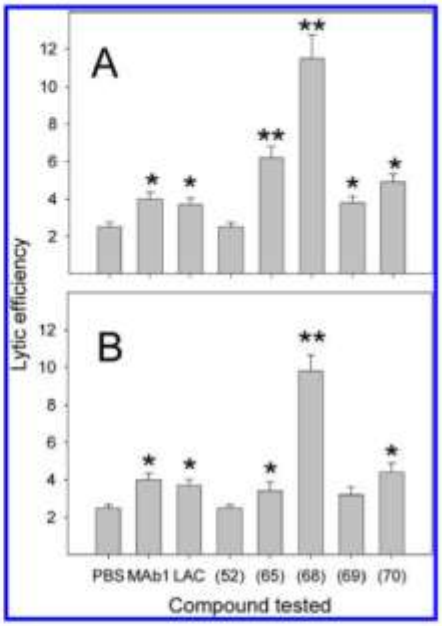


Figure 6. Natural killing of tumor infiltrating lymphocytes isolated from mice treated with the indicated compounds. Killing of B16 melanoma (A) and NK resistant mastocytoma P815 (B) targets is shown. Statistical significance of changes compared to PBS control is marked by asterisks as described in Experimental Section.

Table 2. Initial Assessment of the Antitumor Properties of the Synthesized Compounds Using the Mouse B16S Melanoma Model

	tumor size (cm^2)	
compd	day 26	day 30
PBS	0.6 ± 0.2	1.3 ± 0.3
MAb1	0.6 ± 0.1	1.1 ± 0.2
MAb2	0.4 ± 0.1	0.8 ± 0.2
LAC	0.4 ± 0.1	0.7 ± 0.1
65	0.5 ± 0.1	0.8 ± 0.2
68	0.1 ± 0.1	0.2 ± 0.1
70	0.6 ± 0.2	0.8 ± 0.2

expression of CD69 optimized to have a high physical and biochemical stability,[42] direct binding assays including NMR titrations, equilibrium dialysis, and fluorescence quenching measurements revealed that binding of GlcNAc to the dimeric CD69 protein proceeded in a cooperative fashion with K_d = 0.4 μM.[41] However, after dissociation into the monomeric subunits when no cooperativity in carbohydrate binding could occur, the value of K_d dropped to about 16 μM using the optimized monomeric protein.[41] Recently, high affinity physiological ligand for CD69 has been identified among the highly branched ovomucoid type oligosaccharides.[17] In particular, the pentabranched undecasaccharide having a structure [GlcNAcβ1−2(GlcNAcβ1−4)(GlcNAcβ1−6)Manα1−6][GlcNAcβ1−2(GlcNAcβ1−4)Manα1−3][GlcNAcβ1−4]Manβ1−4GlcNAcβ1−4GlcNAc turned out to be an efficient ligand for CD69 with K_d in the low nanomolar range. This ligand may represent a target structure for $CD69^+$ NK cells, since it could be detected at the surface of NK sensitive targets cell lines and could be induced by stress in NK resistant targets as well.[17] This undecasaccharide may thus represent a suitable initial compound targeting killer lymphocytes including natural killer cells to tumor sites and keeping the killer lymphocytes active in the inhibiting microenvironment of the tumors. Since this undecasaccharide turned out to be available only in small amounts after complicated isolation from biological material[17] and since the total chemical synthesis of such complex oligosaccharide was shown to be very complicated,[21] considerable experimental efforts have been aimed toward the synthesis and identification of smaller oligosaccharide mimetics.

Here we describe one approach toward the development of efficient oligosaccharide mimetic for CD69 based on the synthesis of *N*-acetyl-D-hexosamine-based homooligosaccharides and their dimerization through several types of chemical linkers in order to prepare bivalent compounds suitable for CD69 cross-linking. The initial screening of GlcNAc and GalNAc based homooligosaccharides, both linear and branched, identified the fully branched tetrasaccharide **52** as the best ligand for CD69 with affinity in the nanomolar range, thus approaching the affinity of the much more complex ovomucoid derived undecasaccharide and that of the classical high affinity ligand, $GlcNAc_{23}BSA$ neoglycoprotein. The α-benzyl derivative of this compound was an even better inhibitor, indicating the possibility of additional hydrophobic interaction near the carbohydrate recognition site of CD69.[18] This tetrasaccharide was then used as a lead for the development of an efficient dimerization protocol and optimalization of the linker used for dimerization (length, hydrophobicity, rigidity). The previously developed assay measuring the amount of soluble CD69 precipitated after interaction with the bivalent ligand[45] has been used to provide a simple test of

binding efficiency of the dimerized GlcNAc tetrasaccharides. From these tests, dimerized tetrasaccharide **68** bearing a hydrophilic flexible linker emerged as the most promising compound. Further biological evaluation revealed this compound to be effective in cellular activation of $CD69^{high}$ NK cells, which led to the ability to enhance natural killing in vitro, to decrease the rate of tumor growth in vivo, and to keep the tumor infiltrating lymphocytes activated even in aggressive tumor microenvironment, as revealed after ex vivo examination of their cytotoxicity. Moreover, this compound compared favorably with other reagents cross-linking the CD69 target receptor; it was active in much smaller concentration than the specific monoclonal antibodies against this receptor, and most probably it would also be much less immunogenic.

The exact role of CD69 antigen in NK cell biology is not fully understood. First, CD69 is expressed on many leukocyte subsets[48] and thus cannot be considered as an NK cell specific receptor. Second, the work by Sanchez-Madrid and colleagues using $CD69^{-/-}$ mice provided a clear indication for the negative role of this antigen in tumor cell killing, since the mice lacking surface expression of CD69 were more resistant to tumors.[49] On the other hand, the work of Moretta et al. emphasized that monoclonal antibodies (and perhaps ligands) that are strongly bound to CD69 can activate CD69 positive cells even in the absence of other (antigen dependent) proliferation pathways.[50] Our present work provides a strong support for such a possibility and advocates the dimerized tetrasaccharide **68** as a strong candidate for further systematic experiments looking at the details of its efficacy in other experimental tumor models and the details of the molecular mechanisms of its action.

Conclusions

We prepared efficient mimetics of natural oligosaccharide ligands for CD69, a widespread receptor triggering leukocyte activation. Homooligomeric fully branched GlcNAc tetrasaccharide **52** proved to be the most efficient ligand and could be used as a lead for the development of effective activators of $CD69^+$ NK cells. Compound **68**, GlcNAc tetrasaccharide dimerized through a hydrophilic flexible linker, turned out to be active in enhancing natural killing in vitro, decreasing the growth of tumors in vivo, and increasing cytotoxic activity of tumor infiltrating lymphocytes examined ex vivo. This compound thus represents a strong candidate for an efficient carbohydrate-based immunomodulator with a promising antitumor potential.

Experimental Section

General Chemistry. All reagents and solvents were purchased from Sigma-Aldrich and used as received. Analytical samples were dried at 6.5 Pa and 25 °C for 8 h. Melting points were determined with a Kofler apparatus and are uncorrected. Optical rotations were measured on Rudolph Research Analytical AUTOPOL IV polarimeter at 20 °C. $[\alpha]_D$ values are given in $deg \cdot cm^3 \cdot g^{-1}$. Elemental analyses were performed using a PerkinElmer 2400 series II CHNS/O elemental analyzer. IR spectra were recorded on a Bruker Equinox 55 FTIR spectrometer, and wavenumbers are given in cm^{-1}. NMR spectra were recorded using Bruker Avance spectrometer at 500.1 MHz (^1H) and 125.8 MHz (^{13}C) in $CDCl_3$, using TMS as an internal standard for ^1H NMR spectra and $CDCl_3$ as a standard for ^{13}C NMR spectra. Chemical shifts are given in ppm (δ scale) and coupling constants (J) in Hz. For unambiguous assignment of signals in ^{13}C NMR spectra, the heterocorrelated 2D NMR spectra were measured by the HSQC technique[51] if necessary. Positive-ion FAB mass spectra were acquired on a BEqG geometry mass spectrometer ZAB-EQ (VG Analytical). Complex mixtures were analyzed using LCQ mass spectrometric detector (Finnigan) coupled to a HPLC system for LC/MS applications, equipped with an ion-trap analyzer. +APCI ionization was used for recording spectra. TLC was carried out on Merck aluminum sheets silica gel 60 F254, and column chromatography was carried out on Fluka silica gel 60 (40−63 μm). Analytical RP HPLC was performed using Waters Alliance HPLC system (PDA 996 detector) equipped with a column (150 mm × 3.9 mm) filled with Nova-Pak C18 (4 μm, Waters). Purity of all synthesized compounds was determined to be ≥95% by RP HPLC using a water−methanol gradient. Preparative RP HPLC was performed with a Knauer system equipped with a column (250 mm × 25 mm) filled with LiChrosorb RP-18 (5 μm, Merck).

2-Acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy-D-galactopyranose (21). Hydrogenolysis of compound **20** (50 mg, 0.08 mmol) according to general procedure D afforded 25 mg (74%) of an α/β-anomeric mixture of compound **21**. $[\alpha]_D$ −98 (*c* 0.2, H_2O). FAB MS calcd for $C_{16}H_{28}N_2O_{11}$ 424.2, found *m/z* 425.1 $[M + H]^+$. For $C_{16}H_{28}N_2O_{11}$ (424.4): calcd 45.28% C, 6.65% H, 6.60% N; found 45.13% C, 6.76% H, 6.51% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy-D-galactopyranose (32). Hydrogenolysis of compound **31** (85 mg, 0.08 mmol) according to general procedure D afforded 42 mg (85%) of an α/β-anomeric mixture of compound **32**. $[\alpha]_D$ +10 (*c* 0.3, H_2O). MS EI calcd for $C_{24}H_{41}N_3O_{16}$ 627.2, found *m/z* 650.5 $[M + Na]^+$. For $C_{24}H_{41}N_3O_{16}$ (627.6): calcd 45.93% C, 6.58% H, 6.70% N; found 46.08% C, 6.53% H, 6.73% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose (37). Hydrogenolysis of compound **36** (100 mg, 0.14 mmol) according to general procedure D afforded 30 mg (50%) of an α/β-anomeric mixture of compound **37**. $[\alpha]_D$ +56 (*c* 0.1, H_2O). FAB MS calcd for $C_{16}H_{28}N_2O_{11}$ 424.2, found *m/z* 425 $[M + H]^+$. For $C_{16}H_{28}N_2O_{11}$ (424.4): calcd 45.28% C, 6.65% H, 6.60% N; found 45.13% C, 6.76% H, 6.48% N. For NMR spectra see Supporting Information.

Benzyl-2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→4)]-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→6)]-2-acetamido-2-deoxy-α-D-glucopyranoside (49). Zemplén O-deacetylation of compound **46** (225 mg, 0.17 mmol) according to general procedure A followed by RP HPLC in in water afforded 132 mg (83%) of compound **49**. $[\alpha]_D$ −41 (*c* 0.2, H_2O). MS EI calcd for $C_{39}H_{60}N_4O_{21}$ 920.4, found *m/z* 943.4 $[M + Na]^+$. For $C_{39}H_{60}N_4O_{21}$ (920.9): calcd 50.86% C, 6.57% H, 6.08% N; found 50.98% C, 6.64% H, 5.96% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→4)]-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→6)]-2-acetamido-2-deoxy-D-glucopyranose (52). Hydrogenolysis of compound **49** (25 mg, 0.03 mmol) according to general procedure D afforded 17 mg (66%) of an α/β-anomeric mixture of compound **52**. $[\alpha]_D$ −25 (*c* 0.2, H_2O). FAB MS calcd for $C_{32}H_{54}N_4O_{21}$ 830.3, found *m/z* 853.5 $[M + Na]^+$. For $C_{32}H_{54}N_4O_{21}$ (830.8): calcd 46.26% C, 6.55% H, 6.74% N; found 46.09% C, 6.53% H, 6.78% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→6)]-2-acetamido-2-deoxy-D-glucopyranose (53). Hydrogenolysis of compound **50** (90 mg, 0.13 mmol) according to general procedure D afforded 2 mg (15%) of an α/β-anomeric mixture of compound **53**. $[\alpha]_D$ −93 (*c* 0.1, H_2O). MS EI calcd for $C_{24}H_{41}N_3O_{16}$ 627.2, found *m/z* 650.3 $[M + Na]^+$. For $C_{24}H_{41}N_3O_{16}$ (627.6): calcd 45.93% C, 6.58% H, 6.70% N; found 45.89% C, 6.53% H, 6.78% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-glucopyranosyl-(1→4)-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→6)]-2-acetamido-2-deoxy-D-glucopyranose (54). Hydrogenolysis of compound **51** (230 mg, 0.32 mmol) according to general procedure D afforded 133 mg (66%) of an α/β-anomeric mixture of compound **54**. $[\alpha]_D$ −65 (*c* 0.2, H_2O). FAB MS calcd for $C_{24}H_{41}N_3O_{16}$ 627.2, found *m/z* 628 $[M + H]^+$. For $C_{24}H_{41}N_3O_{16}$ (627.6): calcd 45.93% C, 6.58% H, 6.70% N; found 45.79% C, 6.73% H, 6.59% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-galactopyranosyl-(1→3)-[2-acetamido-2-deoxy-β-D-galactopyranosyl-(1→4)]-2-acetamido-2-deoxy-D-galactopyranose (61). Hydrogenolysis of compound **60** (28 mg, 0.32 mmol) according to general procedure D afforded 14 mg (64%) of an α/β-anomeric mixture of compound **54**. $[\alpha]_D$ +25.77 (*c* 0.3, H_2O). FAB MS calcd for $C_{24}H_{41}N_3O_{16}$ 627.2, found *m/z* 628 $[M + H]^+$. For $C_{24}H_{41}N_3O_{16}$ (627.6): calcd 45.93% C, 6.58% H, 6.70% N; found 45.89% C, 6.63% H, 6.78% N. For NMR spectra see Supporting Information.

2-Acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→4)]-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→6)]-2-acetamido-2-deoxy-D-glucopyranose Thiourea Dimerized through Hydrophobic Linkers (64−70). Compound **52** (24 mg) was incubated in saturated aqueous NH_4HCO_3 (10 mL) at 30 °C for 7 days.[52] The bulk of NH_4HCO_3 was removed by lyophilization. Dowex 50W 1X2 H^+ (Fluka) was added to a solution of the solid sugar amine in water until the pH of the mixture dropped to 4. The resin with absorbed glycosylamines was filtered off and washed with ice cold water. Pure product was eluted with 2 M NH_3 in MeOH, affording 20.6 mg of the corresponding β-glycosylamine **62** (86%). This compound was dissolved in 2 mL of water and added dropwise within 3 min at room temperature to a stirred mixture of 400 μL of $CSCl_2$ and 200 mg of $NaHCO_3$ in 2 mL of acetone. After 45 min, water (2 mL) was added and acetone was removed in vacuo at 30 °C. Unreacted $CSCl_2$ was extracted with chloroform, and the resulting solution of glycosyl isothiocyanate **63** was lyophilized and immediately used in the subsequent coupling reaction. A solution of **63** (2.4 equiv) in dichloromethane was slowly added to the corresponding diamine (1 equiv) in dichloromethane. After 24 h at room temperature the reaction mixture was purified by column chromatography on silica gel (EtOAc−MeOH, 9:1). Thereafter, the compounds were purified by chromatography on BioGel P2 (BioRad) as described previously.[33] The yields of the individual compounds were 2.1 mg (66%) for **64**, 1.8 mg (62%) for **65**, 2.0 mg (65%) for **66**, 1.7 mg (58%) for **67**, 1.5 mg (50%) for **68**, 1.8 mg (60%) for **69**, and 2.1 mg (67%) for **70**. Analytical data for **64**: ^{1}H NMR 8.14 d (NH), 5.07 d (1H, *J* = 3.0, H-1), 4.82 d (1H, *J* = 8.3, H-1′), 4.75 d (1H, *J* = 8.3, H-1′), 4.70 d (1H, *J* = 8.4, H-1″), 4.68 d (1H, *J* = 8.4, H-1″), 4.68 d (1H, *J* = 8.4, H-1), 4.59 d (1H, *J* = 8.5, H-1‴), 4.58 d (1H, *J* = 8.5, H-1‴), 4.26 m (1H, H-2), 4.06 m (2H, H-2, 3), 3.65−3.86 m (3H, H-2, 2′, 2″), 3.47−4.07 m (18H, 3′, 3″, 3‴, 4′, 4″, 4‴, 5′, 5″, 5‴, 6′, 6″, 6‴), 2.07 m (CH_2). FAB MS calcd for $C_{68}H_{114}N_{12}O_{40}S_2$ 1802.3, found *m/z* 1825.3 $[M + Na]^+$. For $C_{68}H_{114}N_{12}O_{40}S_2$ (1802.3): calcd 45.27% C, 6.32% H, 9.32% N; 3.55% S; found 45.32% C, 6.43% H, 9.28% N, 3.32% S. For NMR spectra see Supporting Information. Analytical data for compounds **65**−**70** are given in the Supporting Information.

Binding and Inhibition Experiments. Soluble dimeric rat NKR-P1A and soluble dimeric human CD69 were expressed in *Escherichia coli*, and purified essentially as described previously.[13,42] These proteins were radioiodinated as reported,[13] with carrier-free $Na^{125}I$ (Amersham) to a specific activity of 10^7 cpm per μg of protein. Binding and inhibition assays were performed as described previously[13] with minor modifications. Briefly, 96-well poly(vinyl chloride) microplates (Titertek Immuno Assay-Plate, ICN Flow, Irvine, U.K.) were coated overnight at 4 °C with 50 μL of Glc*p*NAc$_{17}$BSA (10 μg/mL, Sigma) in TBS + C buffer (10 mM Tris-HCl, pH 8.0, with 150 mM NaCl, 1 mM $CaCl_2$, and 1 mM NaN_3). Plates were blocked with 1% BSA (Sigma) in TBS + C for 2 h at 4 °C and incubated with ^{125}I-NKR358 corresponding to half of the saturation amount and the indicated dilutions of the tested compounds in a total reaction volume of 100 μL. Plates were washed three times with TBS + C, drained, and dried, and 100 μL of a scintillation solution was added to each well. Radioactivity in wells was determined in a β-counter Microbeta (Wallac). All experiments were performed in duplicate. The results are average values from duplicate experiments within the range indicated by error bars.

NMR Titrations. All NMR experiments were run at 300 K in a Bruker Avance 600 MHz spectrometer equipped with a cryogenic H/C/N TCI probehead. ^{1}H−^{15}N HSQC spectra of 0.3 mM [^{15}N]-labeled wild-type CD69 protein were used as a routine check of protein folding and stability. The sample buffer consisted of 10 mM MES, pH 5.8, with 49 mM NaCl, 1 mM NaN_3, and 10% D_2O. During NMR titration, a 0.1 mM solution of the unlabeled wild-type CD69 protein was titrated. In an initial experiment, aliquots of the GlcNAc ligand corresponding to 25%, 50%, 75%, 100%, 200%, and 500% of saturation were added and signals of the free GlcNAc ligand were observed at 2.2 ppm in the 1D proton spectra and used for the estimation of the free ligand concentration. In a separate experiment aimed at estimating the binding constant, smaller ligand additions were used as equivalence was approached. The protein was titrated to 75% of the estimated number of binding sites, after which the amount of ligand was increased in increments of 5% of the estimated number of binding sites until the equivalence point was reached. All spectra were processed using the software NMRPIPE.[54] The dissociation constant K_d, defined as $K_d = (c_P - c_L + [L])[L]/(c_L - [L])$, was obtained by a nonlinear fitting of the [L] vs c_L titration curves (Figure 1A). Volume changes during titration were accounted for.

Equilibrium Dialysis. Oligosaccharide **52** was labeled using NaB^3H_4 (specific activity of 500 GBq/mmol) and was prepared as described previously[20] and diluted with the unlabeled compound according to the required specific activity. To set up equilibrium dialysis experiments, a rotating apparatus with glass blocks containing separate sealable chambers with external access was used as described previously (*5*). Then 200 μL aliquots of 0.1 μM solutions of CD69 proteins in 10 mM MES, pH 5.8, with 49 mM NaCl and 1 mM NaN_3 were incubated with varying amounts of ligand at 278 or 300 K for 48 h. After equilibration, 100 μL aliquots were withdrawn from the control and from the protein-containing chambers. The total ligand concentration was determined by liquid scintillation, and the bound ligand was calculated as the difference between the amount of GlcNAc in the chamber containing the protein and the control chamber. The results were calculated and plotted according to Scatchard as described previously.[11]

Tryptophan Fluorescence Quenching. Tryptophan fluorescence quenching experiments were performed according to the described methodology[51] with minor modifications. In initial experiments, 100 nmolar aliquots of CD69 protein were pipetted into multiple wells of a UV Star plate (Greiner, Germany) and mixed with 10-fold serial dilutions of the GlcNAc ligand. Incubation proceeded for 1 h at room temperature, and then the fluorescence of tryptophane residues was measured in duplicate wells using the bottom fluorescence measurements on a Safire2 plate reader (Tecan, Austria) with the following settings: λ_{ex} = 275 nm, λ_{em} = 350 nm, excitation and emission slits were set to 5 and 20, respectively, and the fluorescence gain was manually set to 66. When the lowest concentration of ligand that still caused the quenching of tryptophane fluorescence was found, detailed dilutions of the ligand by 10% saturation steps were performed, and the concentrations of free and bound ligand were calculated as described previously.[55]

Precipitation Assays with the Soluble CD69 Receptor. Each ligand was dissolved in water at 20 nM corresponding to 10-fold the K_d value. The ^{125}I-labeled protein (20 nM, 50 μL)[13] was

added to each sample (50 μL) in 96-well microtiter plates. Mixtures were incubated at 4 °C for 30 min, and then a 20% (v/v) solution of PEG 8000 was added (100 μL). The mixture was left to precipitate for 1 h at 4 °C. After centrifugation (10 min, 4 °C, 1800g_{av}), the supernatant was carefully removed and a 10% (v/v) solution of PEG 8000 (100 μL) was added. This procedure was repeated three times to wash the precipitate. After additional centrifugation and supernatant removal, the precipitates were dried overnight at 37 °C.

Preparation of $CD69^{low}$ and $CD69^{high}$ Lymphocytes. Peripheral blood mononuclear cells were obtained from standard blood fraction enriched in leukocytes (buffy coats from the local blood transfusion service) after dilution with RPMI1640 medium and centrifugation over Ficoll-Paque. Cells were incubated overnight in complete RPMI1640 in plastic cell culture dishes to allow the adherent cells to attach. Collected nonadherent fraction of PBMC (N-PBMC) contained mostly lymphocytes (T, B, and NK cells). Lymphocytes from donors expressing less than 5% of CD69 were designated $CD69^{low}$. Lymphocytes from donors with more then 20% CD69 positive cells were further activated by incubation at a density of 2×10^6 cells/mL in complete RPMI1640 medium for 4 h with PMA (50 ng/mL) and ionomycin (500 ng/mL). This procedure increased the surface expression of CD69 to 75−85%, as analyzed by flow cytometry using monoclonal antibody against CD69 labeled with phycoerythrin. Such lymphocytes were designated as $CD69^{high}$.

Inositol Phosphate Production. [^{3}H]Inositol phosphates were separated and quantified by the methods described previously.[11] Incorporation of [^{3}H]inositol into phospholipid was achieved by incubating human $CD69^{low}$ or $CD69^{high}$ lymphocytes obtained as described above (10^7 cells/ml) with 100 μL of [^{3}H]inositol (1.48 TBq/mol, 37 MBq/ml; GE Healthcare) for 3 h at 37 °C, followed by extensive washing, and resuspension at 10^6 cells/mL. An amount of 50 μL of this suspension containing 5×10^5 cells in complete RPMI with 10 mM Hepes, pH 7.4, was mixed with 50 μL of the tested compounds (60 nM), and the mixture was incubated at 37 °C for indicated times. Antibodies were added in saturating concentrations (10 μg/mL). Reaction was stopped by rapid transfer of the reaction mixture to 100 μL of 10% trichloroacetic acid. Reaction was neutralized by the addition of 50 μL of triethylamine, and 20 μL of 50% aqueous slurry of Dowex 1X8, 100−200 mesh (Sigma), in formate form was added. The supernatant was collected, and inositol bisphosphates and inositol trisphosphates were eluted by the addition of 50 μL of 0.3 and 0.6 M ammonium formate, pH 7.0, respectively. The eluant was dried in thin-walled 96-well plate, and the radioactivity was counted in 100 μL of biodegradable counting scintillant (GE Healthcare) using the Microbeta counter (Wallac).

Monitoring of Intracellular Calcium. Human $CD69^{low}$ or $CD69^{high}$ lymphocytes were loaded with the calcium-sensitive fluor Indo-1 by incubating 10^7 cells/mL with 5 pM Indo-1AM (Molecular Probes, Eugene, OR) in complete RPMI with 25 pM 2-ME at 37 °C. Cells were washed twice and resuspended at 5×10^6 cells/mL in medium. The fluorescence of the cell suspension was monitored with a Safire2 spectrofluorimeter by using an excitation wavelength of 349 nm and emission wavelength of 410 nm. The setting of the instrument was calibrated for each experiment by lysing the Indo-1-loaded cells with Triton X-100 (0.07%) for maximum fluorescence. The minimum fluorescence was determined after the addition of 10 mM EGTA and sufficient Tris base to raise the pH to >8.3. Intracellular calcium concentration was calculated using the formula $[Ca^{2+}]$ (nmoles) = $250[(F - F_{min})/(F_{max} - F)]$ where F is the measured fluorescence and 250 (nM) is the dissociation constant of Indo-1. Signaling was measured in the absence of extracellular calcium in medium containing 1 mM EGTA. Tested compounds were added in 2 min by rapid mixing of 10 μL of 300 nM solutions of these compounds with 90 μL of cellular suspensions, and monitoring continued for an additional 10 min. At 10 min after the beginning of the experiment (8 min after the addition of compounds), ionomycin (Sigma) was added to a final concentration of 1 μM. Antibodies were added in saturating concentrations (10 μg/mL).

Natural Killing. The standard ^{51}Cr release test was performed as described previously.[11] Briefly, 10^4 chromium-labeled target cells in 100 μL of complete RPMI 1640 were mixed in triplicate with 150 nM solutions of the tested compounds, or control compounds, in 50 μL of RPMI 1640 in round-bottomed 96-well plates. Antibodies were added in saturating concentrations (10 μg/mL). Thereafter, the appropriate amount of effector cells ($CD69^{high}$ lymphocytes obtained by Ficoll−Isopaque separation;[12] see above) was added to 100 μL of complete RPMI 1640, and the plate was incubated at 37 °C for 3 h. Then 50 mL of 1% Triton X-100 was added into the maxima release wells, and the incubation continued for another 1 h. Plates were cooled on ice bath, and 100 μL of the supernatant was used for radioactivity measurements. The percentage of specific lysis was calculated using the formula % = [(exp − spont)/(max − spont)] × 100, where exp is the counts in experimental wells, spont is the counts in wells containing medium instead of the effector cells, and max is the counts in wells containing 1% Triton X-100. Complete killing curves were constructed, from which the lytic unit counts for individual experiments were calculated. Lytic efficiency was defined as the inverse of the lytic unit count.

Animal Tumor Therapies. Young (6−8 week old) female C57BL/6 mice were purchased from Charles River (Montreal, Quebec, Canada) and handled under the guidelines of Institute Animal Care Protocol. Mice were shaved in the right flank area and were given injections sc with 2.5×10^4 viable B16F1 cells (low metastasis variant, ref 56) in a final volume of 100 μL of PBS. Ten days after the injection of tumor cells, 1 μmol of the tested compound, or 10 mg of mAb, was injected. Tumor growth was followed by Vernier caliper measurement every other day from day 7 after injection. All of the experiments included 10 mice/group. Tumor area was calculated according to the formula $A = (ab)/2$, where a is the largest superficial diameter and b is the smallest superficial diameter. The tumor infiltrating lymphocytes were isolated from animals at day 30,[45] and their cytolytic activity was measured as described above using the appropriate mouse tumor cells as targets.

Statistical Analyses. Statistical analyses were calculated by Student's t test. P values of ≤0.05 were considered as significant (*, $p \le 0.05$; **, $p \le 0.01$).

Acknowledgment. The authors thank to Markéta Vančurová for her helpful comments on the manuscript. This work was supported by grants from Ministry of Education of Czech Republic (Grants MSM_21620808, 1M0505, and AVOZ50200510 to K.B., Grant Z4 055 0506 to M.L., and Grants MSM0021622413 and LC06030 to V.S.), from Czech Grant Agency (Grants 303/09/0477 and 305/09/H008 to K.B. and Grant 203/09/P024 to P.P.), from Grant Agency of Czech Academy of Science (Grants KAN200520703 and 200100801 to M.L.), and by EU Project Spine 2 (Contract LSHG-CT-2006-02/220 to K.B.).

Supporting Information Available: Synthesis and analytical data of compounds **4**−**10**, **15**−**20**, **23**−**31**, **33**−**36**, **40**−**48**, **50**, **51**, **55**−**60** and analytical data for target compounds **21**, **32**, **37**, **49**, **52**−**54**, **61**, and **65**−**70**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Biron, C. A.; Nguyen, K. B.; Pien, G. C.; Cousens, L. P.; Salazar-Mather, T. P. Natural killer cells in antiviral defense: function and regulation by innate cytokines. *Annu. Rev. Immunol.* **1999**, *17*, 189–220.

(2) Moretta, A. The dialogue between human natural killer cells and dendritic cells. *Curr. Opin. Immunol.* **2005**, *17*, 306–311.
(3) Trinchieri, G. Biology of natural killer cells. *Adv. Immunol.* **1989**, *47*, 187–376.
(4) Bottino, C.; Castriconi, R.; Moretta, L.; Moretta, A. Cellular ligands of activating NK receptors. *Trends Immunol.* **2005**, *26*, 221–226.
(5) Moretta, A.; Bottino, C.; Vitale, M.; Pende, D.; Cantoni, C.; Mingari, M. C.; Biassoni, R.; Moretta, L. Activating receptors and coreceptors involved in human natural killer cell-mediated cytolysis. *Annu. Rev. Immunol.* **2001**, *19*, 197–223.
(6) Lanier, L. L. NK cell recognition. *Annu. Rev. Immunol.* **2005**, *23*, 225–274.
(7) Moretta, L.; Bottino, C.; Pende, D.; Vitale, M.; Mingari, M. C.; Moretta, A. Different checkpoints in human NK cell activation. *Trends Immunol.* **2004**, *25*, 670–676.
(8) Dwek, R. A. Glycobiology: toward understanding the function of sugars. *Chem. Rev.* **1996**, *96*, 683–720.
(9) Seeberger, P. H. Exploring life's sweet spot. *Nature* **2005**, *437*, 1239.
(10) Varki, A. Biological roles of oligosaccharides—all of the theories are correct. *Glycobiology* **1993**, *3*, 97–130.
(11) Bezouška, K.; Yuen, C. T.; Obrien, J.; Childs, R. A.; Chai, W. G.; Lawson, A. M.; Drbal, K.; Fišerová, A.; Pospíšil, M.; Feizi, T. Oligosaccharide ligands for NKr-P1 protein activate Nk Cells and cytotoxicity. *Nature* **1994**, *372*, 150–157.
(12) Pospíšil, M.; Vannucci, L.; Horváth, O.; Fišerová, A.; Krausová, K.; Bezouška, K.; Mosca, F. Cancer immunomodulation by carbohydrate ligands for the lymphocyte receptor NKR-P1. *Int. J. Oncol.* **2000**, *16*, 267–276.
(13) Bezouška, K.; Vlahas, G.; Horváth, O.; Jinochová, G.; Fišerová, A.; Giorda, R.; Chambers, W. H.; Feizi, T.; Pospíšil, M. Rat natural killer cell antigen, NKR-P1, related to C-type animal lectins is a carbohydrate-binding protein. *J. Biol. Chem.* **1994**, *269*, 16945–16952.
(14) Bezouška, K.; Nepovím, A.; Horváth, O.; Pospíšil, M.; Hamann, J.; Feizi, T. CD69 antigen of human lymphocytes is a calcium-dependent carbohydrate-binding protein. *Biochem. Biophys. Res. Commun.* **1995**, *208*, 68–74.
(15) Plihal, O.; Byrtusová, P.; Pavlíček, J.; Mihók, L.; Ettrich, R.; Man, P.; Pompach, P.; Havlíček, V.; Hušáková, L.; Bezouška, K. The isoforms of rat natural killer cell receptor NKR-P1 display a distinct binding of complex saccharide ligands. *Collect. Czech. Chem. Commun.* **2004**, *69*, 631–644.
(16) Pavlíček, J.; Kavan, D.; Pompach, P.; Novák, P.; Lukšan, O.; Bezouška, K. Lymphocyte activation receptors: new structural paradigms in group V of C-type animal lectins. *Biochem. Soc. Trans.* **2004**, *32*, 1124–1126.
(17) Bezouška, K. Carbohydrate and non-carbohydrate ligands for the C-type lectin-like receptors of natural killer cells. A review. *Collect. Czech. Chem. Commun.* **2004**, *69*, 535–563.
(18) Krist, P.; Herkommerová-Rejnochová, E.; Rauvolfová, J.; Semeňuk, T.; Vavrušková, P.; Pavlíček, J.; Bezouška, K.; Petrus, L.; Křen, V. Toward an optimal oligosaccharide ligand for rat natural killer cell activation receptor NKR-P1. *Biochem. Biophys. Res. Commun.* **2001**, *287*, 11–20.
(19) Rohlenová, A.; Ledvina, M.; Šaman, D.; Bezouška, K. Synthesis of linear and branched regioisomeric chitooligosaccharides as potential mimetics of natural oligosaccharide ligands of natural killer cells NKR-P1 and CD69 lectin receptors. *Collect. Czech. Chem. Commun.* **2004**, *69*, 1781–1804.
(20) Pavlíček, J.; Sopko, B.; Ettrich, R.; Kopecký, V., Jr.; Baumruk, V.; Man, P.; Havlíček, V.; Vrbacký, M.; Martinková, L.; Křen, V.; Pospíšil, M.; Bezouška, K. Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry* **2003**, *42*, 9295–9306.
(21) Eller, S.; Schuberth, R.; Gundel, G.; Seifert, J.; Unverzagt, C. Synthesis of pentaantennary *N*-glycans with bisecting GlcNAc and core fucose. *Angew. Chem., Int. Ed.* **2007**, *46*, 4173–4175.
(22) Bongat, A. F. G.; Demchenko, A. V. Recent trends in the synthesis of *O*-glycosides of 2-amino-2-deoxysugars. *Carbohydr. Res.* **2007**, *342*, 374–406.
(23) Holeman, A.; Seeberger, P. H. Carbohydrate diversity: synthesis of glycoconjugates and complex carbohydrates. *Curr. Opin. Biotechnol.* **2004**, *15*, 615–622.
(24) Bezouška, K.; Kavan, D.; Mrázek, H.; Manglová, D.; Bašová, P.; Křen, V. Glycomic analysis of tumor cells reveals the molecular determinants critical for their sensitivity for natural killing and apoptosis. *FEBS J.* **2008**, *275* (Suppl. 1), 81.
(25) Paulsen, H. Advances in selective chemical synthesis of complex oligosaccharides. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 155–173.
(26) Banoub, J.; Boullanger, P.; Lafont, D. Synthesis of oligosaccharides of 2-amino-2-deoxy sugars. *Chem. Rev.* **1992**, *92*, 1167–1195.
(27) Veselý, J.; Ledvina, M.; Jindřich, J.; Trnka, T.; Šaman, D. Synthesis of 2-amino-2-deoxy-β-D-galactopyranosyl(1 → 4)-2-amino-2-deoxy-β-D-galactopyranosides: using various 2-deoxy-2-phthalimido-D-galactopyranosyl donors and acceptors. *Collect. Czech. Chem. Commun.* **2004**, *69*, 1914–1938.
(28) Farkaš, J.; Ledvina, M.; Brokeš, J.; Ježek, J.; Zajíček, J.; Zaoral, M. The synthesis of *O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1→4)-*N*-acetylnormuramoyl-(-α-aminobutanoyl-D-isoglutamine. *Carbohydr. Res.* **1987**, *163*, 63–72.
(29) Ledvina, M.; Zyka, D.; Ježek, J.; Trnka, T.; Šaman, D. New effective synthesis of (*N*-acetyl- and *N*-stearoyl-2-amino-2-deoxy-β-D-glucopyranosyl)-(1→4)-muramoyl-L-2-aminobutanoyl-D-isogluta-mine. Analogs of GMDP with immunopotentiating activity. *Collect. Czech. Chem. Commun.* **1998**, *63*, 577–589.
(30) Wilstermann, M.; Kononov, L. O.; Nilsson, U.; Ray, A. K.; Magnusson, G. Synthesis of ganglioside lactams corresponding to G(M1)-ganglioside, G(M2)-ganglioside, G(M3)-ganglioside, and G(M4)-ganglioside lactones. *J. Am. Chem. Soc.* **1995**, *117*, 4742–4754.
(31) Hansson, J.; Garegg, P. J.; Oscarson, S. Syntheses of anomerically phosphodiester-linked oligomers of the repeating units of the *Haemophilus influenzae* types c and f capsular polysaccharides. *J. Org. Chem.* **2001**, *66*, 6234–6243.
(32) Berkin, A.; Szarek, W. A.; Kisilevsky, R. Synthesis of 4-deoxy-4-fluoro analogues of 2-acetamido-2-deoxy-D-glucose and 2-acetamido-2-deoxy-D-galactose and their effects on cellular glycosaminoglycan biosynthesis. *Carbohydr. Res.* **2000**, *326*, 250–263.
(33) Westerlind, U.; Hagback, P.; Duk, M.; Norberg, T. Synthesis and inhibitory activity of a di- and a trisaccharide corresponding to an erythrocyte glycolipid responsible for the NOR polyagglutination. *Carbohydr. Res.* **2002**, *337*, 1517–1522.
(34) Nilsson, U.; Ray, A. K.; Magnusson, G. Efficient syntheses of 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-galactopyranosyl and 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-α-D-galactopyranosyl chloride. *Carbohydr. Res.* **1990**, *208*, 260–263.
(35) Veselý, J.; Ledvina, M.; Jindřich, J.; Šaman, D.; Trnka, T. Improved synthesis of 1,2-trans-acetates and 1,2-trans ethyl 1-thioglycosides derived from 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-D-hexopyranosides. *Collect. Czech. Chem. Commun.* **2003**, *68*, 1264–1274.
(36) Crich, D.; Dudkin, V. Why are the hydroxy groups of partially protected *N*-acetylglucosamine derivatives such poor glycosyl acceptors, and what can be done about it? A comparative study of the reactivity of *N*-acetyl-, *N*-phthalimido-, and 2-azido-2-deoxy-glucosamine derivatives in glycosylation. 2-Picolinyl ethers as reactivity-enhancing replacements for benzyl ethers. *J. Am. Chem. Soc.* **2001**, *123*, 6819–6825.
(37) Fowler, P.; Bernet, B.; Vasella, A. A ^{1}H-NMR spectroscopic investigation of the conformation of the acetamido group in some derivatives of *N*-acetyl-D-allosamine and -D-glucosamine. *Helv. Chim. Acta* **1996**, *79*, 269–287.
(38) Vasella, A.; Witzig, C. Glycosylidene carbenes. Part 22. α-D-Selectivity in the glycosidation by carbenes derived from 2-acetamido-hexoses. *Helv. Chim. Acta* **1995**, *78*, 1971–1982.
(39) Kantoci, D.; Keglevic, D.; Derome, A. E. A convenient synthetic route to the disaccharide repeating-unit of peptidoglycan. *Carbohydr. Res.* **1987**, *162*, 227–235.
(40) Maloisel, J. L.; Vasella, A. Synthesis of *N*-acetylallosamine-derived disaccharides. *Helv. Chim. Acta* **1992**, *75*, 1491–1514.
(41) Kavan, D.; Kubíčková, M.; Bálý, J.; Vaněk, O.; Hofbauerová, K.; Mrázek, H.; Razbeský, D.; Bojarová, P.; Křen, V.; Zídek, L.; Sklenář, V.; Bezouška, K. Cooperation between subunits is essential for high affinity binding of *N*-acetyl-hexosamines to dimeric soluble and dimeric cellular forms of human CD69. *Biochemistry* [Online early access]. DOI: 10.1021/bi100181a. Published Online: April 6, 2010.
(42) Vaněk, O.; Nálezková, M.; Kavan, D.; Borovičková, I.; Pompach, P.; Novák, P.; Kumar, V.; Vannucci, L.; Hudeček, J.; Hofbauerová, K.; Kopecký, V., Jr.; Brynda, J.; Kolenko, P.; Dohnálek, J.; Kaderávek, P.; Gorčík, L.; Žídek, L.; Sklenář, V.; Bezouška, K. Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice. *FEBS J.* **2008**, *275*, 5589–5606.
(43) Semeňuk, T.; Krist, P.; Pavlíček, J.; Bezouška, K.; Kuzma, M.; Novák, P.; Křen, V. Synthesis of chitooligomer-based glycoconjugates and their binding to rat natural killer cell activation receptor NKR-P1. *Glycoconjugate J.* **2001**, *18*, 817–826.
(44) Chung, S.; Parker, J. B.; Bianchet, M.; Amzel, L. M.; Stivers, J. T. Impact of linker strain and flexibility in the design of a fragment-based inhibitor. *Nat. Chem. Biol.* **2009**, *6*, 407–413.
(45) Bojarová, P.; Křenek, K.; Wetjen, K.; Adamiak, K.; Pelantová, H.; Bezouška, K.; Elling, L.; Křen, V. Synthesis of LacdiNAc-termi-

nated glycoconjugates by mutant glycosyltransferases. A way to new glycodrugs and materials. *Glycobiology* **2009**, *19*, 509–517.
(46) Gerosa, F.; Tommasi, M.; Scardoni, M.; Accolla, R. S.; Pozzan, T.; Libonati, M.; Tridente, G.; Carra, G. Structural analysis of the CD69 early activation antigen by two monoclonal antibodies directed to different epitopes. *Mol. Immunol.* **1991**, *28*, 159–168.
(47) Vannucci, L.; Fišerová, A.; Sadalapure, K.; Lindhorst, T. K.; Kuldová, M.; Rossman, P.; Horváth, O.; Křen, V.; Krist, P.; Bezouška, K.; Luptovcová, M.; Mosca, F.; Pospíšil, M. Effect of *N*-acetylglucosamine-coated glycodendrimers as biological modulators in the B16F10 melanoma model in vivo. *Int. J. Oncol.* **2003**, *23*, 285–296.
(48) Testi, R.; D'Ambrosio, D.; De Maria, R.; Santoni, A. The CD69 receptor: a multipurpose cell surface trigger for hematopoietic cells. *Immunol. Today* **1994**, *15*, 479–483.
(49) Sancho, D.; Gomez, M.; Sanchez-Madrid, F. CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol.* **2004**, *26*, 136–140.
(50) Moretta, A.; Poggi, A.; Pende, D.; Tripodi, G.; Orengo, A. M.; Pella, N.; Augugliaro, R.; Bottino, C.; Ciccone, E.; Moretta, L. CD69-mediated pathway of lymphocyte activation: anti-CD69 monoclonal antibodies trigger the cytolytic activity of different lymphoid effector cells with the exception of cytolytic T lymphocytes expressing T cell receptor α/β. *J. Exp. Med.* **1991**, *174*, 1393–1398.
(51) Gronberg, G.; Nilsson, U.; Bock, K.; Magnusson, G. Nuclear-magnetic-resonance and conformational investigations of the pentasaccharide of the forssman antigen and overlapping disaccharide, trisaccharide, and tetrasaccharide sequences. *Carbohydr. Res.* **1994**, *257*, 35–54.
(52) Likhosherstov, L. M.; Novikova, O. S.; Derevitskaya, V. A.; Kochetkov, N. K. A new simple synthesis of amino sugar β-o-glycosylamines. *Carbohydr. Res.* **1986**, *146*, C1–C5.
(53) Lindhorst, T. K.; Kieburg, C. Glycocoating of oligovalent amines. Synthesis of thiourea-bridged cluster glycosides from glycosylisothiocyanates. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1953–1956.
(54) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.
(55) Chipman, D. M.; Grisaro, V.; Sharon, N. The binding of oligosaccharides containing *N*-acetylglucosamine and *N*-acetylmuramic acid to lysozyme. *J. Biol. Chem.* **1967**, *242*, 4388–4394.
(56) Poste, G.; Doll, J.; Hart, I. R.; Fidler, I. J. In vitro selection of murine B16 melanoma variants with enhanced tissue-invasive properties. *Cancer Res.* **1980**, *40*, 1636–1644.

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 *Aspergillus 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*.