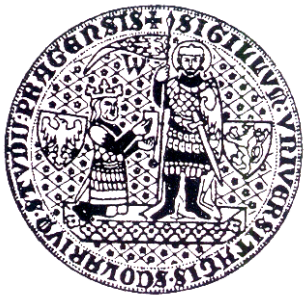
Charles University in Prague
1st Faculty of Medicine
Institute of Inherited Metabolic Disorders



IS *Caenorhabditis elegans* A SUITABLE MODEL ORGANISM FOR THE STUDY OF HUMAN LYSOSOMAL ENZYMOPATHIES?
A STUDY OF FABRY, SCHINDLER, POMPE AND MUCOPOLYSACCHARIDOSIS IIIC DISEASES IN *C. elegans* AND HUMANS, RESPECTIVELY.

Ph.D. thesis

Jana Uřinovská

Supervisor: Prof. MUDr. Milan Elleder, DrSc.

Consultants: MUDr. Martin Hřebíček

MUDr. Jakub Sikora, Ph.D.

Prague 2008

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. MUDr. Milan Elleder, DrSc. and consultants MUDr. Martin Hřebíček and MUDr. Jakub Sikora, Ph.D. for their scientific approach, guidance, understanding, and for the opportunity to perform and finish my Ph.D. studies at the Institute of Inherited Metabolic Disorders.

I would also like to thank MUDr. Marta Kostrouchová, CSc. for her advice and the laboratory support for *C. elegans* experiments.

I express my thanks to MUDr. Jakub Sikora, Ph.D., RNDr. Robert Dobrovolný, Ph.D., Ing. Filip Majer and RNDr. Karel Jelínek, Ph.D. for phylogenetic analyses and homology modeling and to RNDr. Dusan Cmarko, Ph.D. for performing immunogold electron microscopy.

I thank all my colleagues at the Institute, namely to Jana Sovová, Ing. Jitka Hlavatá, Ing. Helena Poupětová, RNDr. Jana Ledvinová, CSc, Mgr. Eva Brožová, Ph.D., Mgr. Katka Šimečková, Ph.D. and Lenka Brtvová, for their help and friendship.

Finally, I wish to thank my family for their love and permanent support.

This work was supported from grants 303/02/1324 and 303/03/H065 from the Grant Agency of the Czech Republic and from research projects VZ111100003 and 0021620806 from Ministry of Education, Youth and Sports of the Czech Republic.

Contents:

2. Introduction

2.1. *Caenorhabditis elegans* - a model organism

2.1.1. Main features

Caenorhabditis elegans (C. elegans) is a small (~1 mm long), free-living, soil nematode feeding primarily on bacteria. *C. elegans* has a constant number of somatic cells. Adult hermaphrodites and males have 959 and 1031 somatic nuclei, respectively. Full cell lineage is described in *C. elegans.*

Worms are reproducing with a life cycle of about 3 days under optimal environmental conditions. Mature adults are fertile for 4 days and their overall life span is about 17 days after reaching adulthood. *C. elegans* has five holocentric autosomes and sex chromosome (X). Sex is determinated chromosomally. Hermaphrodites are diploid for all autosomes and X chromosomes (XX). Males, on the contrary have only one copy of X chromosome (XO). Recombination occurs in males´ sperm and in both sperm and oocytes in hermaphrodites. Males arise by spontaneous non-disjunction of sex chromosomes during meiosis in hermaphrodite germ line with frequency of about 1:500 worms [1-3].

C. elegans can be easily cultivated under laboratory conditions and enables easy genetic manipulation. Self-fertilization of the hermaphrodites allows for homozygous worms to generate genetically identical progeny and male mating facilitates the isolation and maintenance of mutant strains as well as propagation of mutations between strains. Mutant animals can be obtained by chemical mutagenesis or exposure to ionizing radiation [4, 5]. The strains can be kept as frozen stocks for long periods of time. Due to these properties and its simplicity *C. elegans* became a useful model organism for concerted genetic, ultrastructural, behavioral and developmental studies in eukaryotic organisms [1, 2, 6]. It was the first of the multicellular eukaryotic organisms for which complete genomic sequence became known. About thirty-six percent of genes in *C. elegans* have homology to human genes, including those involved in different human pathology states [7].

RNA-mediated interference (RNAi) technique allows to inhibit *C. elegans* genes very easily [8, 9]. High throughput RNAi techniques were used for inhibition of virtually all *C. elegans* genes [10-14]. Besides RNAi, it is relatively easy to generate and observe the expression of transgenes, either with fluorescent or histochemical molecular tags, in the transparent body of the worm [15, 16].

2.1.2. Life cycle

C. elegans develops from embryo (558 cells stage) through four larval stages (L1, L2, L3 and L4) into adult (Figure 1) within three days. Normally the eggs hatch into larvae after twelve hours. All larval stages are interrupted by molts during which a new cuticle is synthesized and the old one is shed. Dauer larvae (dauers) represent a specific larval stage that develops instead of the normal L3 larvae under unfavorable conditions. Dauers can survive several months and than molt into normal L4 larvae under optimal conditions [17, 18]. Hermaphrodites lay about 300 eggs during their fertile period.

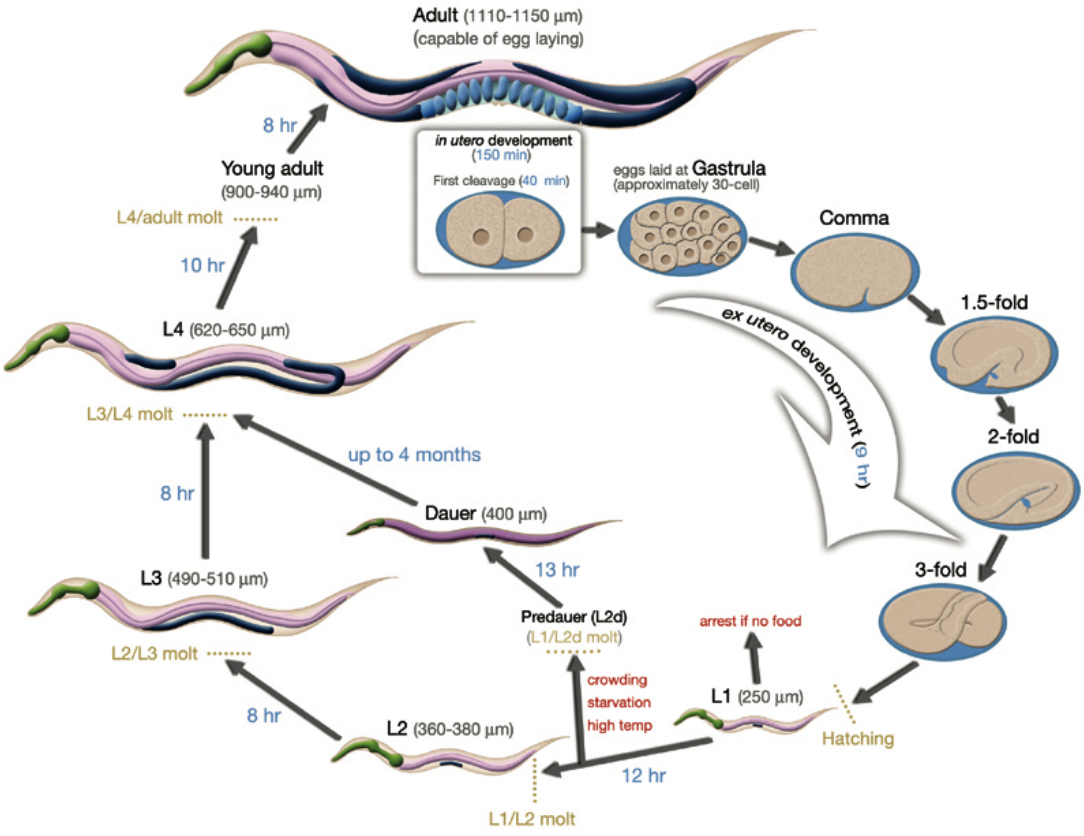


Figure 1: Life cycle of *C. elegans* (at 22 °C)**.**

The picture shows all embryonic and larval stages including adult hermafrodite. The length of time that the animal spends at a certain stage is marked in blue and the length of the animal at certain stage is stated in grey in round brackets.
The image was downloaded from publically available www domain Wormatlas [19] http://www.wormatlas.org/handbook/anatomyintro/anatomyintro.htm

2.1.3. Development

2.1.3.1. Embryogenesis

Embryogenesis in *C. elegans* has two distinct periods: proliferation and organogenesis/morphogenesis [20].

The proliferation stage comprises of multiple mitotic cell divisions which result in formation of the first larval stage [6]. Initial cell divisions generate six distinct cells that are called founder cells: AB, E, MS, C, D, and P_4. Each founder cell goes through a series of synchronous and symmetrical divisions to generate a predetermined number of cells [21].

During organogenesis/morphogenesis terminal differentiation of cells occurs and the general body plan, formed at the end of the embryogenesis is attained. During elongation embryos take form of fully differentiated animals, tissues and organs included. Embryonal elongation includes several morphological forms: the comma, 1.5-fold, 2-fold and 3-fold stage (Figure 1). L1 cuticle is formed in 3-fold stage [20].

2.1.3.2. Post-embryonic development

Post-embryonic development is initiated by feeding of larvae after hatching. In the presence of food, cell divisions continue and postembryonic developmental program begins 3 hours after hatching [22]. The animals normally pass through four larval stages to reach adulthood (Figure 1). Germ line proliferates and mature gonads are formed during L4 stage. The number of somatic cell nuclei increases to 959 in hermaphrodites and 1031 in males [18].

If the embryos hatch in the absence of nutrients the development may get arrested until environmental conditions improve [23, 24].

2.1.4. Genetics

2.1.4.1. The genome

C. elegans genome consists approximately of 97 million base pairs (one-thirtieth the size of the human genome), including 15 kb of mitochondrial DNA. The complete genomic sequence was published in December 1998 by the international *C. elegans* Genome Sequencing Consortium [7]. All data about *C. elegans* are assembled and available to scientific community in databases such as *Caenorhabditis elegans* WWW Server [25], WormBook [26], ORFeome project database WorfDB [27, 28], WormAtlas [19] and others.

The central *C. elegans* database is WormBase, initiated in the year 2001 and continuously updated every 21 days [29]. It´s contents doubled in the last three years [30-33].

C. elegans genome contains 23 693 confirmed or predicted protein coding genes [34]. The worm genome is very compact with one gene per 5kb on average. The genome is predicted to contain only 27 % of intronic sequences and 27 % putative exons [7]. Many of the predicted protein products have homologs in other organism. The average predicted protein similarity between *C. elegans* and *Homo sapiens* is 36 %, which is more than with any other multicellular organism (except other nematodes). Many genes show similarity as high as 97 % [7, 35, 36].

The number of splicing isoforms per gene is low. More than 90 % of alternatively spliced genes have only one or two isoforms [37], the overall current number of alternatively spliced isoforms is 3515 [34].

There are two types of pseudogenes in *C. elegans* genome – processed and unprocessed. The first type of pseudogenes results from reverse transcription of mRNA into cDNA and its reinsertion into gDNA. Unprocessed pseudogenes are formed by gene duplication and the following loss of function due to nucleotide substitutions, deletions, insertions, frameshifts or other DNA alterations [38].

2.1.4.2. Operons and *trans*-splicing

C. elegans and other nematodes have approximately 15 % of genes organized into operons, a property which makes them distinct from other higher organisms. Operons may contain from 2 to 8 genes [39, 40]. The genes in the polycistronic operons are usually separated by tens of base pairs. Polycistronic RNAs are processed into monocistronic mRNAs by internal cleavage and polyadenylation of their 3´ends. The 5´end of the downstream mRNA is simultaneously formed by *trans*-splicing [41]. During the *trans*-splicing, a short RNA leader sequence is attached to the 5´end of mRNA. There are two types of splice leader RNAs – SL-1 and 2. SL1 is predominantly *trans*-spliced to 5´ends of monocistronic genes and first genes in operons. On the contrary, SL2 is *trans*-spliced to downstream genes in the operons [42, 43].

2.1.5. Anatomy

The anatomy of *C. elegans* is very simple. Worms can be observed by phase contrast light microscopy techniques, especially efficient is Nomarski differential interference contrast

(DIC). The anatomical description was completed at the electron microscopy level in the last two decades [6, 19, 20, 44].

2.1.5.1. Basic anatomy

Worms have unsegmented, tubular body that is tapered at the ends and has outer and an inner tube, which are separated by the pseudocoelomic space. The outer tube (body wall) consists of cuticle, hypodermis, excretory system, neurons and body wall muscles. The inner tube comprises of pharynx, intestine and gonad (Figure 2).

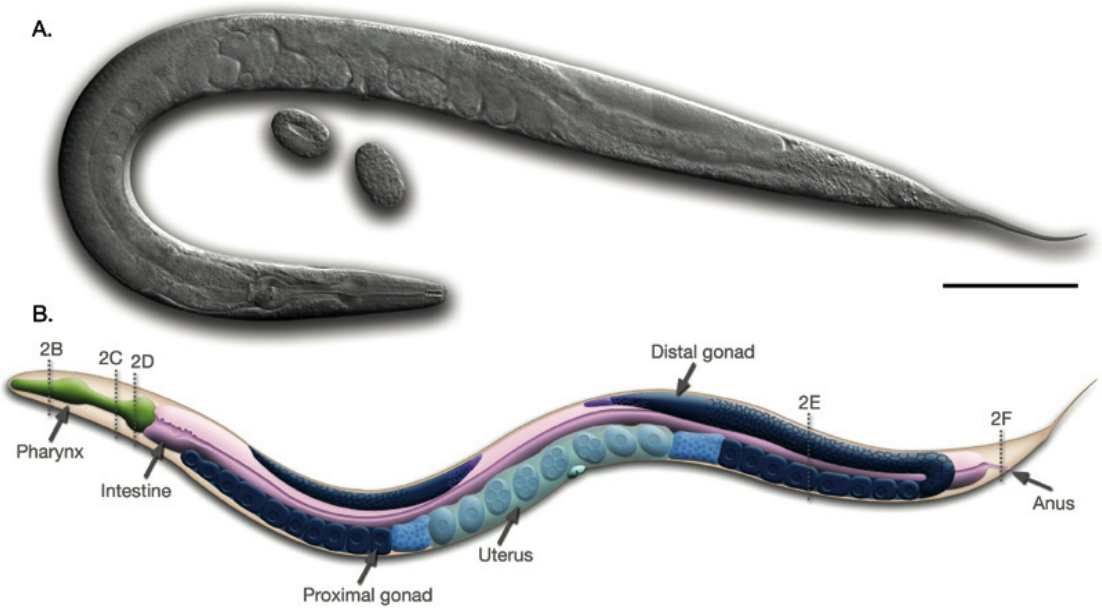


Figure 2: Anatomy of an adult hermaphrodite

A. DIC image (scale bar 0,1 mm) and **B.** schematic drawing.
The image was downloaded from publically available www domain Wormatlas [19] http://www.wormatlas.org/handbook/anatomyintro/anatomyintro.htm

2.1.5.1.1. Cuticle and Epidermis

An external transparent cuticle, which is composed of collagens, covers the body wall. Collagens are secreted by polarized epidermal cells. Many of epidermal cells are multinucleate, arising by cell fusions during development [6].

2.1.5.1.2. The body wall muscle system

Four longitudinal rows of body wall muscle cells are situated under the epidermis and are located subventrally and subdorsally. Individual cells do not form syncitia, which are typical for vertebrates, but are separated and mononucleated. A thin basal lamina separates the

musculature from the epidermis and nervous tissue. The functional contractile unit is analogous to the vertebrates' sarcomere. Thick filaments contain myosin and overlap with two sets of thin actin filaments [3, 45].

2.1.5.1.3. The nervous system

The nervous system in *C. elegans* has 302 neurons in the adult stage and consists of a circumpharyngeal nerve ring, dorsal and ventral nerve cords and a variety of sensory receptors and ganglia. The majority of *C. elegans* neurons are localized in the head around pharynx. The processes from most neurons travel in either the ventral or dorsal cord and project to the nerve ring in the head [6].

2.1.5.1.4. The alimentary system

The gastrointestinal system is formed by a single tube which contains pharynx, intestine, rectum and anus. It also includes related tissues such as valves, glands, muscles.

The mouth opens into the bilobed muscular pharynx, which grinds and pumps food through the tubular intestine in the direction of rectum and anus. *C. elegans* intestine is a tube formed by 20 large epithelial cells. These cells are organized in bilaterally symmetric pairs around the lumen. The whole intestine is made of nine rings. Intestinal cells are mononuclear at hatching; and become larger, binucleated and polyploid during postembryonic development [46]. Intestinal cells secrete digestive enzymes into the lumen, absorb nutrients and serve as storage depots [6]. Some intestinal granules are strongly autofluorescent when irradiated with 300 to 400 nm light. These granules are visible early in the nematode´s development and can serve as cell-type-specific markers [6].

2.1.5.1.5. The reproductive system

The reproductive system of adult hermaphrodites consists of two symmetrically arranged, U-shaped gonad arms. Each arm contains a distal ovary, proximal oviduct, spermatheca and leads into the uterus. Germ line nuclei are produced in ovarial syncytium. Germ cells sequentially proceed through the mitosis and meiotic prophase to enter diakinesis in the oviduct immediately prior to fertilization. Individual nuclei become enclosed by plasma membrane and form oocytes, which enlarge and mature as they pass through the oviducts. Both oviducts terminate at the spermatheca carrying about 150 sperms. The fertilized eggs (the zygotes) are stored in the uterus before laid via the vulva [6, 47].

The reproductive system of adult males consists of a single testis, which is connected with rectum by vas deferens. Vas deferens opens to the body exterior at the cloaca with associated sensory spicules. The males use the spicules during mating to locate the hermaphrodite vulva and to hold it open during sperm transfer [48, 49].

2.1.5.2. The coelomocyte system

C. elegans has three pairs of coelomocytes located in the pseudocoelomic cavity (ventral anterior, ventral posterior and dorsal). Coelomocytes contain large distended rough endoplasmic reticulum and a number of vacuoles of various sizes [50]. Each coelomocyte is covered by its own basal lamina.

Coelomocytes can uptake fluid-phase markers such as india ink, rhodamine-dextran and fluorescein isothiocyanate (FITC)-BSA from the pseudocoelom [50]. The coelomocytes were suggested to act as scavenger cells due to their ability to actively endocytose fluid from the pseudocoelom [51]. It is hypothesized that the coelomocytes may represent a primitive immune system in *C. elegans*, nevertheless their function is not essential for the animal's survival or fertility. Coelomocytes represent suitable cell type for studies of endosomal-lysosomal system [36, 50, 51].

2.1.6. The lysosomal system and lysosomal proteins in *C. elegans*

C. elegans has a fully developed lysosomal system and the lysosomes take up exogenous macromolecules in the same way as lysosomes in mammalian cells [36, 51-60]. Kostich and co-workers identified LMP-1 protein in *C. elegans*, with sequence and presumed structural similarity to mammalian LAMPs (lysosomal associated membrane proteins) and CD68 protein [61]. LMP-1 is the only nematode´s membrane protein with the vertebrate lysosomal targeting sequence at its C terminus. *C. elegans* possesses a variety of proteases and peptidases. The representatives of lysosomal acid glycosidases and phosphatases in *C. elegans* are comparable to that found in the mammalian cells [36, 62-64]. They belong to the same or related protein families and share similar protein structure. The presence of simple and complex glycoconjugates (glycolipids and glycoproteins) containing various sugar substrates has been reported [65, 66] as well as the presence of glycosaminoglycans (GAGs) [67-69]. All these compounds are degraded in the worm's endosomal – lysosomal system.

C. elegans requires cholesterol for proper growth and development. However, it is incapable to synthesize its own cholesterol and must obtain it from the diet [70, 71]. Intracellular cholesterol trafficking in C. elegans is directly dependent on the proper function

of late endosomal/lysosomal membrane protein(s), which are homologous to human NPC1 protein [72] (see section 2.1.8).

2.1.7. RNA-mediated interference (RNAi) technique and *C. elegans*

In 2006, Andrew Fire and Craig Mello obtained the Nobel Prize for discovery of double-stranded RNA triggered suppression of gene activity in a homology dependant manner in *C. elegans* – a phenomenon designated as RNA interference [73]. The molecular mechanism of RNAi is under continuous study and has been a subject of numerous review articles [74, 75]. The delivery of double-stranded RNA (dsRNA) results in the specific and potent inactivation of genes containing homologous sequences [8, 9, 76]. RNAi may be transmitted to the progeny of F1-F2 generation. The inactivation of genes results from the degradation of their endogenous mRNA. DsRNA is capable to cross cell membranes and intercellular environment. In *C. elegans* it is possible to initiate RNAi by injection of dsRNA into germ syncytium, by soaking the worms in dsRNA solution or by feedding them with *Escherichia coli* expressing dsRNA [77].

2.1.8. *C. elegans* as a model organism for human lysosomal storage disorders

C. elegans was repeatedly shown to be a suitable model for many human diseases [78-82], including lysosomal storage disorders such as Niemann-Pick type C disease [72] and juvenile neuronal ceroid lipofuscinosis [83]. Many human genes associated with lysosomal function and/or involved in LSDs have *C. elegans* orthologs [36]. It has been reported that *C. elegans* genome includes two distinct acid sphingomyelinase (ASM) genes [84], contrary to humans with only single ASM. Mutant strain of C. *elegans* with deficiency of the aspartyl protease cathepsin D has been described [63]. Number of works concentrated on *C. elegans* ortholog (CUP-5) of human mucolipin-1 [50, 57, 58, 85, 86]. Mucolipin-1 is a non-selective cation channel in the lysosomal membranes that is modulated by pH and its deficiency causes lysosomal storage disorder - mucolipidosis type IV. CUP-5 is essential for lysosomal biogenesis in *C. elegans*. In addition, the functions of LAMP/CD68 like protein in *C. elegans* including its deficiency [61] were thoroughly evaluated. Novel information acquired by studies of *C. elegans* orthologs of these non-catalytic human lysosomal proteins provided important insights into biological function of these proteins.

The amount of data about the functions of luminal lysosomal hydrolases in *C. elegans* is restricted to sparse reports, some of them are appended to this thesis (appended publication 1 and submitted manuscript).

In conclusion, *C. elegans* is a relevant model organism because it is a simple multicellular organism with short life cycle, it is inexpensive to cultivate and easy to genetically manipulate, its entire cell lineage as well as genome sequence is completely known. Sexual dimorphism of *C. elegans* facilitates genetic manipulation because self-fertilizing hermaphrodites can be used to maintain homogenous strain progeny and males enable genetic crosses. Long-term storage of nematode strains is possible because of the ability to freeze and recover the animals. Its use is generally advantageous due to the simplicity of the organism; on the other hand, it does not allow studies of complex tissue systems.

2.2. Lysosomal biology

2.2.1. Introduction

Christian de Duve used the term *"lysosome"* for the first time in the year 1955 [87]. Lysosomes were initially discovered by cell fractionation assays. Lysosomal morphology was characterized by electron microscopy shortly after [88].

Lysosomes are membrane-bound organelles which are found in all mammalian cells apart from red blood cells. Lysosomes are responsible for the controlled intracellular digestion of macromolecules. This dynamic cellular compartment contains more than 60 luminal hydrolytic enzymes, including proteases, nucleases, glycosidases, lipases, phospholipases, phosphatases and sulfatases. All hydrolases are functioning at acidic pH, which is maintained by vacuolar-type H^+-ATPase and Cl^- channel protein in the lysosomal membrane [89, 90]. The proteome of lysosomal membranes (130-150 different proteins) has been characterized in the year 2007 by Schroder et. al [91] by high throughput protein analysis techniques. The most extensively studied lysosomal membrane proteins are highly glycosylated lysosomal associated membrane proteins (LAMPs) [92-94] and lysosomal integral membrane proteins (LIMPs) [95-97]. High level of glycosylation may protect the lysosomal membrane proteins from the lysosomal proteases in the lumen. Lysosomes are morphologically very heterogeneous [98, 99].

Lysosomes are only an integral part of extremely dynamic endosomal–lysosomal system, which is composed of continually communicating membrane bound vesicles involved in vectorial and regulated cargo transport and recycling.

The characteristics of lysosomes (acidic pH, presence of luminal acid hydrolases and LAMPs and absence of mannose-6-phosphate receptor – M6PR) are shared with a group of cell type specific organelles, which are called lysosome – related organelles (LRO). LROs include melanosomes of melanocytes, lytic granules of T lymphocytes, platelet – dense granules, basophil and azurophil granules of granulocytes. Common pathology of both lysosomes and LROs represents cellular basis for diseases such as Chediak-Higashi, Heřmanský-Pudlák or Griscelli syndromes [100].

2.2.2. Delivery of macromolecules to degradation in lysosomes

There are at least three pathways for entry of substrates to lysosomes – endocytosis, autophagocytosis and phagocytosis. Their individual utilization may be cell type specific.

Endocytosis is general term used for the internalization of extracellular fluid or particles by invagination and pinching off of the plasma membrane [101, 102]. Every eukaryotic cell is capable of endocytosis. Endocytosed molecules are delivered in coated vesicles to small intracellular organelles - early endosomes (EE), which are slightly acidic (pH 6.0 – 6.2). In case of receptor mediated endocytosis, the pH sensitive ligand – receptor complexes dissociate and the receptors are recycled to the plasma membrane. From the EE the substrates pass on into late endosomes (LE). LE is the place where endocytosed substrates meet the lysosomal hydrolases and where the hydrolytic digestion starts. The interior of the LE is more acidic than EE (pH 5.5 – 6.0). Lysosomes are thought to be produced by maturation process from LE and their pH is about 4.5 - 5.5 [90]. The low pH is critical for optimal enzymatic function of luminal hydrolases.

Autophagy as well as endocytosis occurs in all cell types. Autophagy represents a complex system of signalling and executive pathways that provide the cell with disposal mechanism of intracellular contents (proteins, membranes, organelles). The autophagic process can be activated either by stress conditions such as starvation or occurs as constitutive or regulative process. There are currently three recognised molecular variants of autophagy [103, 104]: macroautophagy/lysosomal system (MALS), microautophagy and chaperone – mediated autophagy (CMA). MALS designates a process when an impaired organelle(s) or portion of cytoplasm is delimited by autophagosomal membrane [105, 106]. The exact origin of matrix membrane for autophagosome constitution still remains unresolved, nevertheless its proteome has been defined in mammalian cells [107]. At the end of the process the autophagosomes fuse with LE and the degradation of the contents by lysosomal hydrolases begins. The hydrolytic products are recycled by a combination of passive diffusion and

specific transporters back into the cytosol where they can be reutilized [108]. Microautophagy and chaperone – mediated autophagy degrade cytosolic proteins by two different mechanisms. Microautophagy is a process, in which the lysosomal membrane itself engulfs a part of cytoplasm or small organelles such as peroxisomes (pexophagy). Chaperone – mediated autophagy (CMA) [109, 110] is, on the contrary, highly selective, because it requires a targeting sequence signal (KFERQ-like motifs) [108, 111] in the substrate protein. Molecular chaperones in the cytosol and in the lysosomal lumen (hsc70 and ly-hsc70) [112, 113] and a lysosomal membrane receptor system (LAMP2a isoform of LAMP2 protein) [114, 115] are essential for the direct transport of substrate proteins across the lysosomal membrane. CMA is activated under stress conditions, such as nutrient deprivation, oxidative stress and exposure to toxic compounds.

The last major entry pathway into lysosomes is *phagocytosis* for degradation of extracellular material such as microorganisms and dead cells. Phagocytosis is a process restricted to specialized immune system related cells such as macrophages, neutrophils and dendritic cells. The phagosomes fuse with LE or lysosomes and the ingested material is then degraded [116].

2.2.3. Transport of lysosomal proteins to lysosomes

Lysosomal hydrolases and membrane proteins are synthesized in the rough endoplasmic reticulum (ER) and are transported through the Golgi apparatus to the *trans* Golgi network (TGN). The transport vesicles arise by budding off from the TGN and deliver these proteins to late endosomes, which serve as molecular sorters in the endosomal-lysosomal system.

Most of the soluble lysosomal enzymes are synthesized with N-linked high mannose-type oligosaccharides [117]. The proteins move by vesicular transport from the rough ER to the Golgi apparatus where they go through various post-translational modifications. The most important modification for lysosomal targeting and sorting is the formation of the mannose-6-phosphate (M6P) recognition marker. The M6P recognition signal is generated in a two step reaction [118, 119]. In the first step, the lysosomal enzymes are phosphorylated by UDP-*N*-acetylglucosamine: lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase (phosphotransferase). This enzyme transfers *N*-acetylglucosamine-1-phosphate from UDP-*N*-acetylglucosamine to the C6-hydroxyl group of a selected mannose residue on the high mannose-type oligosaccharide. In the following step, *N*-acetylglucosamine residues are cleaved off by the enzyme *N*-acetylglucosamine-1-phosphodiester α-*N*-acetylglucosaminidase

(phosphoglycosidase). The second reaction exposes the M6P recognition marker and allows the lysosomal enzyme to be recognised by mannose-6-phosphate receptors (MPRs). Because most lysosomal hydrolases contain multiple oligosaccharides, they may acquire more than one M6P residue, which thus provides a higher affinity signal for the MPRs.

There are two types of MPRs in mammalian cells [120-122]. Both receptors are membrane–associated glycoproteins and show similar, but not identical, binding specifities toward phosphorylated oligosaccharides. The first MPR has a M_r of 215,000 and its binding of ligand is independent of divalent cations (cation - independent MPR, CI-MPR). The second MPR has M_r of 46,000 and requires divalent cations for optimal ligand binding (cation – dependent MPR, CD-MPR). These two receptors further differ in their subcellular localization, quarternary structure, half life and in the effect of pH on ligand binding. Both bind ligand optimally at pH 6.0 - 6.3 (pH in the TGN) and have a sharp decrease in binding at pH values below 6.0. However, the CI-MPR binds ligand avidly at pH 7.4 while the CD-MPR is incapable to bind ligand at neutral pH.

The receptor proteins bind to lysosomal hydrolases on the luminal side of the TGN membrane and to adaptins participating in the assembly of clathrin coats on the cytosolic side. Cargo proteins can make an important contribution to their own sorting and formation of coated vesicles [123]. The vesicles deliver their contents to LE, where the low pH induces dissociation of lysosomal enzymes from MPRs. The hydrolases are afterwards dephosphorylated and MPRs are recycled back to the Golgi or to the plasma membrane. Palmitoylation is required for the recycling of the lysosomal sorting receptors back to the TGN [124]. Transport of MPRs to the LE or back to the TGN or plasma membrane is specified by signal peptides in the cytoplasmic tail of the MPRs [125-128]. Some newly synthesized lysosomal enzymes escape the binding to MPR in the TGN and are secreted. The CI-MPR localized at the plasma membrane is capable of binding the escaped M6P-containing lysosomal enzymes and return them by receptor – mediated endocytosis to lysosomes via early and late endosome.

Some lysosomal hydrolases such as cathepsin B [129], acid α-glucosidase [130] and lysosomal acid phosphatase (LAP) [131] are delivered to the LE as precursor forms and they are processed in the acidic milieu of endosomes or lysosomes by autocatalytic proteolysis or other proteases.

In addition to MPR–mediated trafficking there are alternative mechanisms of delivery of soluble lysosomal enzymes to the endosome-lysosome system. This finding is supported by the studies in patients with I-cell disease (ICD, mucolipidosis II, or ML II) [132] and of MPR

knockout – mice. All cells and tissues of ICD patients are deficient in the phosphotransferase activity, which means that lysosomal enzymes lack their M6P recognition signal and the lysosomal hydrolases are secreted. In spite of this finding not all cells are deficient in lysosomal enzyme content. Similarly, studies performed in MPRs - deficient mice [133] demonstrated cell type – specific MPR independent mechanisms for the transport of lysosomal hydrolases.

Prosaposin, a precursor of four lysosomal saposins (A – D), acid sphingomyelinase (ASM) and GM2 activator protein, all belonging to the saposin – like protein (SAPLIP) family, share a common intracellular receptor - sortilin for sorting from TGN to the lysosomes [134-137]. Sortilin also functions as an alternative sorting receptor for cathepsin D and H [138]. Beta-glucocerebrosidase employs LIMP2 as a receptor molecule in the lysosomal membrane [139]. In addition, acid phosphatase most probably utilizes mannose-phosphate mediated recycling pathway to lysosomes from cytoplasmic membrane [140].

The lysosomal membrane proteins (LMP), such as LAMPs, LIMPs and other, are sorted from the TGN to the lysosomes by an M6P - independent pathway. The LMP leave the TGN in clathrin – coated vesicles distinct from those that transport the M6P – tagged hydrolases and they are delivered to lysosomes on the basis of either tyrosine or di-leucine – targeting signals in their cytoplasmic C-terminal tails [131, 141, 142]. The signals are recognised by four heterotetrameric adaptor protein complexes (AP1 – 4) or by three monomeric adaptor proteins (GGA1 - 3). APs and GGAs play a significant role in the Golgi – endosome sorting, endocytosis and vesicle budding [123]. APs are involved in the transport of lysosomal membrane proteins (LAMPs, LIMPs and others) [94, 143] and sometimes M6PRs [144]. GGAs mediate sorting of M6PRs and sortilin [145, 146].

The lysosomal system and lysosomal proteins in *C. elegans* are discussed in section 2.1.6.

2.3. Lysosomal storage diseases

2.3.1. Main features

Lysosomal storage diseases (LSDs) are a group of inherited metabolic diseases (IMD, inborn errors of metabolism), which are characterised by an accumulation of undegraded material in the lysosomes. The global incidence of LSDs as a group is about 1:6000 newborns, but individually they are mostly very rare [147].

There are 48 different storage disorders described to date [148]. The most recently characterized LSDs at the molecular level are MPSIIIC (appended publication 2), ceroid lipofuscinosis 9 - CLN9 deficiency [149], cathepsin D deficient NCL [150] and MFSD8 deficient NCL [151]. Majority of LSDs are inherited in an autosomal recessive manner, with the exception of several X-linked disorders (see below).

LSDs are characterised by incomplete degradation of macromolecules and accumulation of partially degraded material in the lysosomes. The accumulation of undegraded material has been considered for a long time as the predominant reason of cellular and tissue damage and subsequent organ dysfunction. Current concept of pathogenesis of lysosomal storage accepts a paradigm of complex systemic pathologic consequences due to single protein deficiencies. The predominant fraction of LSDs is caused by the deficiency of lysosomal hydrolases or their activators, however, several storage disorders result from the deficits of proteins, which are not directly involved in hydrolysis of molecules within the lysosome (e.g. Niemann-Pick type C1 a C2 (NPC 1) [152-154], Mucolipidosis type II (I-cell disease) [155, 156] and Mucolipidosis type IV (ML IV) [157] or Dannon disease [158].

The occurrence of lysosomal storage disorders is truly pan-ethnic, nevertheless certain ethnic groups with tendency to socio-economic clustering (Ashkenazi Jews and others) have higher incidence of these disorders.

LSDs have a broad spectrum of clinical phenotypes. Furthermore a single disorder can vary in the age of onset and severity of symptoms including central nervous system (CNS) manifestation. Common symptoms of many LSDs include organomegaly, CNS dysfunction, bone abnormalities and coarse facial features.

The predominating approach, beside others, to the treatment of LSDs is based on substitution of the defective protein with its active counterpart. Lysosomal proteins can reach lysosomes from the extracellular environment by receptor – mediated endocytosis via CI-MPR or other receptors such as mannose receptor. Enzyme replacement therapy (ERT) is based on this finding and was first successfully used in patients with Gaucher disease [159]. Today the ERT is utilized for treatment of Gaucher disease [160], Fabry disease [161, 162], Hurler disease (MPS I) [163], Pompe disease [164, 165] and enzymes for treatment of other LSDs are currently tested in clinical trials. ERT has been demonstrated to improve non-neuropathic symptoms but is currently virtually ineffective for neurological affection. The important aim of current research efforts is to achieve higher effectiveness of ERT [166, 167].

Bone marrow transplantation also provides replacement of defective protein and was successfully applied in some forms of mucopolysaccharidosis, in non-neuropathic Gaucher

disease [168] and other disorders. However, it has high morbidity and mortality and poor donor availability which limit its clinical use. The targeting of the therapeutic enzymes to deficient cells in the central nervous system due to the blood-brain barrier (BBB) still represents a major pharmacokinetic problem for ERT.

Gene therapy is an attractive therapeutic alternative. Gene therapy strategies are under development for both non-neurological and neuronopathic LSDs employing various animal models of LSDs [169-173]. In spite of the numerous studies over the decade there have still remained many limitations (transient expression, blood-brain barrier, humoral responses) for effective gene therapy of LSDs in humans.

Additional currently used and explored approach is substrate reduction therapy (SRT) which modulates the availability of substrate by its modification or inhibition of its synthesis. SRT can improve the clinical symptomatology especially in patients with significant (relatively high) residual activity. For example, an inhibitor of a glucosylceramide synthase (N-butyldeoxynojirimycin) is effective in treatment of non-neuronopathic Gaucher disease [174, 175]. Oral administration of another inhibitor (1-ethylendioxyphenyl-2-palmitoylamino-3-pyrrolidinopropanol, EtDO-P4) to Fabry disease mice resulted in the reduction of the overall amount of accumulated globotriaosylceramide [176].

The most current therapeutic strategy - enzyme enhancement therapy (EET) [159], intends to stabilize and increase the residual activity/function of the mutated protein. EET is efficient in the cases in which certain mutations cause protein misfolding and aggregation, instability and /or altered trafficking to lysosomes. Molecular basis of EET is the use of low-molecular-weight chaperones [177], which include substrate analogues, competitive inhibitors, or other modulators that can specifically and reversibly bind and stabilize the mutant proteins and thus diminish their endoplasmatic reticulum – associated degradation (ERAD). EET can be effective in the treatment of the neuropathic LSDs due to low-molecular-weight of these compounds (pharmacological chaperons), which enables them to cross BBB. Therapeutic use of specifically acting chaperones has been, so far, described in two LSDs - D-galactose or 1-deoxygalctonojirimycin in Fabry disease [177, 178] and N-(*n*-nonyl) deoxygalactonojirimycin in Gaucher´s disease [177].

2.3.2. Lysosomal storage diseases of our interest

We have focused on LSDs with impaired metabolism of glycans, namely on Fabry, Schindler, Pompe diseases and muccopolysaccharidosis (MPS) type IIIC. We hypothesized that the worm, which has active glycan metabolism, will have orthologs corresponding to

human enzymes deficient in the above disorders. The protein families of glycosidases which are deficient in the first three disorders are well characterised (see below) while the protein deficient in MPSIIIC was unknown and its enzymatic action was only partially characterized. The gene deficient in MPSIIIC was recently identified in our lab (appended publication 2) and, as will be shown later, there is no apparent ortholog in the worm.

2.3.2.1. Fabry disease

Fabry disease (MIM no. 301500) is an X-linked recessive inborn error of glycosphingolipid catabolism resulting from the deficient activity of the lysosomal hydrolase α-galactosidase A (α-D-galactoside galactohydrolase; EC 3.2.1.22; α-GAL A). This enzymatic deficiency results in the progressive deposition of neutral glycosphingolipids with terminal α-galactosyl moieties in the lysosomes of endothelial, perithelial, and smooth-muscle cells of blood vessels and in other tissues such as heart, kidneys or eyes [179].

Glycosphingolipids (GSL) are important constituents of the cytoplasmic membranes as well as intracellular membranes (e.g. Golgi and lysosomal membranes) and have a high metabolic turnover [179, 180]. The cells with the highest turnover, such as endothelial cells, perithelial cells and smooth-muscle cells of the cardiovascular and renal system, are the ones most affected in patients with Fabry disease. The major accumulated compound is globotriaosylceramide (Gb_3Cer) followed by galabiosylceramide (Ga_2Cer) [181]. The abnormal accumulation of blood group B glycosphingolipids with terminal α-galactosyl moieties has been reported in Fabry patients with B or AB blood types [182, 183]. However, the level of accumulation of blood group B glycosphingolipids is much lower compared to Gb_3Cer and Ga_2Cer.

Alpha-galactosidase A gene has been localized to a region on the long arm of the X chromosome, Xq22. The gene consists of seven exons and six introns and its length is approximately 12 kbp. *Alpha-GAL A* gene contains a high number of repetitive sequences. There are 12 *Alu* repetitive elements distributed along the gene including 3´ flanking region. These repetitive sequences form about 30 percent of the gene, which makes *α-GAL A* gene one of the most *Alu* – rich genes with approximately 1 *Alu* per 1 kb. Several regulatory elements were identified in the 5´ flanking region of *α-GAL A* gene including one TATA, five CCAAT box and some other regulatory elements [184]. The promoter also contains CpG enriched regions (CpG islands) that are typically found in housekeeping genes, where they are implicated in transcriptional regulation and in maintaining inactivation of X chromosome

linked genes [185]. The full-length cDNA has 1393 bp and contains 60 bp of 5′ untranslated region, the initiation codon, and the entire open reading frame, which encoded a precursor protein of 429 amino acids including 31 amino acids long signal peptide. An unusual feature of *α-GAL A* cDNA is the absence of 3′ untranslated region. The polyadenylation signal sequence is immediately followed by the TAA termination codon [186]. RNA editing of human *α-GAL A* has been reported [187], however this finding has never been reconfirmed by another group [188]. Isolation and characterization of the human *α-GAL A* cDNA and DNA have enabled identification of the pathogenic variations causing Fabry disease. Over 240 mostly private mutations have been identified including partial gene rearrangements, point mutations and splice–junction defects. Approximately 70 percent of the mutations are missense or nonsense mutations. The disease is therefore very heterogenous at the molecular level [179, 189].

Alpha-GAL A belongs to the glycoside hydrolase family 27, clan D [190, 191]. Members of the family have been identified in animals, plants and microorganisms and they all share similar active site and reaction mechanism. The X–ray crystallographic structures of human α-GAL [192, 193] and rice α-GAL [194] have been determined. Preliminary structure of α-GAL I from *Mortierella vinacea* has also been reported [195]. Native human α-GAL A is a protein of about 101 kDa that forms a homodimer. The enzyme is relatively heat–labile glycoprotein. Alpha-GAL A monomer contains four potential N-glycosylation sites. Variant amounts of sialic acid residues in the carbohydrate sidechains result in different tissue forms of the enzyme [179]. Human α-GAL A has a homodimeric quarternary structure and each monomer unit is composed of two distinct domains. Domain I contains the active site and adopts a $(\beta/\alpha)_8$ barrel structure - a common motif in many glycoside hydrolases. Domain II has eight antiparallel β strands, packed into two β sheets in a β sandwich fold that contains a Greek key motif [193].

The clinical manifestations of Fabry disease apparently result mainly from progressive deposition of Gb_3Cer, while exact molecular mechanisms leading to tissue damage are largely unknown [196]. Clinical onset usually occurs during childhood or adolescence, but it may be delayed until the second or third decade. Common features of the disease in affected hemizygotes are nephropathy, cardiopathy, angiokeratomas, clouding of the cornea and lens, hypohidrosis and pain of hands and feet, joints, muscles and abdomem. Stroke, seizures, heart disorders (ischemia and infarction) and kidney disorders with proteinuria and chronic renal failure develop in the third or fourth decade of life and result in premature mortality. Late–

onset variants with predominantly neurological, cardiac or renal manifestations have also been described. So-called, “cardiac variant” of Fabry disease correlates with higher residual α-GAL A activities and certain mutations and is without clinical renal impairement [179]. Several mutations have been found in variant phenotypes [197]. The finding of patients with the same mutation but different phenotypes, suggest that other modifying factors may contribute to the phenotypic variability of Fabry disease [179]. All hemizygous males suffer from Fabry disease as well as majority of heterozygous females who may express moderate or even severe forms of the disorder, because of the X chromosome localization of the *α-GAL A* gene. The more severe forms are associated with skewed X chromosome inactivation [198].

Diagnosis of suspected hemizygotes is confirmed biochemically by demonstration of deficient α-GAL A activity in plasma, leukocytes or cultured cells. Fabry patients have increased levels of Gb_3Cer in plasma or urinary sediment. Hemizygotes with classical phenotype usually have no detectable α-GAL A activity when measured with the addition of α-*N*-acetylgalactosaminidase inhibitor - *N*-acetylgalactosamine [199]. Reliable detection of heterozygotes is based on DNA analysis. X-chromosome inactivation can be studied by analysis of DNA methylation of the X chromosome at the androgen receptor gene HUMARA [200].

Until recently the treatment of Fabry patients has been solely symptomatic. Few years ago two recombinant α-galactosidases A, agalsidase α (Replagal^R) and agalsidase β (Fabrazyme^R), have been approved for the ERT of Fabry disease (see section 2.3.1). It was shown that ERT is well tolerated, decreases the risk of additional complications and relieves the pain attacks. Unfortunately, in patients with fully developed symptoms it only slows the disease progression [161, 162, 201, 202] and it remains to be seen whether early treatment will prevent or mitigate symptoms. The effects of ERT on the tissue and cellular level in Fabry disease are currently under intensive study. Recent reports demonstrate variable distribution patterns of the recombinant enzyme throughout the lysosomal system of different cell types [203]. The prenatal diagnosis is used for prevention of the disease in the affected families.

2.3.2.2. Schindler disease

Schindler disease (MIM no. 609241-2) is a rare autosomal recessive disorder caused by the deficient activity of α-*N*-acetylgalactosaminidase (α-N-acetyl-D-galactosamide: N-acetylgalactosaminohydrolase; EC 3.2.1.49; α-NAGA). The enzymatic defect leads to the

accumulation of sialylated- and asialoglycopeptides, glycosphingolipids (e.g. blood group A glycosphingolipids) and oligosaccharides with α-*N*-acetylgalactosaminyl residues in various tissues and fluids [204].

Early studies suggested that human α-*N*-acetylgalactosaminidase, which was previously referenced as lysosomal α-galactosidase B, is a glycoform of α-galactosidase A. Purified enzymes had similar physical properties, including subunit molecular mass (~ 46 kDa), homodimeric structure, and amino acid sequences. However subsequent studies demonstrated kinetic, structural, and immunologic differences which indicated that these enzymes (α-GAL A and α-NAGA) were not glycoforms, but products of two different genes [205-207]. The two genes differ in the number of exons and in the number, position and orientation of *Alu* repeats. Exons 2 – 7 of the *α-NAGA* gene show considerable similarity to the first six exons of *α-GAL A* gene, contrary to little homology between the *α-GAL A* exons 7 and *α-NAGA* exons 8 and 9. Because of the remarkable amino acids identity (49 %) and similarity between the two genes (63 %), Wang and co-workers suggested that these two genes evolved by duplication and divergence from a common ancestral gene [208, 209].

Alpha-NAGA gene has been localized to the chromosomal region 22q13.1 – 13.2. Its gene consists of nine exons ranging in length from 95 bp (exon 5) to 649 bp (exon 9). The 3.6 kb full – length cDNA and the entire 14 kbp gene have been isolated and sequenced [208, 210]. A precursor peptide of α-NAGA consists of 411 aminoacids including a 17 aminoacids signal peptide.

Human α-NAGA is a relatively thermostable enzyme with a native molecular weight of about 110 kDa and homodimeric structure. α-NAGA is a member of the glycoside hydrolase family 27 clan D like α-GAL A. Alpha-NAGA monomer contains six potential N-glycosylation sites and four of them are conserved in identical positions as in human and murine α-GAL. Alpha-NAGA, likewise α-GAL A, has complex, high mannose, and hybrid-type oligosaccharide structures that are similarly phosphorylated [211]. The structure of chicken α-NAGA has been determined by X – ray crystallography [212] and was found to be highly similar to the structure of human α-GAL A (see above).

Schindler disease is clinically heterogenous and three phenotypically distinct forms of the disorder were identified [204]. Type I disease is characterized by normal development in the first year of life, followed by rapid neurodegenerative course. Type II disease is an adult-onset form characterized by angiokeratoma corporis diffusum and mild intellectual impairment. Type III disease is an intermediate and variable form with manifestations ranging

from epilepsy and mental retardation in infancy to autistic presentation in early childhood. All three forms of the Schindler disease have very low levels of α-NAGA activity and identical patterns of urinary oligosaccharide and glycopeptide accumulation. The fundamental cellular pathogenesis of the disease is still unclear.

Suspected patients can be initially screened by the analysis of the urinary oligosaccharides and glycopeptides. The diagnosis is confirmed by the determination of α-NAGA activity in various tissues or body fluids (plasma, leukocytes, cultured lymphoblasts or cultured skin fibroblasts) and in case of prenatal diagnosis in chorionic villi or cultured amniocytes.

Currently there is no specific treatment for the Schindler disease available.

2.3.2.3. Glycogen storage disease type II (Pompe disease)

Glycogen storage disease type II (GSD II), also named Pompe disease or acid α-glucosidase deficiency (MIM no. 232300), is an autosomal recessive disorder of glycogen metabolism resulting from the deficient activity of the lysosomal hydrolase acid α-glucosidase (alpha-D-glucoside glucohydrolase; EC 3.2.1.20; GAA). The hydrolysis of glycogen by GAA in lysosomes is one of its degradation pathways. Glycogen is delivered to lysosomes by no other means but autophagy [213]. The enzyme deficiency results in intralysosomal accumulation of glycogen of normal structure in various tissues and organs, mainly in cardiac and skeletal muscles and in hepatocytes [130].

GAA gene is localized at the long arm of chromosome 17, 17q25.2-q25.3 and its structure including mRNA splicing has been delineated in detail [214, 215]. The gene consists of 20 exons and its length is approximately 20 kbp. The first exon is non – coding and exons 10 and 11 contain the evolutionarily conserved catalytic site domain. The promoter has features characteristic of a housekeeping gene. The GC content is high (about 80 %) and obvious TATA and CCAAT motifs are lacking. The cDNA is over 3.6 kb long, with 2859 nucleotides of coding sequence. More than 20 single nucleotide polymorphisms (SNPs) have been reported in the coding region of *GAA* gene. These normal genetic variants were suggested to interact with mutations and modulate the resulting phenotype [130]. Approximately 300 mutations [216] have been identified including point mutations, splice – junction defects, small deletions and insertions and large deletions.

GAA is a glycoprotein that catalyzes the hydrolysis of α-1,4- and α-1,6-glucosidic bonds. Human GAA, as deduced from the cDNA sequence, consist of 952 amino acids and

the molecular weight of the non-glycosylated protein was estimated to be 105 kDa. The primary structure of GAA indicates that there are seven potential N-glycosylation sites. All of these sites can be glycosylated and at least two of the oligosaccharide side chains are phosphorylated [217]. The enzyme is additionally modified by both N- and C- terminal proteolytic cleavage, primarily within lysosomes. These modifications are essential for catalytic activity [130, 218].

GAA belongs to the glycosyl hydrolases family 31 (GH31) which includes many others hydrolases from different eukaryotic and prokaryotic organism, all having significant sequence similarity [190, 219-222]. Human genome contains at least four other paralogs belonging to the family GH31 – sucrase-isomaltase (SUIS) and maltase-glucoamylase (MGA) with acid pH optimum and glucosidase II alpha subunit (GANAB) and neutral α-glucosidase C (GANC) functioning at neutral pH. All five proteins share common GH31 active side sequence [223]. Orthologs of GAA have been identified in vertebrates, including mammals and birds, *C. elegans*, yeast, prokaryotes and plants [130]. YicI from *Escherichia coli* [219], MalA from *Sulfolobus solfataricus* [224] and N-terminal subunit of human maltase-glucoamylase (NtMGA) [225] are the only enzymes of family GH31, for which the three-dimensional (3D) structure has been determined. The catalytic domain is composed of $(\beta/\alpha)_8$ barrel structure (for details about GH31 structural properties see appended submitted manuscript). High level of structural similarity provides convenient conditions for understanding influence of various mutations on the protein functions in humans [219].

Clinical presentation of GSD II spans a range of clinical phenotypes, all of which include varying degrees of myopathy but differ in the age of onset, extent of accumulation of glycogen in organs and rate of progression. The most severe is the classic infantile–onset disease, characterized by cardiomegaly, hypotonia, hepatomegaly and death usually before 2 years of age. The other extreme of the phenotypic spectrum is slowly progressive adult–onset variant dominated by proximal myopathy. Between these two forms, there is a heterogenous continuum of clinical phenotypes, variously termed childhood, juvenile, or muscular variants. Patients usually succumb to respiratory failure.

Clinical diagnosis is confirmed by the demonstration of deficient GAA activity mainly in muscle, fibroblasts, lymphocytes, leukocytes, urine, and in case of prenatal diagnosis in chorionic villi or cultured amniotic cells. Acarbose, as a potent inhibitor of SUIS and MGA, is used for the direct and specific measuerement of GAA activity [226]. Residual enzyme activity can overlap with activity found in healthy heterozygote carriers. The ultimate diagnostic tool is direct DNA – based detection of pathogenic variations in *GAA* gene.

Prior to 2006 effective therapy was not available, although supportive therapy can have significant impact on patients with late – onset disease. Clinical trials with recombinant GAA (aglucosidase alfa, Myozyme®) showed prolonged survival, reversal of cardiomyopathy and motor gains in infantile patients. On the basis of these results Myozyme was approved for the ERT of Pompe disease [165]. Suitable gene therapy is under investigation [173, 227, 228]. The successful development of gene therapy in GAA-knockout mice indicates that this approach could be applicable for treatment of GSD II.

There are several animal models of the GSD II (e.g. cattle, dog, quail, cat, sheep and mouse), either occurring naturally or developed by genetic engineering [130]. Based on the data that demonstrate the importance of glycogen metabolism in nematodes [229, 230], we reasoned that *C. elegans* may be an additional suitable model for GSD II.

2.3.2.4. Mucopolysaccharidosis III C (Sanfilippo syndrome C)

The mucopolysaccharidoses (MPS) are a family of inherited lysosomal storage disorders caused by deficiencies of enzymes catalyzing the degradation of glycosaminoglycans (GAG), such as dermatan sulphate, heparan sulphate, keratan sulphate, chondroitin sulphate or hyaluran. The undegraded or partially degraded GAGs are then accumulated in lysosomes, which results in cell, tissue and organ dysfunction, and increased excretion of GAGs in urine. There are 11 known enzyme deficiencies that result in 8 distinct MPS (MPS I – VII and IX) [231].

The mucopolysaccharidosis III C (MPS IIIC), also called Sanfilippo syndrome C (MIM no. 252930), is a rare autosomal recessive disorder of mucopolysaccharide catabolism caused by the inherited deficiency of the lysosomal membrane enzyme heparin acetyl-coenzyme A (CoA):α-glucosaminide *N*-acetyltransferase (N-acetyltransferase, HGSNAT; EC 2.3.1.78), which leads to the intralysosomal storage of heparan sulphate in all organs and its excretion in urine.

Heparan sulphate is a polysaccharide which is synthesized and transported to the cell surface or extracellular matrix as a component of proteoglycans, which fulfil diverse biologic functions such as binding and activation of growth factors and/or may participate in the regulation of cell adhesion and migration [232, 233].

Until recently, N-acetyltransferase gene had not been identified and its protein structure characterized. Ausseil and co-workers mapped N-acetyltransferase gene to an 8.3 cM (16Mbp) interval in the pericentromeric region of chromosome 8 [234]. The region contained about 87 predicted or known genes. Biochemical studies showed that N-

acetyltransferase is an integral membrane protein and its incubation with acetyl[^{14}C]CoA in the absence of the acceptor of the acetyl group resulted in radioactive labelling of a 120 kDa polypeptide, which suggests that it represents a subunit containing the enzyme active site. This finding was further supported by the absence of this labeled polypeptide in the cells of MPS III C patients [235]. The N-acetyltransferase gene has been identified recently [236] (appended publication 2). The gene maps to 8p11.1 and encodes a 73 kDa protein with predicted multiple transmembrane domains and glycosylation sites (the transmembrane protein 76 (TMEM 76), identical to HGSNAT).

The N-acetyltransferase is the only known lysosomal enzyme that is not a hydrolase. It catalyzes the acetylation of the terminal glucosamine residues of heparan sulphate that cannot be hydrolyzed directly. It is inferred that the enzyme acetylates itself on the cytoplasmic side of the lysosomes by transfer of acetyl group from acetyl-CoA. This part of the reaction occurs at neutral pH. Acetyl group is then transferred to the luminal side and is used to acetylate heparan sulphate [237]. This two – step mechanism has been further confirmed by a finding that cells from some MPS III C patients can catalyze the first part of the reaction but not the second, whereas cells from other MPS III C patients lack both activities [238].

All Sanfilippo disorders (A, B, C a D) are clinically similar. MPS III C is characterised by severe CNS degeneration (progressive neuropsychiatric problems, mental retardation associated with hearing loss). Somatic symptoms include hepatomegaly, short stature, joint stiffness, coarse facial features and hypertrichosis are mostly mild compared to other mucopolysaccharidoses. Clinically, the onset usually occurs between the second and sixth year of life [231, 239].

The diagnosis of MPSs is based on the analysis of urinary GAGs - a method that still continues to be used as a valid screening test. The precise diagnosis is confirmed by the enzymatic activity assay [240, 241]. The diagnostic algorithms including those for prenatal testing in affected families are similar to other lysosomal enzymatic deficiencies combining biochemical and lately also DNA based approaches.

2.3.3. Lipid rafts and composition of lysosomal membrane

It has been shown that N-acetyltransferase is associated with lipid rafts in lysosomal membrane [242]. Lipid rafts [243, 244] are dynamic and heterogenous membrane microdomains that are enriched in cholesterol and glycosphingolipids and have specific lipid and protein composition. Lipid rafts are also called detergent resistant membranes (DRMs),

because these domains are not readily solubilised in non-ionic detergents such as Triton X-100 or 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS). DRMs were first described in the cytoplasmic membrane and lately also in other intracellular compartment, such as ER, GN, mitochondria and lysosomal-endosomal system [242, 245-247]. Lipid rafts play active roles in various physiological processes including signal transduction, vesicle trafficking (sorting in TGN and endosomes) [248], cell adhesion and motility, entry of pathogenic viruses or bacterial toxins such as cholera toxin, which binds the DRM-enriched ganglioside GM1 [249, 250]. Proper constitution of lysosomal membrane microdomains is necessary for chaperone-mediated autophagy [115]. LAMP 2A, a receptor of CMA, is concentrated in DRMs at the lysosomal membrane and its dynamic association with these microdomains and distribution between the membrane and lysosomal lumen plays a key role in CMA regulation. Additional information about the functions and composition of lysosomal membrane DRMs and their roles in molecular pathology of LSDs is continually becoming available.

3. Aims of the study

Study of cellular pathology of human lysosomal diseases is very important and interesting because it enables us to understand the related biological processes in cells. Not everything is practicable to study on patients, so it is important and needful to use animal models. However, work with mammalian animal models is on overall very complicated and demanding. *C. elegans* is a simple multicellular organism which was previously used to successfully study several human diseases including some lysosomal storage disorders caused by deficiency of lysosomal non-catalytic proteins. It represents ideal intermediate stage between study of tissue culture and mammalian models.

Subject of this Ph.D. thesis was to study selected deficits of lysosomal glycoconjugates degradation (Fabry, Schindler, Pompe and Mucopolysaccharidosis IIIC diseases) in term of cell biology, biochemistry and genetics and find out if *C. elegans* is suitable model for lysosomal enzymopathies. First task was to find and characterize the true *C. elegans* orthologs and then perform systematic RNA-mediated interference assays of these orthologous genes with subsequent characterization of phenotype of interfered organisms. Not all lysosomal proteins have their orthologs in *C. elegans* and one example is human Acetyl-CoA: glucosaminide N-acetyltransferase (HGSNAT), deficit of which leads to MPS IIIC. Based on this finding we could not study this lysosomal membrane protein in *C. elegans*. I participated in the discovery of the gene coding HGSNAT and other aims of this study were to perform basic bionformatic analysis and characterize tissue and cellular distributions of HGSNAT mainly at microscopical level.

This work is composed of three sections and partial goals of each part are listed below.

3.1. Aims of the first part of the study

- To identify *C. elegans* orthologs of human α-galactosidase and α-N-acetyl-galactosaminidase which are presumed to share a common ancestor.
- To verify the predicted gene and characterize it on the cDNA level.
- To study its biological functions during *C. elegans* lifetime by RNA-mediated interference.
- To study its expression pattern by using extrachromosomal GFP array.
- To study the phylogeny of the orthologous genes.

- To study expression of *C. elegans* ortholog in wild-type and knockout mouse fibroblasts.
- To evaluate a suitability of *C. elegans* as a model organism for studying two lysosomal enzymopathies - Fabry and Schindler diseases.

3.2. Aims of the second part of the study

- To identify *C. elegans* orthologs of human acid α-glucosidase.
- To verify the predicted or partially confirmed genes and characterize them on the cDNA level.
- To study their biological functions during *C. elegans* lifetime by RNA-mediated interference.
- To characterize acquired deletion strains generated by *C. elegans* Gene Knockout Consortium.
- To study expression pattern of orthologous genes with acidic pH optimum by using extrachromosomal GFP arrays.
- To study the phylogeny of the orthologous genes.
- To evaluate a suitability of *C. elegans* as a model organism for studying glycogen lysosomal storage (Pompe disease).

3.3. Aims of the third part of the study

- To verify the sequence of the predicted gene of human acetyl-coenzyme A: α-glucosaminide N-acetyltransferase (HGSNAT) on the cDNA level.
- To identify splice variants of HGSNAT and human tissues in which is HGSNAT expressed.
- To perform basic bioinformatic analyses.
- To identify and characterize *C. elegans* ortholog of human HGSNAT.
- To study the cellular distribution of human HGSNAT and its relation to lysosomal microdomains by immunofluorescence microscopy.

4. Materials and methods

4.1. General techniques

4.1.1 Preparation of genomic DNA and cDNA

Mixed stage *C. elegans* culture and cultured skin mouse or human fibroblasts were used for preparation of total RNA [251] which was reverse-transcribed using SuperScriptII reverse transcriptase (Invitrogen) and oligo dT18 according to protocol supplied by the manufacturer (Invitrogen). Genomic DNA was isolated from cultured mouse or human skin fibroblasts by standard phenol extraction method [252].

4.1.2 Polymerase chain reaction (PCR)

PCR reactions were performed according to generally accepted protocols [253] using specific primers (section 10.1, Tables 8-11) and templates (gDNA, cDNA, plasmid DNA, and Human Multiple Tissue cDNA (MTC) Panel 1 - CLONTECH Laboratories, Inc). We used combination of DNA polymerases - KlenTaq (GeneAge) and DeepVent (New England Biolabs). PCR amplifications of each DNA fragment were done under specific conditions (first denaturation step for 1 min at 94 ^{o}C, then 35 cycles, each composed of three steps: denaturation step for 15 or 20 s at 94 ^{o}C, annealing step for 55 s at 57 – 67 ^{o}C and extension step for approximately 1 min per 1 kb at 68 – 72 ^{o}C, the terminative extension step for 10 min at 68 – 72 ^{o}C). All PCR reactions were performed on thermal cyclers DNA Engine PTC-200, DNA Engine tetrad PTC-225 and DNA Engine DYAD^{TM} (MJ Research).

4.1.3 Cloning of DNA, isolation of plasmids and DNA sequencing

Cloning into TOPO vectors (pCR 2.1, pCRII and pCR-XL) was performed according to manufacturer´s protocol with TOPO TA cloning kits (Invitrogen). T4 DNA Ligase was used (Invitrogene, Fermentas) for cloning into other vectors (L4440, pPD95.75, pPD95.67 – provided by Dr. A. Fire - Stanford University, and pCMVTag1). Gel purified DNA fragments (High Pure PCR Product Purification Kit, Roche) were ligated overnight into previously linearized vectors (see particular sections) at 16 ^{o}C. Transformation of competent cells TOP10 (Invitrogen) was performed by heat shock at 42^{o}C for 30s. Cells were spread after transformation on LB agar (LB medium with agar (15g per liter)) plates with appropriate antibiotic for selection and were allowed to grow overnight at 37 ^{o}C. Single colonies were

picked up and cultured overnight in LB medium (1% Tryptone, 0,5% Yeast Extract, 1% NaCl, pH7.0) with appropriate antibiotic. Plasmid DNAs were isolated by JETquick Plasmid Miniprep Spin Kit (Genomed) or in case of transfection by EndoFree Plasmid Maxi Kit (Qiagen) according to manufacturer´s protocols. We analyzed the plasmids for inserts by restriction analysis and/or by PCR screening. DNA sequences of inserts and whole costructs were verified by direct sequencing on the automated fluorescent sequencer Li-Cor (LI-COR, Inc.), AlfExpress (Pharmacia Biotech) or ABI 3100 – Avant sequencer (Applied Biosystems, USA).

4.1.4 Signal sequence, transmembrane domains and potentional N-glycosylation sites prediction

All predictions mentioned in this section were performed by CBS prediction servers [254]. Signal peptide sequences were predicted by SignalP (version 3.0) server and TargetP (version 1.1) server was used for prediction of intracellular protein targeting. Transmembrane domains prediction in HGSNAT was performed by TMHMM (version 2.0) server and potentional N-glycosylation sites in HGSNAT sequence was predicted by server NetNGlyc (version 1.0) which is designated for prediction of N-linked glycosylation sites in human proteins. All these predictions were done under predefined default values.

4.1.5 Total protein determination

Concentration of total protein in the samples was determinated by Hartree method (modification of Lowry method) [255]. Concentration series of standard solution bovine serum albumin (BSA) was used for calibration. 5 or 10 μl aliquots of sample homogenates were used. Absorbance was measured on the UV/VIS spectrofotometer (Unicam UV 2-100, U.K.), wavelength 650 nm.

4.1.6 Microscopic techniques

Worms and cultured mouse and human skin fibroblasts were evaluated under epifluorescence microscope (Nikon, TE 2000E) equipped with Nomarski DIC optics and confocal laser scanning head C1*si* (for detailed microscope setup see http://udmp.lf1.cuni.cz/). EZ-C1 (Nikon), Imaris Personal (Bitplane, Switzerland) and Image J (NIH, Bethesda) software packages were used for image analysis and 3D rendering. Confocal images were deconvolved with theoretical PSF using Huygens Professional (SVI, The Netherlands) Software. The potential spectral contamination of GFP signal (excited by 488 nm laser) by

unspecific intestinal autofluorescence of worms was eliminated by the use of single value decomposition/linear unmixing algorithm [256] applied to spectrally acquired confocal images. This approach is demonstrated on the Figure 20.

4.2. *C. elegans* related techniques

4.2.1 *C. elegans* strains and cultivation

The wild-type Bristol N2 and other strains of *C. elegans* were cultured under standard laboratory conditions as described previously [2]. Worms were maintained on nematode growth medium (NGM) agar (17g of Bacto Agar, 2.5 g of Bacto Pepton, 3 g of NaCl, 1 ml of Cholesterol (5 mg/ml in Ethanol), H_2O to 1l, after autoclaving were added 25 ml of 1 M KH_2PO_4, 1 ml of 1 M $CaCl_2$ and 1 ml of 1 M $MgSO_4$) in Petri dishes and were fed on *E.coli* strain OP50. RRF3 strain [257] and GS1912 strain [50] were used for some RNAi experiments. These nematode strains were provided by the *C. elegans* Genetics Center [258] (CGC). Deletion mutants of *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) were generated by the *C. elegans* Gene Knockout Consortium [259].

4.2.2 Single worm PCR (swPCR)

Individual worms were transferred to 3 ul drop of lysis buffer (50 mM KCl, 10 mM TRIS-HCL pH 8,3, 2,5 mM $MgCl_2$, 0,45% Tween 20, 0,01% gelatine and 200 μg/ml proteinase K) in the cap of 0,5 μl PCR tube. The tube was spun, afterwards, to move the drop with worm to its bottom. Samples were frozen for at least 10 min at -80 ^oC and then heated for 1 hour at 60 ^oC, 15 min at 95 ^oC and cooled down to 4 ^oC. DNA from a single worm prepared by this way was used as the template for an appropriate PCR reaction (see section 4.1.2).

4.2.3 BLAST searches

Wormbase (2002-2008 consecutive and freeze versions, http://www.wormbase.org/) databases were repeatedly searched for human α-GAL, α-NAGA and GAA orthologs in C. elegans using BLASTP program set at default values [260]. Amino acid sequence of human lysosomal α-GAL (GenBank acc. no. NP_000160), α-NAGA (GenBank acc. no. NP_000253) and GAA (GenBank acc. no. NP_000143) were used as a query sequences (GenBank [261]).

4.2.4 cDNA amplification, sequencing and gene prediction confirmation

Total RNA was isolated from mixed stage N2 cultures [251] and reversely transcribed with oligo dT-T7 (5′- AATACGACTCACTATAG) primer and Superscript reverse transcriptase (Invitrogen). The coding sequences of R07B7.11 (*gana-1*) and D2096.12, D2096.3 (*aagr-1*), R05F9.12 (*aagr-2*), F40F9.6 (*aagr-3*) and F52D1.1 (*aagr-4*) were PCR amplified in partially overlapping fragments with intragenic primers designed according to available Wormbase [29] and Worfdb [27] predictions and EST alignments. SL1 (5′-GGTTTAATTACCCAAGTTTGAG) and SL2 (5′-GGTTTTAACCCAGTTACTCAAG) together with gene specific anti-sense primers were used to amplify 5′ UTR of the cDNAs and to evaluate the mode of *trans* splicing of the appropriate mRNAs. The 3′ ends of cDNAs were amplified with T7 primer and gene specific sense primers to characterize polyadenylation sites. PCR products were cloned with TOPO TA cloning kit (Invitrogen) into the pCR 2.1 vector. Positive clones were sequenced on the Li-Cor automated fluorescent sequencer, resultant sequences were aligned with corresponding genomic cosmid sequences in the AlignIR software (Li-Cor) to evaluate splicing boundaries and overall gene organisation.

4.2.5 RNA mediated interference

The entire coding sequence of *gana-1* was used for RNAi experiments. For RNAi assays of particular *aagr*s we selected regions of *aagr-1-4* cDNAs with the lowest similarity/identity between otherwise homologous sequences (for details see appended submitted manuscript). The selected regions of *aagr-1-4* were as follows: 1500-2226 bp (*aagr-1*), 1-1161 bp (*aagr-2*), 1925-3054 bp (*aagr-3*) and 55-799 bp (*aagr-4*) where adenine from ATG initiation codon was marked as nucleotide +1. The selected cDNA sequences were amplified by PCR with gene specific primers.

PCR products were cloned into the double promoter pCRII-TOPO vector (Invitrogen) for the purposes of successive microinjection mediated RNAi experiments. Resultant constructs were separately linearized and used as templates for in-vitro transcription employing In Vitro Transcription KIT (Promega) using T3 and SP6 RNA polymerases. Single stranded antisense RNA molecules were annealed for 30 minutes at room temperature to generate double stranded RNAs (dsRNAs) which were then treated with DNase I (New England Biolabs). Standard protocols were used for dsRNA microinjection [9] of L4/young adult hermaphrodites′ germline syncytium. 8-10 worms were microinjected with dsRNA or

water (control) in each experiment. F_1 progeny of injected animals was screened for morphological RNAi effects.

PCR products were cloned into L4440 double T7 promoter vector (supplied by Dr. Andrew Fire, Standford University) for RNAi feeding experiments. In case of *aagr-1-4* the PCR products were first cloned into the pCRII-TOPO vector (Invitrogen) and checked by sequencing. The *aagrs* clone were then digested from the pCRII-TOPO vector (*aagr-1* – dual *EcoRI*, *aagr-2* and *-3 – AvaI/HindIII* and *aagr-4* – dual *AvaI* digestion). Acquired fragments were gel purified and ligated into L4440 vector. The plasmids were transformed into *E. coli* strain HT115 [8] carrying IPTG inducible T7 RNA polymerase. Worms were fed on induced HT115 *E. coli* carrying the relevant L4440 construct. Worms fed on HT115 carrying L4440 plasmid without insert were used as a control. F_1 and F_2 progeny was screened for morphological phenotypes.

Interfered nematodes were treated as described in section 4.2.6 for biochemical measurements.

4.2.6 Determination of α-galactosidase, α-N-acetylgalactosaminidase, α-glucosidase and β-hexosaminidase (control) activities

Enzyme activities were determined in the crude homogenate of worms. Prior to all activity measurements worms were washed from culture plates and repeatedly rinsed and centrifuged in M9 buffer (6g of Na_2HPO_4, 3g of KH_2PO_4, 5g of NaCl, 0,5g of $MgSO_4.7H_2O$, H_2O to 1l) and finally pelleted by centrifugation. 4-methylumbelliferyl (MU)-α-D-galactopyranoside (1mM), 4-MU-α-D-N-acetylgalactosaminide (1mM) and 4-MU-β-D-glucopyranoside (2,5 mM) in the McIlvaine buffer (0.1M citrate/0.2M phosphate buffer, pH 4,7) and 4-MU-α–D-glucopyranoside (1,26 mM) in the McIlvaine buffer (pH 4,0 or 6,5), were used as enzyme substrates. Reaction mixtures (20 µl of enzyme substrate and 10 or 20 µl of sample solution) were incubated at 37°C (α-Gal for 30 min, α- NAGA for 60 min, α-Glu for 60 min and β-Hex for 15 min). Reactions were stopped by 600 µl of 0.2 M glycine/ NaOH buffer (pH 10,6) [199, 262]. The activities were also determined at 25°C in normal N2 strain homogenates. For some measurements of α-galactosidase and α-N-acetylgalactosaminidase activities inhibitors N-acetyl-D-galactosamine (0,1 M) (Sigma), D-galactose (0,1 M) (Sigma) or D-glucose (0,1 M) (Sigma) were added and for selected measurements of α-glucosidase activity selective inhibitor acarbose (8 µM) (Sigma) [226] was added to the reaction mixture. Fluorescence signal of the 4-methylumbelliferone was measured on the luminiscence

spectrofotometer LS 50B (Perkin-Elmer) (excitation 365 nm and emission 448 nm). 4-methylumbelliferone (1μM) was used for calibration (blank, 0,2, 0,4, and 0,6 μM solutions). All measurements were performed in doublets.

4.2.7 Transgenic GFP expression

Mixed stage N2 nematode culture genomic DNA was used as a template to amplify inserts including regulatory and part or whole coding sequence of the analysed genes. The PCR products were cloned in-frame with GFP coding sequence of the vectors pPD95.67 containing a nuclear localization sequence (NLS) and/or pPD95.75 without NLS (both vectors supplied by Dr. Andrew Fire, Standford University). Extrachromosomal transgene expressing lineages were generated by co-injecting plasmid DNA along with the marker plasmid pRF4 (containing the *rol–6(su1006)* mutant collagen gene) at the total concentration of 100 ng/μl into L4/young adult germline syncytium of N2 worms [263]. Presence of the transgene was verified both in plasmid DNA and RNA by PCR and sequencing. The transgenic progeny was screened for GFP signal by epifluorescence microscopy (see section 4.1.6).

4.2.7.1 Preparation of *gana-1* GFP fusion construct

The entire coding region of the *gana-1* gene together with 3 kb of its 5´ upstream sequences was amplified by a nested PCR reaction using DyNAzyme EXT^{TM} PCR system (Finnzymes) for preparation of long PCR products and two pairs of gene specific primers (section 10.1, Table 8). The internal primers contain overhang with restriction sites of SphI and SalI respectively. The amplified 4709 bp long PCR product was cloned into pCR-XL-TOPO vector using TOPO XL PCR cloning kit (Invitrogen). The SphI and SalI restriction fragment of *gana-1* was recloned in frame with GFP to the corresponding restriction sites in the digested GFP reporter vector pPD95.67. Accuracy of the *gana-1*::GFP transgene was confirmed by sequencing.

4.2.7.2 Preparation of the *aagr-1* and a*agr-2* GFP fusion constructs

In the first step was amplified the 6 kb fragment spanning 459-6599 bp of cosmid D2096. The PCR product was used as a template for three second round PCR reactions with specific sense primers aligning at 711-730 (construct pJS6), 2942-2962 (construct pJS4) and 5446-5463 nucleotide (construct pJS1) of the D2096 cosmid sequence and containing overhang with restriction site of *PstI*. As the second primer was used universal antisense primer aligning at 6220-6241 nucleotide of the D2096 cosmid with overhang containing

restriction site of *BamHI* (primers in Table 9, section 10.1). Acquired PCR products 5531 bp (pJS6), 3300bp (pJS4) and 796bp (pJS1), respectively, were cloned using TOPO XL PCR cloning kit (Invitrogen) into pCR-XL-TOPO vector. The *PstI* and *BamHI* restriction fragments were ligated into *PstI*/*BamHI* linearised pPD95.75 and/or pPD95.69 vectors in-frame with open reading frame of GFP.

To study the expression of *aagr-1* we amplified the 3720bp fragment spanning 6242-9961 nucleotide of D2096 cosmid sequence. For PCR amplification we used primers containing C-terminal *SmaI* restriction site. Resultant PCR product was cloned into pCR-XL-TOPO vector and the construct was then digested by *SmaI* and *HindIII*. This 2106 bp long fragment contained complete intron 1 and 1340 bp of *aagr-1* ATG upstream regulatory sequence. The fragment was ligated into *HindIII/SmaI* linearised pPD95.75 vector in-frame with GFP open reading frame (construct pJS8).

The *aagr-2* GFP fusion construct that included 3431 bp of *aagr-2* regulatory sequence was prepared similarly. The first step included amplification of the 3862 bp fragment spanning overlapping R05F9 and ZK546 cosmids. Resultant PCR product was used as a template for second round PCR with 40533-40553 nucleotide (R05F9) and 11999-12019 nucleotide (ZK546) primers with overhang containing *BamHI* restriction sites. The amplified PCR product was cloned into pCR-XL-TOPO vector and the *BamHI* digestion fragment was recloned into *BamHI* linearised pPD95.75 vector in-frame with GFP open reading frame (construct pJS9).

4.2.8 Alkalization of lysosomes

Mixed stage cultures (pJH3 and N2 (control)) were harvested from NGM OP50 plates and washed three times with H_2O. Worms were pelleted by centrifugation (max. 1000 RPM, 2 min.) between the washes. Afterwards worms were treated with either NH_4Cl or concanamycin A (CON A), the agents which are known to specifically increase pH in the acidic cellular compartments [90, 264]. Pelleted animals were resuspended in 0, 10, 25, 50, 75 and 100 mM aqueous solutions of NH_4Cl. Small aliquots of worms were examined after 30 min, 2, 4, 6, 8 and 24 hours. Alternatively, animals were soaked in 0, 10, 20, 50, 100, 150, 200 nM solutions of CON A in aqueous media. Small aliquots of worms were evaluated after 30 min, 1, 2, 3, 6 and 24 hours.

Specimen were mounted on the agarose pads and examined using epifluorescence microscope (see section 4.1.6).

4.2.9 Immunofluorescence staining of worms

Mixed stage pJH3 and N2 cultures were harvested from NGM OP50 plates and washed thoroughly in M9 buffer to deplete intestinal bacteria. Worms were pelleted by centrifugation (1000 RPM, 2 min.) between the washes. The fixation and imunnofluorescence staining procedures were based on the approaches of Nonet et al. [265]. Animals were fixed overnight in 4% paraformaldehyde in 100mM sodium/potassium phosphate buffer. After fixation the pellets were washed three times in 1× PBS, and incubated in reduction solution (1% Triton X-100, 100mM Tris (pH7,0), 1% β–Mercaptoethanol) overnight at 37°C in order to reduce the cuticle. After the reduction step, the worms were washed 5 times in 1× PBS, and then were incubated for 5 hours in 900U/ml collagenase type IV (Sigma) diluted in Krebs-Ringer solution (119 mM NaCl, 25 mM $NaHCO_3$, 11.1 mM glucose, 1.6 mM $CaCl_2.H_2O$, 4.7 mM KCl, 1.2 mM KH_2PO_4, 1.2 mM $MgSO_4.7H_2O$, pH 7.4). The reduction/digestion step was performed twice. After washing (3 times in 1× PBS) the pellets were stored for further processing in AbA buffer (1 × PBS, 0.1% Triton X-100, 1% BSA, 0,05% NaN_3). All antibody incubations were performed overnight at room temperature in AbA buffer. Primary antibody (polyclonal rabbit anti-GFP IgG (Abcam)) was diluted 1:500. Secondary antibody (goat anti-rabbit IgG Alefa Fluor 488 labeled (Molecular Probes)) was diluted 1:1000. The worms were washed three times with AbB buffer (1×PBS, 0.1% Triton X-100, 0.1%BSA, 0.05%NaN_3) between incubations. Specimens were mounted onto the agarose pads with additional DAPI or SYTOX orange (Molecular Probes) staining of the nuclei. Microscopic examination was performed as described in section 4.1.6.

4.2.10 Western blotting

Worms from mixed stage pJH3 and N2 cultures were homogenized by sonication and the concentration of protein was measured by Hartree method (see section 4.1.5). Proteins (25 – 50 μg of total protein per lane) were separated by SDS-PAGE gradient gel (4 to 20 % polyacrylamide) and consecutively transferred onto nitrocellulose membrane by semi-dry blotting using semi-dry blotting apparatus PHERO-multiblot (Biotec-Fischer GmbH, Reiskirchen, Germany) according to a common protocol [266]. Membranes were treated according to a generally accepted Western blotting protocol with chemiluminiscence detection (SuperSignal, West Pico) [266]. Primary antibody was rabbit polyclonal anti-GFP IgG (Abcam, dilution 1:5000), the secondary antibody was goat anti-rabbit IgG/Px (Pierce, diluted 1:20000).

4.2.11 Isolation and characterization of deletion mutants

Deletion mutants of *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) were kindly generated by *C. elegans* Gene Knock-out Consortium (http://celeganskoconsortium.omrf.org/). Both deletion strains were backcrossed repeatedly (at least 3 times) according to established protocols [267]. Backcrossed worms were PCR screened for the presence of homozygous deletions with N2 worms as controls (primers in Tables 9 and 10, section 10.1). PCR products were directly sequenced using 3100-*Avant* Genetic Analyzer (Applied Biosystems) to evaluate the size and position of the deletion. The backcrossed worms were checked for morphological or otherwise observable phenotypes and the enzyme activities were measured as described in section 4.2.6.

4.2.12 Multiple protein alignments and phylogenetic tree construction

Consensual translations of the sequenced cDNA´s and evolutionary relevant protein sequences were aligned using ClustalW algorithm [268] and Blosum62 matrix and the alignments were manually edited. Phylip 3.57c and 3.6 software packages [269] were used for phylogenetic tree construction employing maximum likelihood, maximum parsimony and distance matrix methods. For details see appended publications 1 and submitted manuscript.

4.2.13 Homology modeling

4.2.13.1 3D model of GANA-1

X-ray structure of chicken α-NAGA [212], rice α-GAL [194] and human α-GAL [193] were used as structural templates for model construction of GANA-1 using automatic algorithms of SwissModel server [270]. The model was manually inspected and energy was minimized in DeepView computer program [271]. For details see appended publication 1.

4.2.13.2 3D models of AAGR-1-4 and molecular docking

Protein sequence alignments of *E.coli* YicI [219], *S. solfataricus* MalA [224], *H.sapiens* NtMGA [225] and *C. elegans* AAGR-1-4 were calculated using ClustalW algorithm [268] and Blosum62 matrix. Structural models were prepared using MODELLER 9 software [272] with the *E.coli* YicI, *S. solfataricus* MalA and *H.sapiens* NtMGA as templates. Unstructured loops (insertions in AAGR-1 and -2, in special) were energy refined and the final models were selected based on DOPE score (Discrete Optimized Protein Energy) and

visual check. Quality of the models was assessed by PROCHECK [273] and MolProbity [274] programs and individual structural domains were pairwise compared using DaliLite server [275]. Docking of acarviosine was done by Autogrid 4 (generation of atom-specific affinity, electrostatic and desolvation potential maps) and Autodock 4 [276]. For details see appended submitted manuscript.

4.3. Cell culture related techniques

4.3.1 Cell culture

Mouse skin fibroblasts (wild type (wt) and two knockout lines for α-galactosidase and α-N-acetylgalactosaminidase were supplied by Prof. Detlev Schindler, Wurzburg University). Human skin fibroblasts from patients and control individuals were cultured according to standard procedure in DMEM (Sevapharma a.s) supplemented with 10% fetal calf serum (FCS) (PAA laboratories), 2% $NaHCO_3$ (Sevapharma a.s) and 1% gentamycin (Lek Pharmaceutical and Chemical Company) or penicillin/streptavidin (Diagenes s.r.o.). Cell cultures were grown in 25 cm^2 culture flasks (TPP) in 5% CO_2 incubator (Jouan IGO 150). All cell culture procedures were carried out in the laminar flow box (Biohazard, Clean Air CA/RE3).

4.3.2 Expression of gana-1 in eukaryotic expression system

4.3.2.1 Preparation of expression constructs

We used epitope tagging vector pCMV-Tag 1 (Stratagene) designed for gene expression in mammalian cells. *Gana-1* gene was inserted into the pCMV-Tag 1 vector with either FLAG or c-myc epitopes. The expression of the fusion protein can be monitored using tag-specific antibodies.

cDNA acquired by reverse transcription of RNA from mixed stage N2 nematode culture was used as a template for PCR amplification of inserts. Primers harboured overhanging sequences containing recognition sites for selected restriction enzymes (section 10.1, Table 8). Inserts included entire coding sequence of *gana-1*. In case of internal tagging, we amplified the entire coding sequence in two PCR products. The acquired PCR products were cloned into pCR 2.1 vector using TOPO TA cloning kit (Invitrogene). Positive clones were verified by sequencing on ABI 3100 – Avant sequencer (Applied Biosystems, USA).

The fragments of *gana-1* acquired by restriction were recloned *in frame* with appropriate tag epitope to the corresponding restriction sites in the digested pCMV-Tag 1 vector. Sequences of the *gana-1* constructs were evaluated by sequencing and plasmids were isolated by EndoFree Plasmid Maxi Kit (Qiagen). We acquired six *gana-1* constructs. The first was tagged with both epitopes, the Flag (N-terminal) and c-myc (C-terminal). The remaining five constructs were tagged with Flag epitope either on N-terminal or on C-terminal and three are tagged internal on various place of *gana-1* sequence.

4.3.2.2 Transfection of mouse skin fibroblasts

Mouse skin fibroblasts were transfected with the above described constructs (pCMV-Tag1+*gana-1*) using Lipofectamine 2000 reagent (GibcoBRL) or Amaxa nucleofector technology [277] (Amaxa Biosystems).

In case of Lipofectamine transfection cells were allowed to grow to 50-80% confluency in 4-well culture slides (BD Falcon) for transient transfection or in 25 cm^2 culture flasks (TPP) for creating stable cell lines. 2 μl (130 μl) of Lipofectamine reagent were mixed with 45 μl (658 μl) of serum/antibiotic-free medium (DMEM, GibcoBRL) per well (flask), and incubated for 5 minutes at room temperature (RT). The 1,5 μg (20 μg) DNA for transfection was mixed with 45 μl (658 μl) of serum/antibiotic-free medium. Both solutions were mixed and incubated for 30 minutes at room temperature to form of DNA-lipid complexes. After that cells were washed three times with PBS, transfection solution was added and incubated for 5 hours at 37^{o}C and 5% CO_2. Medium was replaced with a fresh medium containing 10% FCS. After 20 hours the transient cell lines were fixed and examined by immunofluorescence staining (for details see bellow). The stable cell lines were allowed to grow for 72 hours and after trypsinization they were cultured in the complex medium with 0,8 μg/ μl of selective antibiotic geneticin (G418) (KRD) to maintained stable cell lines.

We used the Human Dermal Fibroblast Nucleofector Kit (Amaxa Biosystems) according to recommended manufacturer´s protocol (program U-23) for fibroblast transfections. The transfection effectivity was evaluated by co-transfection with pmaxGFP vector. It was estimated to be 10%. We used $1x10^6$ cells of α-N-acetylgalactosaminidase knockout cell line, $5x10^5$ cells of α-galactosidase knockout and wt cell lines and 2,5 μg of DNA (2 μg of control vector) per reaction. Immediately after transfection cells were allowed to grow in 6-well plates with complex medium and were afterwards cultured as described above (Lipofectamine transfection - transient or stable cell lines).

4.3.2.3 Culturing of stable cell lines with therapeutic chaperones

The stable cell lines were cultured in the complex medium with 0,8 μg/ μl of G418 and with one of the therapeutic chaperones, D-galactose (200 mM) (Sigma), N-acetyl-D-galactosamine (100 mM) (Sigma) and deoxygalactonojirimycin (20 μM) (Sigma) respectively, for three to four days. Cells were examined by Western Blotting or immnonofluorescence staining.

4.3.3 Immunofluorescence staining of transfected cells

Transfected cells were fixed in 100% methanol at -20 ^{O}C for 10 min. As primary antibodies monoclonal mouse anti c-myc IgG1 (Sigma, dilution 1:150), polyclonal rabbit anti-Flag IgG (Sigma, dilution 1:100) and monoclonal rat anti mouse CD107a (Lamp-1) (Fitzgerald, dilution 1:200) were used. Secondary antibodies were goat anti-mouse IgG1 Alexa Fluor 488 labeled, goat anti rabbit IgG Alexa Fluor 568 labeled and donkey anti rat IgG Alexa Fluor 488 labeled (all diluted 1:1000, Molecular probes). Microscopical examination was performed as described above.

4.3.4 Western blotting

Cells were homogenized by sonication and the measurement of protein concentration and Western blotting procedure was performed as described above (see section 4.2.10). The proteins (25 μg of total protein per lane) were separated in 12% SDS-PAGE gel. The primary antibody - polyclonal rabbit anti-Flag IgG (Sigma) - was diluted 1:4000 and the secondary antibody - goat anti-rabbit IgG/Px (Pierce) - was diluted 1:10000. Fusion protein C-Terminal FLAG-BAP (Sigma) was used as the positive control (10 ng per lane).

The specifity of antibody against HGSNAT was verified in homogenates of leukocytes by Western blotting. We used polyclonal rabbit anti HGSNAT (2785-6) (diluted 1:3000) as the primary antibody and goat anti rabbit IgG/Px (Pierce) (diluted 1:12000) as the secondary antibody.

4.3.5 Measurement of influence of HGSNAT antibody on HGSNAT and β-hexosaminidase activities

For assays were used artificial substrates tagged with 4-methylumbelliferone (MU) in the McIlvaine buffer (see section 4.2.6). In the case of HGSNAT assay we used 10 μl of 3 mM 4-MU-β-D-glucosaminide in the McIlvaine buffer pH 5.5, 10 μl of homogenates of

leukocytes with concentration 1 μg/μl of proteins and 10 μl of 6 mM acetyl CoA (Sigma) and the reaction mixture was incubated for 17 hours at 37 ^{o}C. The amount of polyclonal rabbit anti HGSNAT antibody (2785-6) was ranged from 0 to 4.5 μg per reaction. Fluorescence signal was measured on the luminiscence spectrofotometer LS 50B (Perkin-Elmer).

4.3.6 Electron microscopy

The fixed specimens were dehydrated with ethanol, embedded in LR White resin, and polymerized for 24 hours at 60 ^{o}C. Ultrathin sections on formvar-carbon-coated gold grids were pre-treated on a drop of 10 % normal goat serum (Sigma) in PBS for 10 min. Sections were then incubated for 17 hours at 4°C with the primary polyclonal rabbit anti HGSNAT antibody (2785-6) diluted 1:20 in PBS with 0.05 % Tween 20 (Sigma) and 0.1 % BSA (Fluka). After washing with PBS-Tween and PBS alone, followed by a repeated treatment with normal goat serum, the sections were reacted with the 6 nm colloidal gold-conjugated secondary antibody goat anti-rabbit (Jackson) in PBS at room temperature for 30 min. The grids were finally thoroughly rinsed, air dried, and stained with uranyl acetate and lead citrate. Observations were carried out with a Tecnai G2 Sphera at 120 kV or JEOL 1200 electron microscopes.

4.3.7 Immunofluorescence staining and colocalization studies of HGSNAT

The immunofluorescence staining was performed according to a common protocol and the specific conditions were adjusted according to primary antibodies and their combinations used (section 10.1, Tables 12 and 13). Cells were allowed to grow to 80-95% confluency in 4-well culture slides (BD Falcon) and then washed three times with 1x PBS (5 min per wash). After washing the cells were fixed in 100% methanol at -20 ^{o}C for 10 min or in 4% paraformaldehyde (PFA) at 4 ^{o}C for 10 min with or without following permeabilization step with 0.1% TRITON X-100 performed at room temperature for 10 min. After the protein blocking step (5% FBS in 1x PBS, 30 min, RT) samples were incubated with primary antibodies diluted in 5% FBS in 1x PBS (for details see Tables 12 and 13, section 10.1) for 1 hour at 37 ^{o}C or overnight at 4 ^{o}C. Appropriate Alexa Fluor labeled antibodies (Molecular probes) were used as secondary antibodies and were diluted 1:1000 and incubated 1 hour at 37^{o}C. Nuclei of cells were stained with DAPI (1:1000, 15 min, and 37 ^{o}C). The cells were washed several times with 1x PBS between each incubation step. Specimens were mounted with aqueous mounting medium IMMU-MOUNT (Thermo Scientific) and microscopical examination was performed as described above (section 4.1.6).

5. Results and discussions

5.1. *Caenorhabditis elegans* as a model organism for selected lysosomal storage diseases

Initial studies were based on comparison of the conservation of sequence of selected lysosomal proteins *in silico* (Table 1) and measuring their activities in mixed N2 *C. elegans* cultures to confirm the presence of lysosomal activities in question and to evaluate the conservation of their function (Table 2).

Human disease	Mutant protein	Expect	identities (%)	positives (%)	*C.elegans* gene
Fabry	**Alpha galactosidase**	**2,3e - 87**	**44**	**60**	**R07B7.11**
Farber	Acid ceramidase	2,1e - 84	40	64	HK11D2.2
Gaucher	Beta glucocerebrosidase	1,2e - 106	43	63	Y4C6B.6
GM1 gangliosiosis	Beta galactosidase	5,6e - 116	41	58	T19B10.3
GM2 gangliosidosis					
Tay-Sachs HEXA	Alpha hexosaminidase	2,6E - 95	40	58	T14F9.3
Sandhoff HEXB	Beta hexosaminidase	9,1E - 84	44	63	T14F9.3
Mucolipidosis IV	Mucolipin	1,2E - 79	38	3	R13A5.1
Niemann-Pick A,B	Acid sphingomyelinase	2,0E - 86	34	56	B0252.2
Schindler	**N-acetylgalactosaminidase**	**3,7E - 87**	**45**	**60**	**R07B7.11**
Aspartylglycosaminuria	Aspartylglycosaminidase	1,7e- 68	46	61	R04B3.2
Fucosidosis	Alpha fucosidase	2,9e - 96	46	63	W03G11.3
Galctosialidosis	Cathepsin A	8,1e - 92	40	61	F41C3.5
alfa- mannosidosis	Alpha mannosidase	1,1e - 157	38	56	F55G10.1
beta-mannosidosis	Beta mannosidase	5,8e - 150	38	56	C33G3.4
MPSIIIB	Alpha N-acetyl glucosaminidase	4,5e - 98	36	54	K09E4.4
MPSIIID	N-acetylglucosamin sulfatase	2,3e - 91	43	60	K09C4.8
MPSVII	Beta glucuronidase	6,1e - 114	41	64	Y105E8B.TT
Pompe	**Alpha glucosidase**	**2,7E - 134**	**36**	**53**	**D2096.3**
Wolman	Acid lipase	3,2e - 81	42	61	Y57E12B.B
cystinosis	Solute carrier family 3 member 1	2,0e - 36	28	49	F2D10.9
sialic storage Salla	Solute carrier family 17 member 5	2,1e - 100	40	60	C38C10.2
CLN1	Palmitoyl-thioesterase 1	1,2E - 81	54	75	F44C4.5
CLN3	CLN3 protein	7,6e - 80	41	61	F07B10.1
Niemann-Pick C1	NPCi protein	8,7e - 87	28	47	F02E8.6

Table 1: Conservation of selected lysosomal proteins between man and *C. elegans* – Blastp search [278].

The initial results showed that there is a good conservation (orthologous genes) and function (enzymatic activity) of the selected lysosomal proteins between man and *C. elegans*, and that *C. elegans* may be a suitable model system for the study of LSDs. Based on the

highest similarity and type of LSD we resorted to further study *C. elegans* orthologs of α-galactosidase, α-N-acetylgalactosaminidase and α-glucosidase whose deficiencies lead to Fabry, Schindler and Pompe disease, respectively.

Enzyme	Activity v [nmol.mg^{-1}.h^{-1}]*
β-N-Acetyl-glucosaminidase	308
β-galactosidase	10
α-mannosidase	130
β-glucuronidase	2.8
β-glucosidase	55.3
β-glucosidase + CBE inhibition	4.6
α-glucosidase pH 4	**53**
α-glucosidase pH 6.5	**67.5**
α-galactosidase	**28.7**
α-galactosidase + GalNAc inhibition	**1**
α-N-Acetyl-galactosaminidase	**317**

Table 2: Activities of selected acid hydrolases in *C. elegans* – incubation at 25 ^{o}C, measured with methylumbelliferyl substrates [279].

5.1.1. *Caenorhabditis elegans* as a model organism for Fabry and Schindler disease (appended publication 1)

5.1.1.1 Blast search and verification of the predicted gene structure

We were able to find only one predicted open reading frame (ORF) orthologous to human α–GAL and α–NAGA by BLAST searches in the complete *C. elegans* genome [29]. It was designated R07B7.11 and it had significant sequence similarity to both human genes, which were previously proposed to evolve from the common ancestral gene (see sections 2.3.2.1 - 2.3.2.2). We have not found any other orthologous gene with significant similarity to either α–GAL or α–NAGA by repetitive searches. Similar results were obtained while searching *C.briggsae* genome [29]. R07B7.11 was later designated *gana-1* (α-GAL and α-NAGA ortholog (appended publication 1)). *Gana-1* consists of 5 exons and 4 introns and was initially annotated as a potential ortholog of human α–NAGA. Two EST clones for this predicted ORF were reported shortly before we started the study, the first in the third exon and the other in the last exon including the stop codon and a part of 5´untranslated region (Figure 3). Open reading frame sequence tag (OST) of the same structure but without UTRs

for *gana-1* appeared during the course of our work in the Worfdb database [27] and confirmed our findings [280].

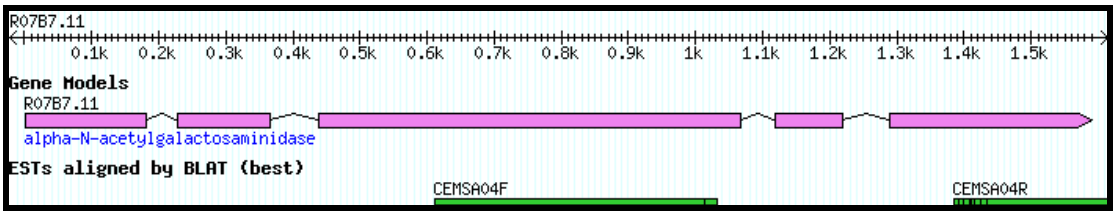


Figure 3: Schematic representation of *gana-1* gene structure (acquired from Wormbase at the beginning of our work [280])**.** The entire length of gDNA is 1681 bp and the spliced cDNA consists of 1356 bp and 91 bp of 3´UTR.

The sequence of the predicted gene was verified by sequencing of the PCR products from genomic DNA and cDNA. We analysed the whole coding region including 3´ and 5´ untranslated regions (UTR). The sequence of *gana-1* followed the predicted mRNA completely. The length of genomic DNA from start to stop codons is 1681 bp. The entire ORF of *gana-1* has 1356 bp and encodes a protein of 451 aminoacids. We have found no alternative splicing by RT-PCR, a feature similar to both human and mouse orthologs and we noted no signs of RNA editing in clones derived from the *gana-1* cDNA. The sequence analysis revelaed SL1 mode of *trans* splicing which suggests that the gene is transcribed either as the only or as the first gene in operon. Physical distances between *gana-1* and the two predicted downstream genes R07B7.12 and R07B7.13 suggest clustering into operon [40, 42]. We found that the gene R07B7.12 codes a predicted protein of unknown function and the gene R07B7.13 codes a nuclear hormone receptor (NHR-206) [281]. As this region has not been annotated as an operon [281] and there is no apparent connection among these three genes, it is unlikely that these genes form one. We did not further explore this question issue because it was not important for our study.

5.1.1.2 Prediction of GANA-1 signal peptide

GANA-1 signal sequence prediction was performed by SignalP server [254]. Prediction result revealed high N-terminal signal peptide prediction probability (1.000) and two possible signal sequence cleavage sites (between residues 15-16 or 21-22).

5.1.1.3 Biochemical studies

We found both enzyme (α–GAL and α–NAGA) activities in the homogenates from *C. elegans* N2 strain using methylumbelliferyl (MU) substrates [279]. The activity of α–NAGA measured after incubation at 37 °C was 430 nmol.mg-1.h-1 with MU-α–N-acetylgalactosaminide as a substrate and the activity of α–galactosidase A using MU-α–galactopyranoside was 43 nmol.mg-1.h-1. Alpha–NAGA and α–GAL activities measured after incubation at 25 °C were 317 and 28.7 nmol.mg-1.h-1 respectively (Table 2). Alpha–NAGA activity predominated over α–GAL activity, which was about 10% of that of α–NAGA.

In accordance with published observations in human enzymes, α–NAGA activity is inhibited by both inhibitors N-acetyl-D-galactosamine (D-GalNAc), and D-galactose (D-Gal). On the other hand, α–GAL activity is inhibited only by D-Gal while D-GalNAc has no inhibitory effect on its activity [282]. We reasoned that if *gana-1* had activity toward both substrates, MU-α–D-galactoside and MU-α–N-acetylgalactosaminide, and only one active site, then both inhibitors - D-GalNAc and D-Gal, respectively, would inhibit both activities.

We performed inhibitory assays in the homogenates from *C. elegans* N2 strain based on this assumption. In the case of α–galactosidase A, the degradation of the MU-α–galactopyranoside was significantly inhibited (up to 95%) in the presence of D-GalNAc, whereas, in the presence of D-Gal the degradation of the same substrate was inhibited up to 75%. In the assay of α–NAGA, the degradation of the MU-α–N-acetylgalactosaminide was inhibited up to 97% by D-GalNAc and up to 90% by D-Gal. No inhibition of α–NAGA and α–GAL A activity by D-glucose was observed.

We observed significant effect of both inhibitors on both activities and mainly the strong inhibitory effect of D-GalNAc on the α–GAL activity, which is not present in human α–GAL. These results support the hypothesis that *C. elegans* has only one enzyme with both α–GAL A and α–NAGA activities and can have only single active site per monomer of the protein, as was predicted. However, because these measurements were not performed with the pure enzyme, they provide only indirect evidence for the above-mentioned hypothesis.

5.1.1.4 RNA-mediated interference

To evaluate the function of *gana*-1 we performed RNA interference (RNAi) experiments. RNAi assays were directed against the complete coding region of *gana-1*.

Initially we used either microinjection or feeding approach of RNAi separately. However, we were not able to observe any abnormal morphological phenotype. We thus decided to use *C. elegans* RRF-3 strain, which is hypersensitive to RNAi [257], but similarly as in N2 strain, RNAi have not revealed any abnormal morphological phenotype. In order to foster RNAi we resorted to combination of microinjection and feeding approaches and biochemically evaluated both activities of *gana-1*. Measurement of *gana-1* activity in four individual experiments (thousands of individual worms) showed simultaneous decrease of both α–GAL and α–NAGA activities in interfered worms (Table 3). In all RNAi experiments both α–GAL and α–NAGA activities decreased proportionally, usually by tens of percent of activity of appropriate controls. The well-known variability in the efficiency of RNAi [8] explains the differences between individual experiments. The activity of β-hexosaminidase, which was used as the control enzyme, did not differ between the control and interfered worms. These results support the specificity of *gana-1* RNAi and our hypothesis that the *gana-1* has both enzyme activities.

Experiment No.	RNAi	**α-GAL activity (nmol/mg*h)**		**α-NAGA activity (nmol/mg*h)**	
		activity	% of control	activity	% of control
I.	***gana-1***	1,2	**67**	26,6	**50**
	control	1,8	100	53,1	100
II.	***gana-1***	11,1	**84**	195,1	**88**
	control	13,3	100	221,7	100
III.	***gana-1***	1,0	**30**	11,8	**19**
	control	3,4	100	61,7	100
IV.	***gana-1***	2,9	**30**	50,7	**24**
	control	9,6	100	212,1	100

Table 3: Alpha-GAL and α-NAGA activities after RNAi of *gana-1*. Table shows parallel decrease of both α-GAL and α-NAGA activities after *gana-1* RNAi compared to control. All measurements were done at 37 ^{o}C.

Both enzyme activities were lower in interfered and control worms cultured on the bacterial strain HT 115 [8] compared to N2 strain cultured on the OP50 strain. We explain this variability by different culture conditions and by the stress caused by the injection and the presence of the plasmid DNA.

To find if the *gana-1* RNAi influences function of coelomocytes (see later) such as uptake of fluid from pseudocoelom and degradation of endocytosed proteins we performed *gana-1* RNAi in *C. elegans* GS1912 strain [50]. GS1912 animals are transgenic worms that express green fluorescent protein (GFP) in body wall muscles. The GFP is secreted into the

pseudocoelomic cavity and consecutively degraded predominantly by the coelomocytes. However, we did not observe any difference between RNAi-treated and control GS1912 worms. These results suggest that *gana-1* RNAi does not affect endocytosis and degradation function of coelomocytes.

RNAi of *gana-1* yielded no abnormal morphological phenotypes, which was not surprising due to variable efficiency of RNAi and well-known fact that even low level of lysosomal hydrolase activities do not result in lysosomal storage. It was reported that only very low residual enzyme activities (about 1 to 2 % of normal activity) lead to the lysosomal storage [283], so the maximum decrease of enzyme activities to the level of 20 % of control activities in the *gana-1* RNAi-treated worms is expected not to cause lysosomal storage phenotype.

5.1.1.5 Expression of *gana-1*

We created transgenic worms carrying expression construct containing the complete coding region of *gana-1* C-terminally tagged with green fluorescence protein (GFP) in order to study the expression of *gana-1* in *C. elegans*. *Gana-1*::GFP was expressed under the control of a 3 kb region of its hypothetical promoter (Figure 4). Although we verified the presence of the *gana-1*::GFP transgene in the transgenic worms on the level of gDNA, cDNA (Figure 5) and protein (Western blotting) (Figure 6), we were not able to observe any GFP signal by fluorescence microscopy under the standard laboratory conditions. Because Western blotting showed the presence of fusion protein of expected size, we assumed that the absence of GFP signal resulted from pH-dependent quenching of GFP fluorescence emission [284].

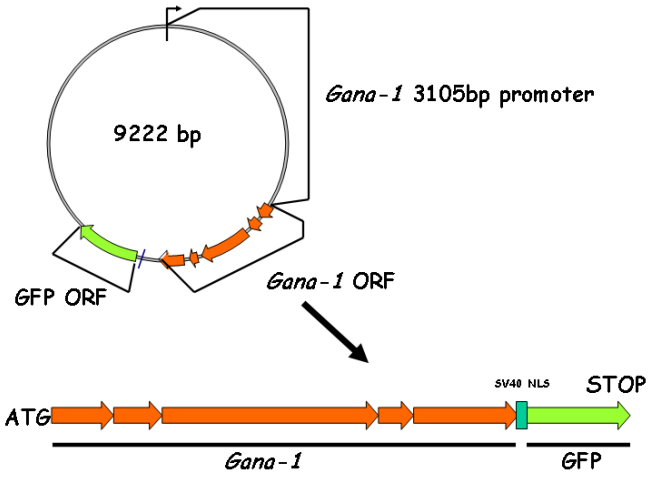


Figure 4: Schematic representation of *gana-1*::GFP extrachromosomal expression construct with appropriate RNA transcript.

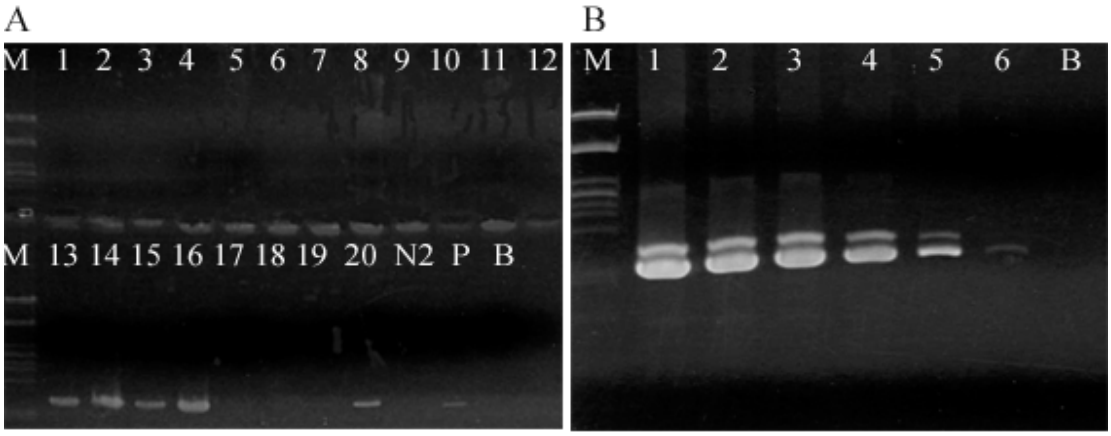


Figure 5: Verification of *gana-1*::GFP transgene presence in transgenic worms on the DNA (A) and RNA (B) levels
M – Lambda DNA/*PstI* Marker, N2 - N2 Bristol strain (negative control), P - *gana-1*::GFP construct (positive control), B - blank
(A) 1-20 – screening of transgenic worms (single worm PCR)
(B) 1-6 – cDNA of transgenic worms (temperature gradient)

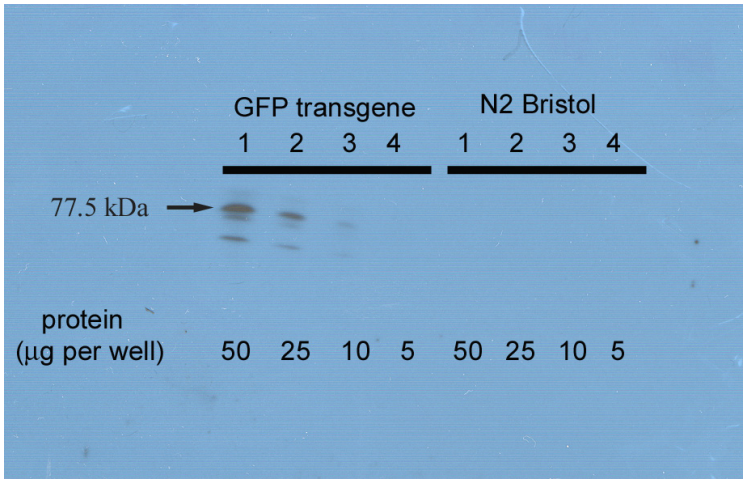


Figure 6: Verification of presence of GANA-1-GFP fusion protein in homogenate of transgenic worms by Western blotting. The size of the GANA-1-GFP fusion protein is 77.5 kDa. N2 Bristol strain was used as negative control.

We used two agents [90, 264] that specifically alkalize acidic cellular compartments in order to enhance the GFP emission and to confirm that the absence of GFP signal was due to the quenching of fluorescence by low pH in the acidic cellular compartment. Worms carrying *gana-1*::GFP transgene were soaked in NH_4Cl or concanamycin A (CON A) solutions. After soaking, we observed a distinct GFP signal in the membrane bound compartment of endocytically active coelomocytes in a small proportion of worms (3–5% of the population) (Figure 7). Intensity of the GFP signal was dependent on the concentration and the time of

incubation of the alkalizing reagent used. We were able to observe the first visible GFP signal after 8 hours of incubation in 100 mM NH_4Cl and already after 2 hours of incubation in 50 nM CON A. We did not see any GFP signal while using lower concentrations of both NH_4Cl and CON A even after prolonged incubations. We consider the reappearance of the GFP signal after treatment of the transgenic worms with compounds increasing pH in acidic cellular compartments as indirect proof of lysosomal localisation of the fusion protein. The GFP signal in coelomocytes had the same coarsely granular pattern as that observed after immunostaining (discussed below).

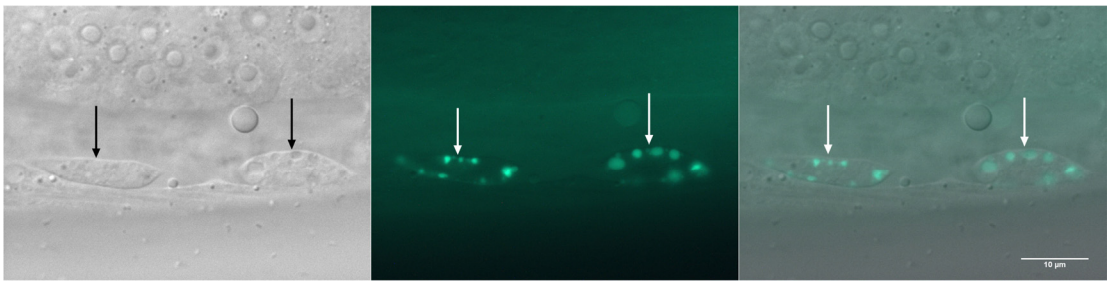


Figure 7: Alkalization of *gana-1*::GFP transgenic worms. Visible GFP signal in vesicular compartments of two coelomocytes (arrows) after CON A treatment (50 nM CON A, 24 hours incubation). DIC, fluorescence and merge image, respectively. Scale bar 10 μm.

Based on these results we decided to detect the transgene product indirectly by GFP antibody in order to better characterise the tissue and intracellular distribution of the fusion protein. Immunofluorescence detection of GFP fusion protein showed a specific and coarsely granular cytoplasmic pattern of transgene expression limited either to body wall muscle cells and sometimes also to coelomocytes (30% of population) or joint body wall muscle, intestinal cells and coelomocytes expression pattern (3-5% of population) (Figure 8). The staining pattern corresponds well with the results of GFP detection in NH_4Cl and CON A experiments. The differences between the results of immunofluorescence and alkalization studies can be explained by limited access of alkalizing agents to other tissues than coelomocytes. This is not surprising because of the well-known function of coelomocytes to actively endocytose fluid from the pseudocoelom (see 2.1.5.2).

We observed the mosaic expression of the transgene in approximately 30% of the population, which corresponded to the common expression efficiency of extrachromosomal array transgenes [285]. Co-localization study with the intestinal autofluorescence granules previously described as secondary lysosomes [286] was hampered by a significant decrease of

inherent intestinal granular autofluorescence caused by the immunofluorescence staining procedure.

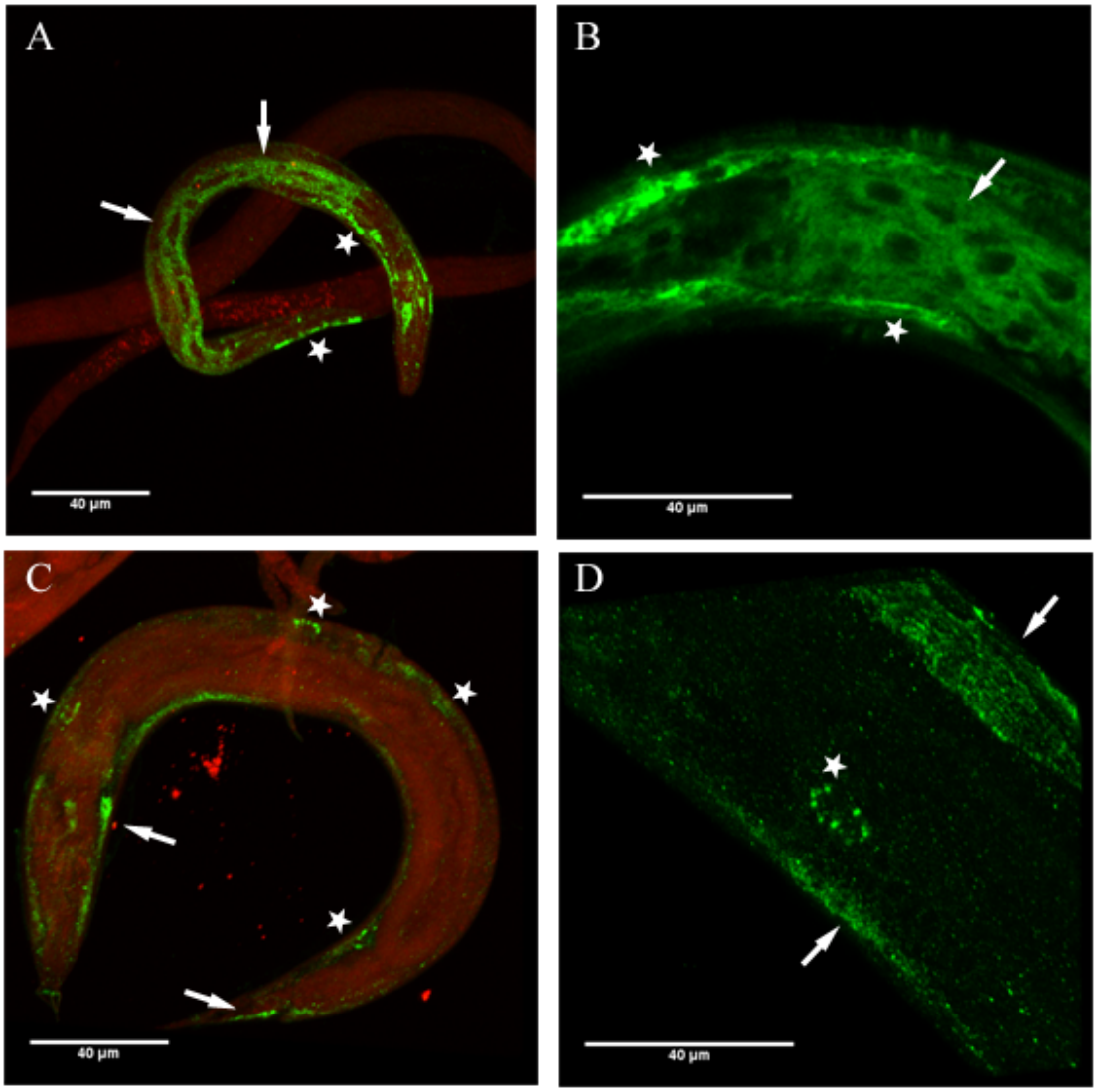


Figure 8: Immunofluorescence detection of GANA-1-GFP fusion protein. (A) and (B) Coarsely granular cytoplasmic expression pattern in intestinal (arrows) and body wall muscle (asterisks) cells. (C) and (D) Identical cytoplasmic expression pattern limited to body wall muscle cells (arrows) and coelomocytes (asterisks). Image (A) show two non-transgenic worms in the background. (A), (C) and (D) images were acquired by 3D rendering of original confocal Z-stacks. Scale bars represent 40 µm.

5.1.1.6 Bioinformatic studies on *gana-1*

This part of the thesis was performed in collaboration with Robert Dobrovolny and Jakub Sikora (Charles University in Prague, 1st Faculty of Medicine, Institute of Inherited Metabolic Disorders).

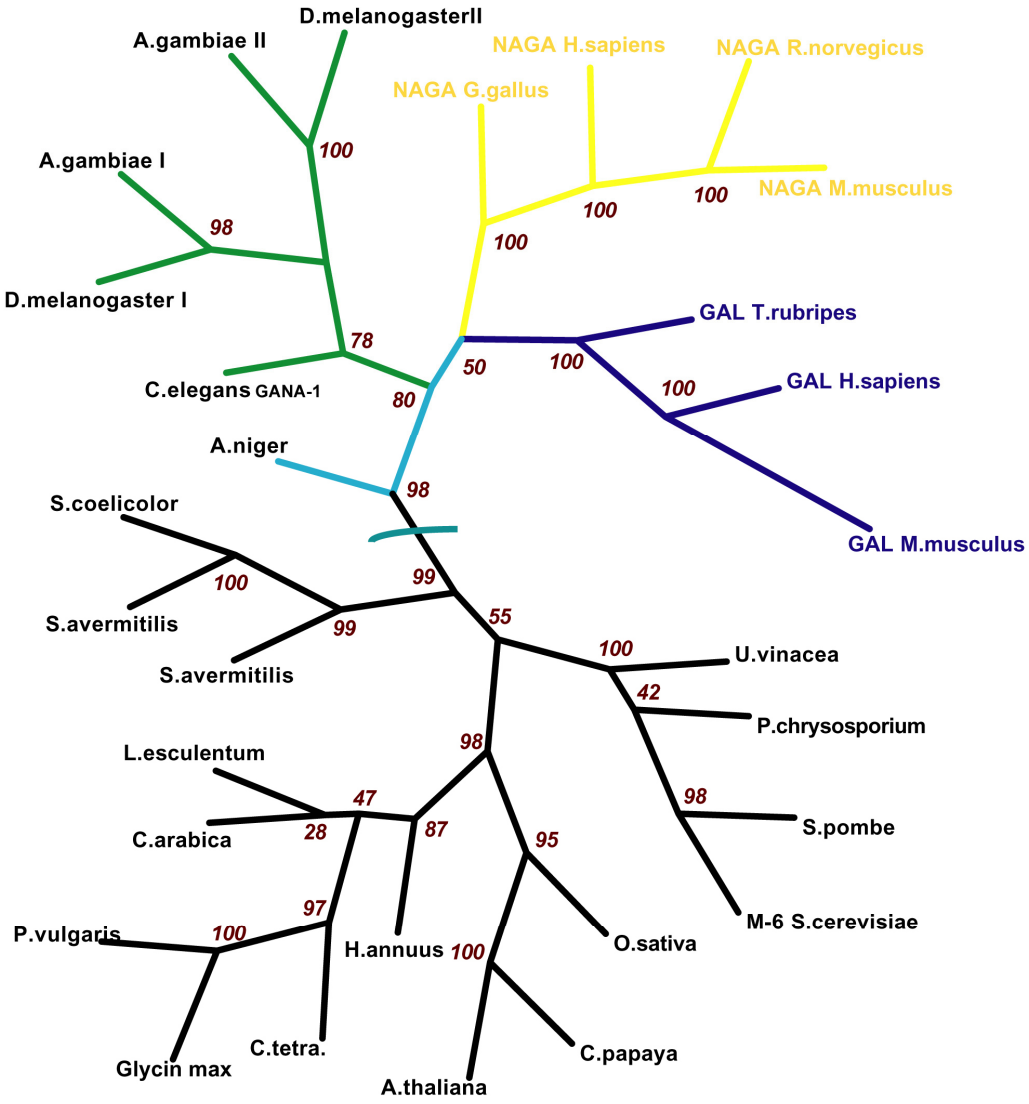


Figure 9: Phylogenetic tree of GANA-1 and other members of melibiase family constructed using maximum likelihood method. Bootstrap values at the nodes evaluate the tree (maximum 100). Cladogram divides proteins into four groups: animal NAGAs (yellow), animal GALs (dark blue), plant/lower organism GALs (black) and *Caenorhabditis/Drosophila/Anopheles* group (green).

The alignment of GANA-1 protein sequence with other selected melibiase (GH27) [190, 191] family members showed striking similarity. GANA-1 had the highest sequence similarity with *Anopheles gambiae* GAL (46,1%) and the lowest similarity with *Streptomyces avertimilis* GAL (22,4 %). The phylogenetic tree (Figure 9) was constructed on the basis of multiple protein alignment. The phylogenetic analysis identified four main clades: animal NAGAs, animal GALs, plant/lower organisms GALs and clade containing sequences of *Drosophila melanogaster, Anopheles gambiae* and *C. elegans*. The branch containing GANA-1 is positioned between higher animal GALs and NAGAs and plant/lower organisms GALs.

Based on these results we suggest that *gana-1* is an ancestor of animal GALs and animal NAGAs. However, the presence of a pair of genes in *Drosophila* and *Anopheles* genomes that are annotated as α–GAL and α–NAGA is not in full agreement with this conclusion. The presence of these genes in the *Caenorhabditis/Drosophila/Anopheles* branch instead of GAL and NAGA clades could be explained by low divergence of these sequences from common ancestral gene or by independent gene duplication in *Drosophila/Anopheles* organisms. Based on the phylogenetic analysis we conclude that *gana-1* indeed evolved from the common ancestor of α–NAGA and α–GAL enzymes.

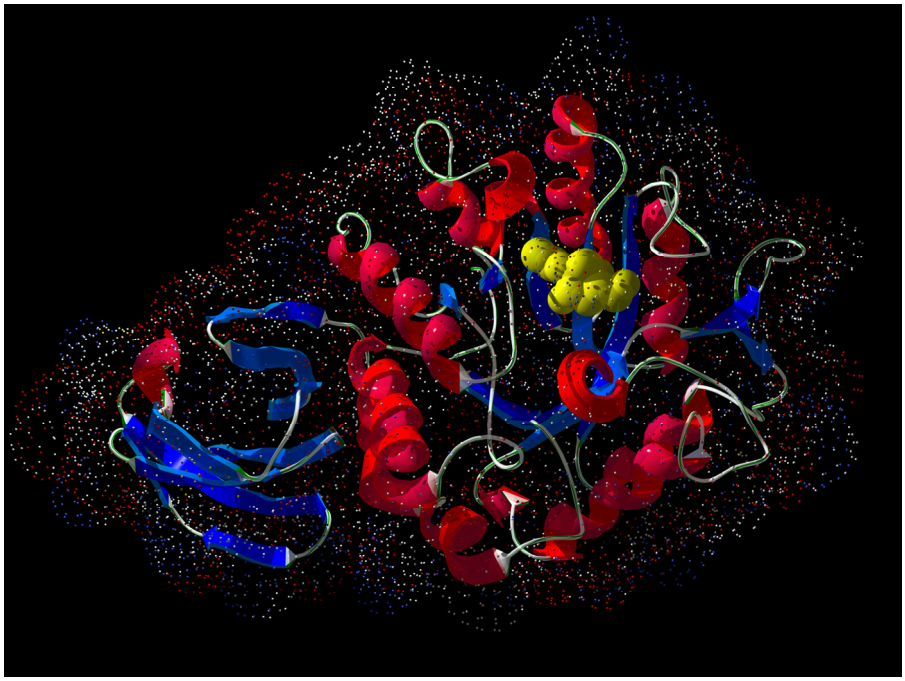


Figure 10: 3D model of GANA-1 monomer based on X-ray structures of chicken α-NAGA [212], rice α-GAL [194] and human α-GAL [193]. Inhibitor N-acetyl-D-galactosamine is placed into the active site (yellow).

Consecutively we performed homology modeling of GANA-1 protein molecule based on crystallographic structures of chicken α–NAGA [212], human α–GAL A [193], and rice α–GAL A [194]. The structural model of GANA-1 has two-domain structure alike the templates (Figure 10). Domain I, which contains predicted active site, adopts $(\alpha/\beta)_8$ barrel structure which represents common motif in many glycoside hydrolases. Less conserved domain II is β domain with β sandwich structure containing Greek key motif. The active site pocket is formed by the same twelve amino acids (W31, D76, D77, Y118, C126, K152, D154, C156, S186, A189, Y190 and R211) as in chicken α–NAGA. This finding infers that GANA-

1 shares the identical catalytic reaction mechanism with chicken α–NAGA [212]. The active site residues are well conserved between α–GALs and α–NAGAs. They differ only in two residues, which discriminate between α–GAL and α–NAGA substrate [189]. In contrast to α–GALs, which do not have α–NAGA activity and are not inhibited by N-acetyl-D-galactosamine because N-acetylated substrates cannot enter their smaller active sites, α–NAGAs can accommodate α–galactose and have some α–GAL activity. The appropriate residues of GANA-1 are corresponding to α–NAGAs, which are consistent with our previous results. For additional details see appended publication 1.

Multiple alignment, phylogenetic analyses and homology modeling employed GANA-1 sequence verified by the above studies.

5.1.1.7 Expression of *gana-1* in eukaryotic expression system (unpublished data)

To characterize the function and expression of GANA-1 *in vivo* and to support our hypothesis that GANA-1 has both α–GAL and α–NAGA activities, we decided to express *gana-1* in normal mouse fibroblasts and mouse fibroblasts deficient in either α–Gal or α–Naga. Our second aim in this part of the study was to find if GANA-1 compensated α–GAL and α–NAGA activities in mammalian cells.

We prepared six expression constructs containing the complete coding region of *gana-1*. These were expressed under the control of CMV promoter in the pCMV-Tag 1 Epitope Tagging Mammalian Expression Vector (Stratagene). The first construct was double labeled with N-terminally tagged FLAG and C-terminally tagged c-myc, the second and the third were tagged either N- or C-terminally with FLAG. The remaining three constructs were internally tagged with FLAG in different loops of the *gana-1* sequence (sites selected on the basis of the modelled structure) (Figure 11).

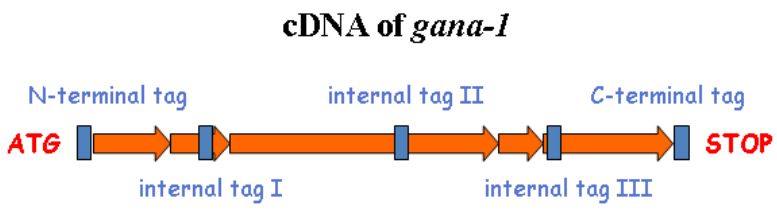


Figure 11: Schematic representation of localisation of tag(s) (Flag and/or c-myc) in the sequence of cDNA of *gana-1* used for expression study in mouse fibroblasts.

Preliminary results of transfection verified functionality of all six expression constructs and we observed granular cytoplasmic pattern of *gana-1* overexpression in all three types of mouse fibroblasts (wild type and both knockouts) (Figure 12). We did not seen any FLAG signal in the fibroblasts transfected with the constructs N-terminally tagged with FLAG which was not surprising due to predicted N-terminal signal peptide that is cleaved off. As is demonstrated in the Figure 12A, we were able to detect only c-myc signal in the double immunolabeled fibroblasts carrying construct with N-terminally FLAG and C-terminally c-myc tags. We did not see any significant difference between expression of individual constructs and also any difference between expression in the wild type and both knockout mouse fibroblasts. Unfortunately, colocalization studies with Lamp 1 did not provide confirmation of the lysosomal localization of the overexpressed *gana-1* (Figure 13). We observed signal of overexpressed *gana-1* disperse in the cytoplasm. Most of the overexpressed *gana-1* was probably accumulated in endoplasmic reticulum but we did not confirm this assumption by colocalization studies using markers of this organelle.

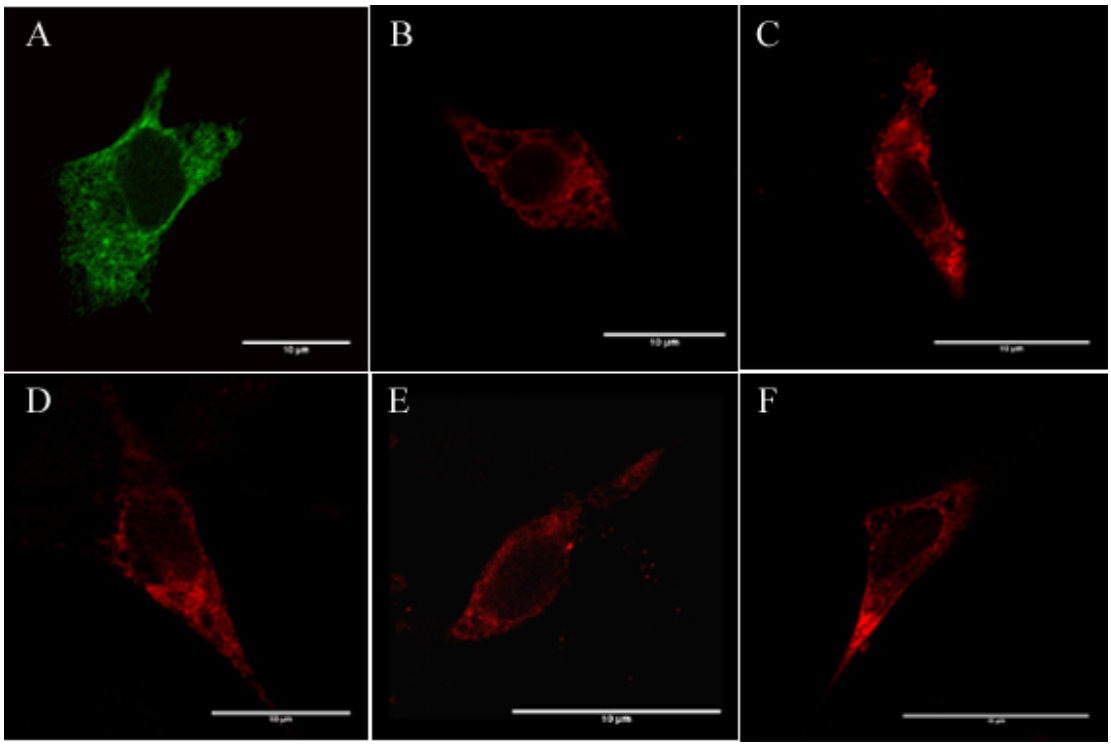


Figure 12: Immunofluorescence detection of *gana-1* fusion protein in mouse fibroblasts. All images show granular cytoplasmic pattern of *gana-1* overexpression. Used expression constructs: (A) double tagged with N-terminal Flag and C-terminal c-myc, (B) and (F) tagged with C-terminal Flag, (C) internal Flag – I, (D) internal Flag – II and (E) internal Flag – III. (A)-(E) wild type mouse fibroblasts, (F) fibroblasts deficient in α-NAGA. Flag signal is shown in red and c-myc in green. Scale bars represent 10 μm.

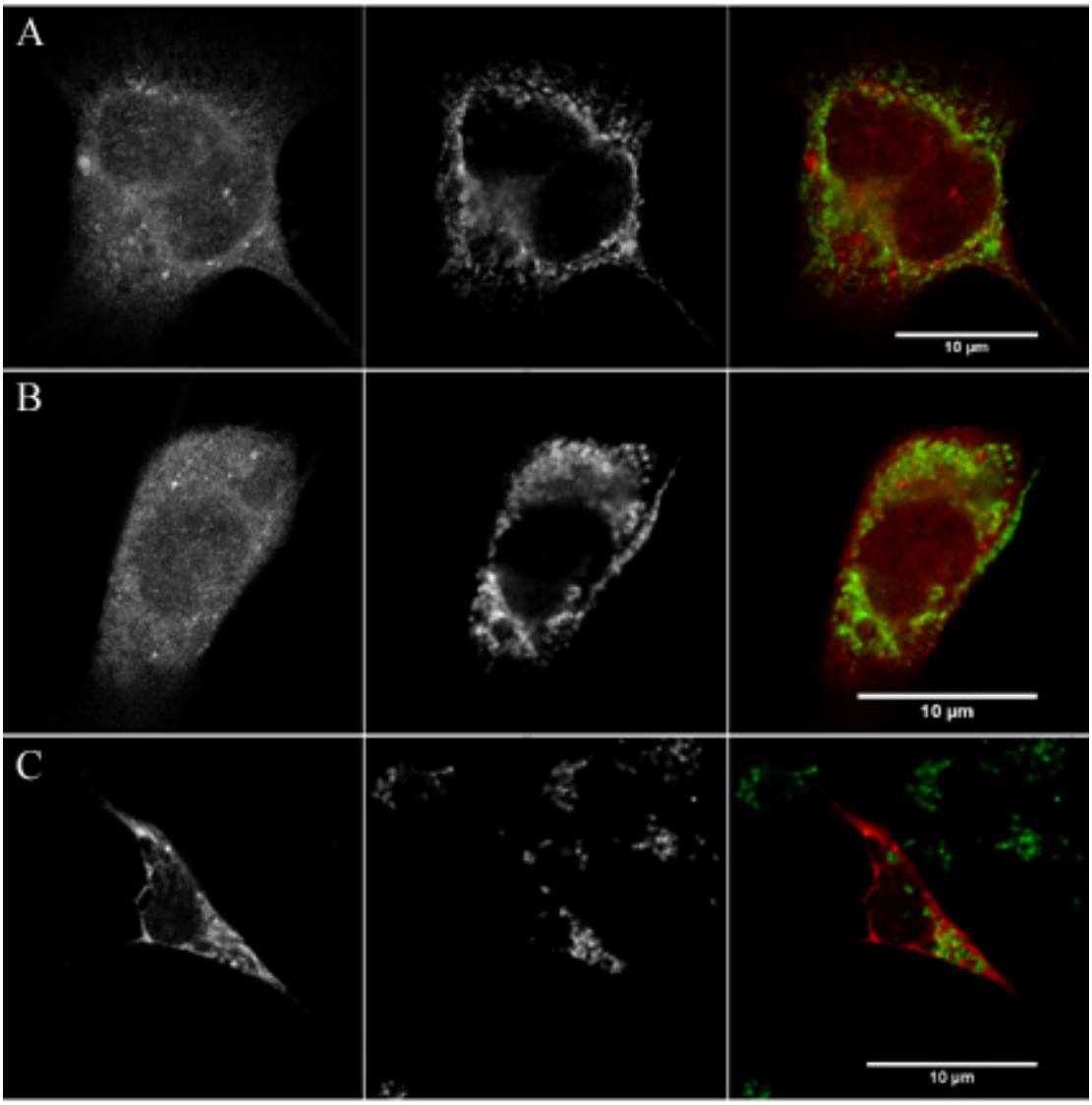


Figure 13: Colocalization studies of *gana-1* fusion protein and lysosomal marker Lamp-1 in transiently transfected mouse fibroblasts. Expression construct with C-terminal Flag tag. (A) and (B) mouse fibroblasts deficient in α-GAL, (C) wild type mouse fibroblasts. Flag signal is shown in red and Lamp-1 in green. Scale bars represent 10 µm.

Because of low efficiency of transient transfection (only about 2%), we tried to obtain stable cell lines expressing *gana-1* under the selection of antibiotic geneticin (G418). The presence of the *gana-1* transgene in the mouse fibroblasts was confirmed on the level of genomic DNA and in case of *gana-1* C-terminally tagged with FLAG also on the level of mRNA (cDNA). However, we were not able to confirm the presence of GANA-1 fusion protein by Western blotting or immunosfluorescence assays with antibody against Flag or c-myc tags. We tried to cultivate the stable cell lines with therapeutic chaperones (see section 2.3.1. – last paragraph), such as D-galactose (200mM), N-acetyl-D-galactosamine (100mM) and deoxygalactonojirimicine (DGNJ, 20µM) for three to four days, but we did not detect any

GANA-1 expression by Western blotting or immunofluorescence assays. We can only hypothesize about the reasons of low GANA-1 expression and/or its enhanced degradation. Our findings can also be attributed to different glycosylation patterns [69] and/or dissimilar signal peptides (see Figure 2 in appended publication 1) in mammalian and *C. elegans* cells.

5.1.2. *Caenorhabditis elegans* as a model organism for Pompe disease (appended submitted manuscript)

5.1.2.1 Blast search and sequence verification of predicted genes

We found by repetitive BLAST searches for *C. elegans* orthologs of human GAA in the Wormpep database four predicted coding sequences or sequences annotated as confirmed by partially overlapping EST clones - D2096.3 (acid alpha glucosidase related, *aagr-1*), R05F9.12 (*aagr-2*), F40F9.6 (*aagr-3*), F52D1.1 (*aagr-4*) and one pseudogen F53F4.8. All four found coding sequences with their overall similarity of the amino acid sequences with the human GAA sequence were annotated as glycoside hydrolase family 31 (GH31) [190, 191] members. We found no predicted or confirmed double active site proteins (such as SUIS or MGA in humans) in the *C. elegans* genome.

We verified the available *in silico* predictions by sequencing of all four genes. The cDNAs were amplified in overlapping PCR fragments including 5´trans spliced ends and 3´untranslated regions including polyadenylation sites (Figure 14). Gene organization (exon/intron boundaries) and the lengths of all four transcripts generally corresponded with the database predictions. In comparison to the available database predictions which annotated two alternative splicing variant of *aagr-3* - F40F9.6a and b[22], we were able to detect only a single splicing variant of *aagr-3* corresponding to F40F9.6b. Because we were not able to amplify any PCR product of the predicted gene F53F4.8, we supposed that F53F4.8 is a pseudogene. This assumption was lately supported by Wormbase (Wormbase Realase WS189).

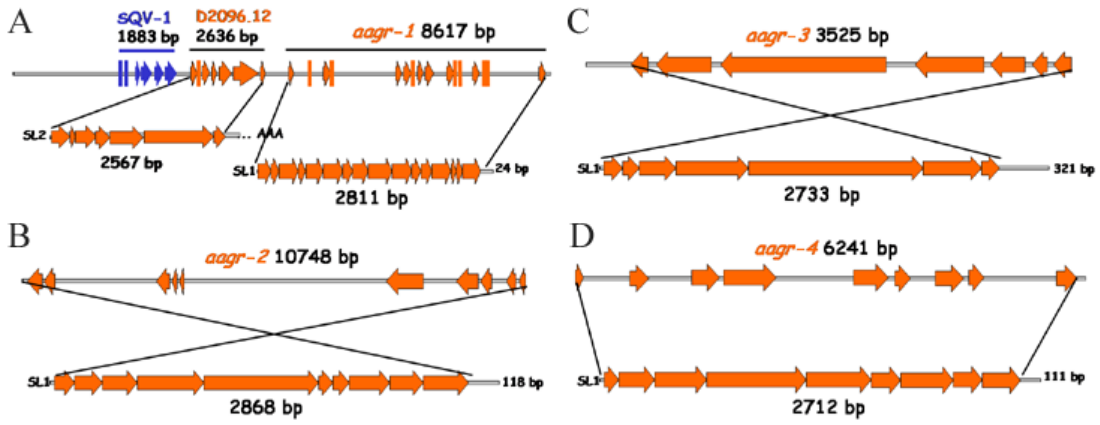


Figure 14: *Aagrs* gene structures. Schematic representations of (A) operon CEOP4284 and *aagr-1*, (B) *aagr-2*, (C) *aagr-3* and (D) *aagr-4* gene structure. The size of the gDNA is shown above the gene structure and the length of the cDNA without the 3´ UTR (shown on the right side of transcript) is given below the transcript.

We found that all four genes (*aagr-1-4*) contain SL1 *trans* splicing element despite short distance (approx. 650 bp) between *aagr-1* and upstream (D2096.12) gene which belongs to CEOP4284 transcriptional operon [43]. The first gene in the CEOP4284 operon is UDP-glucuronic acid decarboxylase (*sqv-1*) [287] which is *trans* spliced in SL1 mode. The sequence analysis of the second gene D2096.12 in the operon verified its gene organization and complete sequence, which was in full agreement with the available prediction and revealed SL2 mode of *trans* splicing. The search in the available protein databases for potential orthologs of D2096.12 revealed the *Caenorhabditis briggsae* gene BP:CBP19293 of unknown function as the highest scoring ortholog. Other less scoring orthologs were mostly gene products of unknown function. Based on these findings, genomic relationship of CEOP4284 operon and *aagr-1* represents local clustering of genes involved in glycosaminoglycan neogenesis (*sqv-1*) and GH31 member involved in carbohydrate catabolism. Based on these results we can only hypothesize about the nature of this operonic function clustering.

5.1.2.2 Signal peptide and intracellular targeting predictions

SignalP predictions revealed presence of N-terminally situated signal peptides with strong probability for all AAGRs (0.996 for AAGR-1, 1.000 for AAGR-2, 0.911 for AAGR-3 and 0.999 for AAGR-4). The lengths of signal peptide sequences correspond with the signal sequence cleavage sites: 22-23 (AAGR-1), 18-19 (AAGR-2), 19-20 or 22-23 (AAGR-3) and 16-17 (AAGR-4). TargetP server predicted all AAGRs as non-secretory and non-mitochondrial proteins.

5.1.2.3 Biochemical studies

GAA activities were assessed in the homogenates from *C. elegans* N2 strain when measured with methylumbelliferyl substrates at two different pH values (4.0 and 6.5) and two different incubation temperatures (25 °C and 37 °C). We assumed that values acquired by measuring according to this protocol at acidic pH represent sum of enzymatic activities of several enzymes that are predicted in *C. elegans*. It is especially maltase-glucoamylase (MGA) that interferes with direct GAA measurement in humans. The overall values measured at 25 °C were 53 and 67.5 nmol $mg^{-1}h^{-1}$ for pH values 4.0 and 6.5 respectively. Activity values measured at 37 °C were 165 and 277 nmol $mg^{-1}h^{-1}$ respectively.

It has been reported that addition of acarbose [226], which is a potent inhibitor of maltase-glucoamylase and sucrase-isomaltase, to the reaction mixture at acidic pH selectively eliminates these interfering activities in the GAA assay. Thus values measured with acarbose at acidic pH should directly reflect acid alpha glucosidase activity. On the other hand the effect of acarbose on the neutral glucosidases has been insignificant. Tables 4-6 show influence of acarbose on GAA activities.

In order to differentiate AAGRs functioning at the acid pH from the ones acting at neutral pH we further performed selective RNAi experiments and deletion mutant analyses as described below.

5.1.2.4 RNA-mediated interference

To evaluate the impact of RNAi on the *C. elegans* phenotype and to characterize the worms GAAs we performed selective separate RNA interference experiments of all four *aagr* genes. We selected the sequence regions for RNAi assays on the basis of the cDNA multiple alignment. To avoid or at least minimize cross-interference between otherwise homologous sequences we tried to find and employ regions with the lowest similarity/identity as possible.

Equally as in case of other lysosomal hydrolase ortholog *gana-1* (appended publication 1) we were not able to observe any morphological or other observable phenotype despite extensive and very detailed examination of the interfered animals. In addition to the all four *aagrs*, we performed RNAi assay directed against the *aagr-1* upstream gene D2096.12 with the same result. The minimal RNAi phenotype was very probably caused by high residual enzymatic activity (see section 5.1.1.4).

Because of the absence of morphological changes and our previous experiences with *gana-1* we resorted to biochemical evaluation of the impact of RNAi treatment on the GAA activities in interfered worms. Measurement of GAA activities at two distinct pH (4.0 and

6.5) and parallel acarbose inhibition showed separation of the neutral and acid glucosidase activities to different AAGRs (Tables 4 and 5). We showed the significant decrease of GAA activity in case of *aagr-1-3* RNAi experiments. The decrease of GAA activities after RNAi of *aagr-4* was minimal and together with results of analysis of *aagr-4* deletion mutant, we suppose that AAGR-4 enzymatic potential is limited. We found the most relevant drop of activity values after RNAi of *aagr-2* at the acidic pH and in case of *aagr-3* RNAi at neutral pH, up to the levels of 20 % of control values. Glucosidase activities after RNAi of *aagr-1* was decreased more dramatically in the acidic pH (average value 68% of controls) compared to neutral pH (86% of control values). These changes were less dramatic compared to decrease of *aagr-2* and *aagr-3* activity after relevant RNAi.

We observed variable influence of acarbose on the GAA activity (Tables 4 and 5) in the individuạl RNAi experiments. These experiments suggested AAGR-2 as the acarbose inihibited acidic AAGR and AAGR-1 as the least acarbose sensitive acidic AAGR. This observation was further supported by the analysis of *aagr-1* deletion mutant (see later). Based on this finding we consider AAGR-1 the most probable true ortholog of human acid α-glucosidase.

Experiment No.		RNAi	α-glucosidase activity [nmol/$mg_{protein}$*h]							
			pH 4 acarbose -		**pH 4 acarbose +**		**pH 6,5 acarbose -**		**pH 6,5 acarbose +**	
			activity	% of control	activity	% of control	activity	% of control	activity	% of control
I.	**acidic**	***aagr-1***	141	**54**	10	**15**	252	**94**	215	**103**
		aagr-2	26	**10**	20	**31**	188	**70**	183	**88**
		control	261	**100**	64	**100**	268	**100**	208	**100**
V.	**neutral**	***aagr-3***	178	**106**	40	**103**	62	**25**	19	**9**
		aagr-4	166	**99**	37	**94**	203	**83**	157	**76**
		control	167	**100**	39	**100**	246	**100**	206	**100**

Table 4: Influence of acarbose on activities of all AAGRs in selected RNAi experiments. Table shows variable impact of acarbose on GAA activities measured at acidic (4.0) or neutral (6.5) pH after selected RNAi experiments of separate *aagr-1-4* compared to controls. Influence of acarbose on neutral activities of AAGR-3 and -4 was insignificant. All measurements were done at 37 ^{o}C with 8 μM acarbose.

Experiment No.	RNAi	α-glucosidase activity [nmol/mg$_{protein}$*h], pH 4,0					
		acarbose -		acarbose +		% of activity inhibited by acarbose after RNAi	
		activity	% of control	activity	% of control		
I.	***aagr-1***	119	**55**	18	**26**	85	acidic
	aagr-2	21	**10**	16	**23**	***22***	
	control	216	**100**	70	**100**	68	
II.	***aagr-1***	83	**61**	12	**31**	85	
	aagr-2	19	**14**	13	**32**	***32***	
	control	137	**100**	40	**100**	71	
III.	***aagr-1***	143	**129**	21	**76**	86	
	aagr-2	30	**27**	20	**72**	***35***	
	control	110	**100**	27	**100**	76	
IV.	***aagr-1***	104	**80**	17	**44**	84	
	aagr-2	34	**26**	22	**57**	***34***	
	control	129	**100**	38	**100**	70	
V.	***aagr-3***	143	**99**	46	**109**	68	neutral
	aagr-4	136	**95**	41	**99**	70	
	control	144	**100**	42	**100**	71	
VI.	***aagr-3***	165	**98**	53	**114**	68	
	aagr-4	155	**92**	44	**95**	71	
	control	168	**100**	47	**100**	72	
VII.	***aagr-3***	198	**104**	64	**115**	68	
	aagr-4	205	**107**	61	**109**	70	
	control	191	**100**	56	**100**	71	
VIII.	***aagr-3***	160	**99**	58	**114**	64	
	aagr-4	154	**95**	46	**91**	70	
	control	162	**100**	51	**100**	69	

Table 5: GAA activities after separate RNAi of *aagr-1-4* and effect of acarbose. Table shows GAA activities after eight RNAi experiments measured at acidic pH with or without 8 µM acarbose. The influence of acarbose is demonstrated by the final column. The biggest inhibition effect of acarbose was observed after RNAi of *aagr-1*. All measurements were done at 37 °C.

5.1.2.5 Isolation and characterization of the *aagr-1* and *aagr-4* deletion mutants

Deletion mutants of *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) were kindly generated by *C. elegans* Gene Knock-out Consortium [259]. We used the PCR primer data provided by the *C. elegans* Gene Knock-out Consortium to analyse the extent of the deletions.

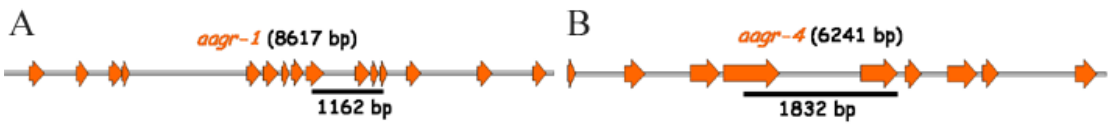


Figure 15: Schematic representation of deletions in the *aagr-1* (RB1790 deletion strain) and *aagr-4* (RB1307 deletion strain) gene. The position and extent of the (A) RB1790 and (B) RB1307 deletions are represented below the relevant gene. The size of the gDNA (in the brackets) is shown for comparison.

The sequence analysis of the gDNA of the *ok2317* strain disclosed 1162 bp long deletion that leaded to the loss of 224 C-terminal base pairs of the exon 9, complete introns 9, 10, 11 and exons 10, 11 and 65 base pairs from intron 12 (Figure 15A). The *ok2317* deletion leads to the loss of 218 amino acids of the primary structure of the AAGR-1 protein. Majority of the missing amino acids (205) form the active domain and cover α4-helix and final 4 βα repetitions in the $(\beta\alpha)_8$–core barrel structure according to the multiple amino acid alignment (see Figure 24 and 25).

Detailed microscopic analysis of the *ok2317* animals did not reveal any abnormal morphological or other observable phenotype. We performed GAA activity measurements at pH values 4.0 and 6.5 and parallel acarbose inhibition in the back-crossed nematode cultures homogenates of *ok2317* strain and the standard N2 Bristol strain in order to evaluate the impact of the deletion on the enzymatic activity of this predicted acidic activity glycosyl hydrolase. We observed that acidic GAA activity significantly decreases as compared to standard N2 Bristol strain and that the residual acidic activity is strongly inhibited by acarbose (Table 6). These results are in full accordance with the activity values acquired after RNAi assays of *aagr-1* in N2 Bristol strain (Tables 4 and 5) and thus support our previous finding that AAGR-1 is the least acarbose sensitive acidic AAGR and thus the most probable ortholog of human acid α-glucosidase.

C. elegans strain	α-glucosidase activity [nmol/$mg_{protein}$*h]						% of activity inhibited by acarbose
	pH 4		**pH 4**		**pH 6,5**		
	acarbose -		**acarbose +**		**acarbose -**		
	activity	% of control	activity	% of control	activity	% of control	
aagr-1 - ok2317	147	**75**	12	**33**	329	**83**	**92**
N2 Bristol	198	**100**	37	**100**	398	**100**	**81**

Table 6: GAA activities of the *ok2317* deletion mutant. Table shows GAA activities of *aagr-1* deletion mutant compared to control N2 Bristol strain. The impact of acarbose is demonstrated in the final column. All measurements were done at 37 °C.

Unfortunately, the *aagr-1* deletion mutant does not clearly replicate GSD type II (Pompe) phenotype. We explain this finding primarily by the existence of the second enzyme (*aagr-2*) with acidic glucosidase activity in *C. elegans* and which probably compensates the deficient acidic glucosidase activity of *aagr-1*. The possibility of compensation is further supported by the results from RNAi experiments of *aagr-1* and *aagr-2,* respectively. Biochemical analyses of *ok2317* mutant strain show that contribution of *aagr-2* to the total

acidic GAA activity is predominant over the contribution of *aagr-1* (only about 25% of total activity). We can also speculate whether the substrate turnover in worms and whether the duration of live cycle of *C. elegans* are sufficient to allow lysosomal storage buildup. Based on this assumption, it is possible that double knockout of both enzymes with acidic GAA activity in *C. elegans* AAGR-1 and AAGR-2 could be suitable model for study Pompe disease and may provide valuable information about the suitability of *C. elegans* deletion mutants of lysosomal hydrolases as appropriate models of human lysosomal storage disorders.

We analysed the *ok1423* strain on the level of the genomic DNA. The sequence analysis revealed 1832 bp long deletion that resulted in the loss of 421 C-terminal base pairs of the exon 4, complete intron 4 and exon 5 and 45 base pairs from intron 5 (Figure 15B). The *ok1423* deletion leads to the loss of 281 amino acids from the conceptual translation of the *aagr-4* gene, including the complete active site and the first 6 βα repetitions in the $(\beta\alpha)_8$ - core barrel according to the multiple amino acid sequence alignment.

We performed detailed microscopic analysis of the *ok1423* animals; however we did not see any abnormal morphological or other observable phenotype. Consequential measurement of GAA activities in the homogenates of deleted *vs.* standard N2 Bristol strain at pH values 4.0 and 6.5 showed the insignificant impact of the deletion on enzymatic activities at both pH levels (Table 7). To investigate the contribution of the two predicted neutral activity enzymes, AAGR-3 and AAGR-4, to the global neutral GAA activity, we performed RNAi assays against *aagr-3* mRNA in *ok1423* back-crossed strain. The acquired activity values measured at neutral pH corresponded with the values obtained by the RNAi experiments against *aagr-3* in N2 Bristol strain. Activity values measured at acidic pH did not differ from the controls (Table 7). Based on these results we conclude that AAGR-4 provides only a minor portion of the global neutral GAA activity in contrast to the dominant contribution of AAGR-3.

C. elegans strain	α-glucosidase activity [nmol/$mg_{protein}$*h]			
	pH 4		pH 6,5	
aagr-4* - *ok1423	134	**82**	264	**95**
N2 Bristol	165	**100**	277	**100**
RNAi of *aagr-3* in *ok1423* strain	260	**110**	99	**26**
control in *ok1423*	238	**100**	375	**100**

Table 7: GAA activities of the *ok1423* deletion mutant. Table shows GAA activities of *aagr-4* deletion mutant compared to control N2 Bristol strain and GAA activities after *aagr-3* RNAi experiment in *ok1423* strain which are comparable to activities values after RNAi of *aagr-3* in N2 worms. All measurements were done at 37 °C.

5.1.2.6 Transcriptional GFP fusions of *aagr-1* and *aagr-2* genes

In order to evaluate the expression patterns of the acid GAA orthologs AAGR-1 and AAGR-2 we prepared transcriptional GFP fusion constructs containing regulatory sequences of the genes making up *aagr-1* and *aagr-2* genes, and CEOP4284 operon.

5.1.2.6.1 Operon CEOP4284 and *aagr-1*

Genomic region including CEOP4284 operon which includes two genes sqv-1 [287] and D2096.12 and downstream gene of our prime interest - *aagr-1* (D2096.3), was divided into four fragments (Figure 16) in order to evaluate the extent of regulatory sequences of the individual genes.

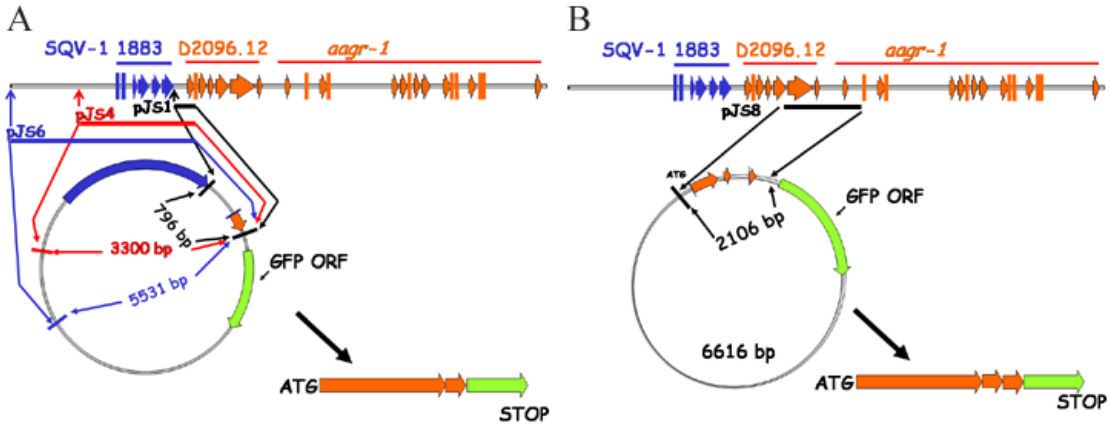


Figure 16: Schematic representation of extrachromosomal GFP expression constructs on operon CEOP4284 and *aagr-1* gene and relevant RNA transcripts. (A) Transcriptional GFP constructs pJS1 (796 bp of the operon nucleotide sequence; shown in black), PJS4 (3300 bp; in red) and pJS6 (5531 bp; in blue). (B) Transcriptional GFP construct pJS8 (2106 bp) including immediate *aagr-1* regulatory sequence.

Transcriptional extrachromosomal GFP constructs covering intergenic region between sqv-1 and D2096.12 and 5 kb of cosmid D2096 genomic region upstream of D2096.12 ATG including entire coding region of sqv-1 (pJS6) revealed broad expression GFP pattern in many types of nematodes´ tissues such as intestinal cells, pharyngeal muscle, coelomocytes, body wall muscle cells, pharyngeal and rectal glands, epidermal cells and rectal sphincter (Figure 17). GFP expression was observed also in a population of head neurons. The expression was detectable during whole development of the nematode from embryonal to adult stage (Figure 17). We prepared two GFP constructs either with nuclear localization sequence (NLS) (vector pPD95.69) or without it (vector pPD95.75). The NLS motif is used to enhance GFP expression. Because the presence of NLS has not changed the GFP expression

pattern compared with its absence in the fusion constructs, we resorted to work further only with constructs without NLS (vector pPD95.75).

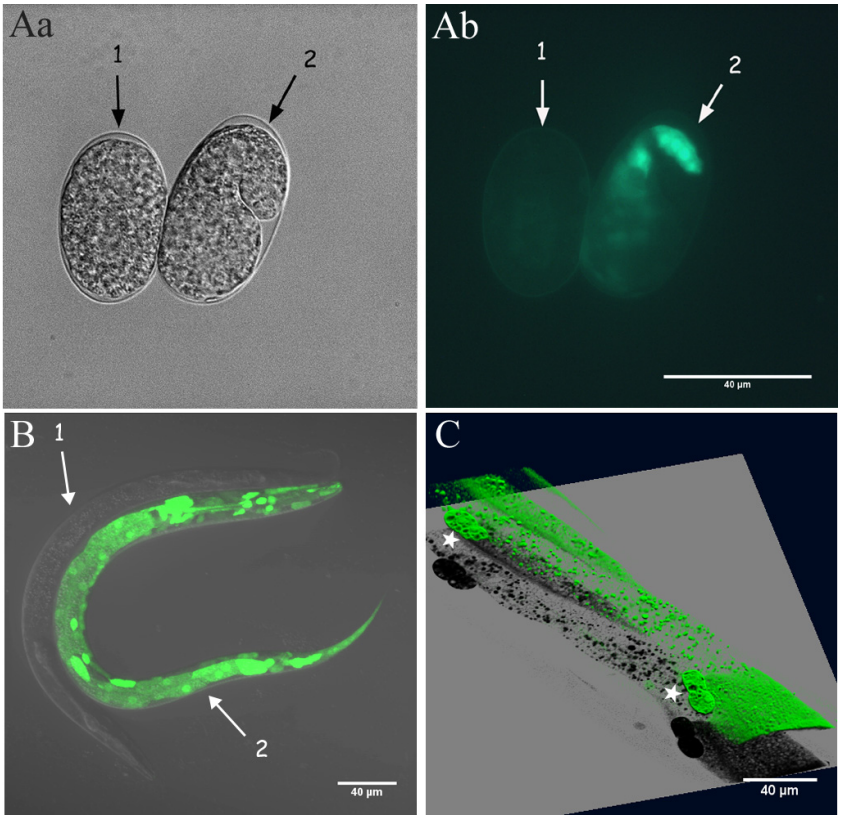


Figure 17: GFP expression of the transgenic worms carrying pJS6 construct. (A) The non-transgenic (1) and pJS6 construct carrying (2) embryos. (a) DIC and (b) fluorescence non-confocal images. (B) The transgenic worm (2) with broad GFP expression pattern in intestinal cells, pharyngeal muscle cells, coelomocytes and head epidermal cells. The non-transgenic worm (1) is shown in the background. (C) Detailed view of the broad GFP expression pattern in intestinal cells, body wall muscles and two pairs of coelomocytes (asterisks). The picture C was acquired by 3D rendering of initial confocal Z-stacks. Scale bars represent 40 µm.

Transcriptional extrachromosomal GFP construct covering 3kb region upstream of D2096.12 ATG (pJS4) showed only intestinal cells and pharyngeal muscle expression pattern (Figure 18) visible during all developmental stages of the nematode.

GFP construct covering 459 bp upstream of D2096.12 ATG including only intergenic region between sqv-1 and D2096.12 (pJS1) revealed GFP expression limited to anal sphincter (Figure 19) and the temporal expression was coincident to pJS4 and pJS6.

The last and for our research the most important GFP construct pJS8 includes 2106 bp of *aagr-1* regulatory sequence, which cover 1340 bp upstream of *aagr-1* ATG and complete intron 1. GFP expression was limited to six coelomocytes and in some cases to intestinal cells (Figure 20) and was observed in later developmental stages of the nematode (from L4 stage).

Earlier temporal expression was not observed even when extensively evaluated with confocal microscope. The expression pattern was similar to the one obtained by expression of *gana-1*::GFP.

GFP expression of constructs covering different lengths of regulatory sequence of CEOP4284 operon and *aagr-1* showed broad spatial and temporal pattern and was dependent on the regulatory sequence length.

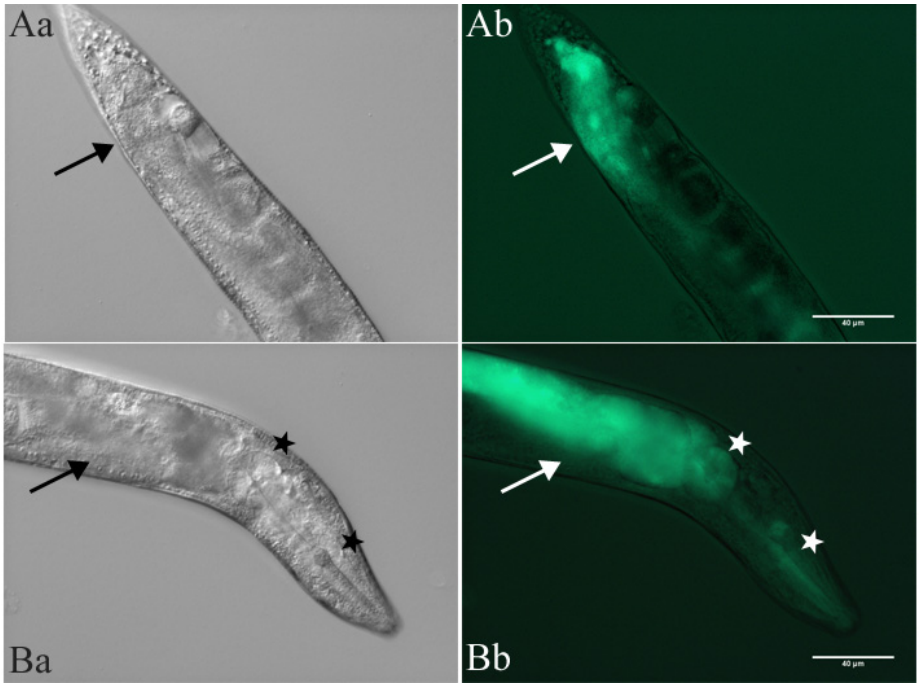


Figure 18: GFP expression of the transgenic worm carrying pJS4 construct. The transgenic worm with GFP expression (A) in intestinal cells (arrow) and (B) in intestinal cells (arrow) and pharyngeal muscles (asterisks). (a) DIC and (b) fluorescence non-confocal images. Scale bars represent 40 μm.

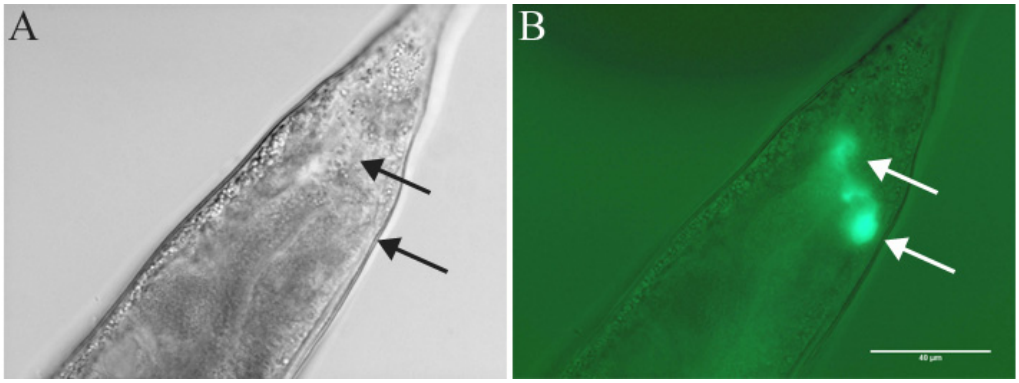


Figure 19: GFP expression of the transgenic worm carrying pJS1 construct limited to anal sphincter (arrows). (A) DIC and (B) fluorescence non-confocal images. Scale bar represents 40 μm.

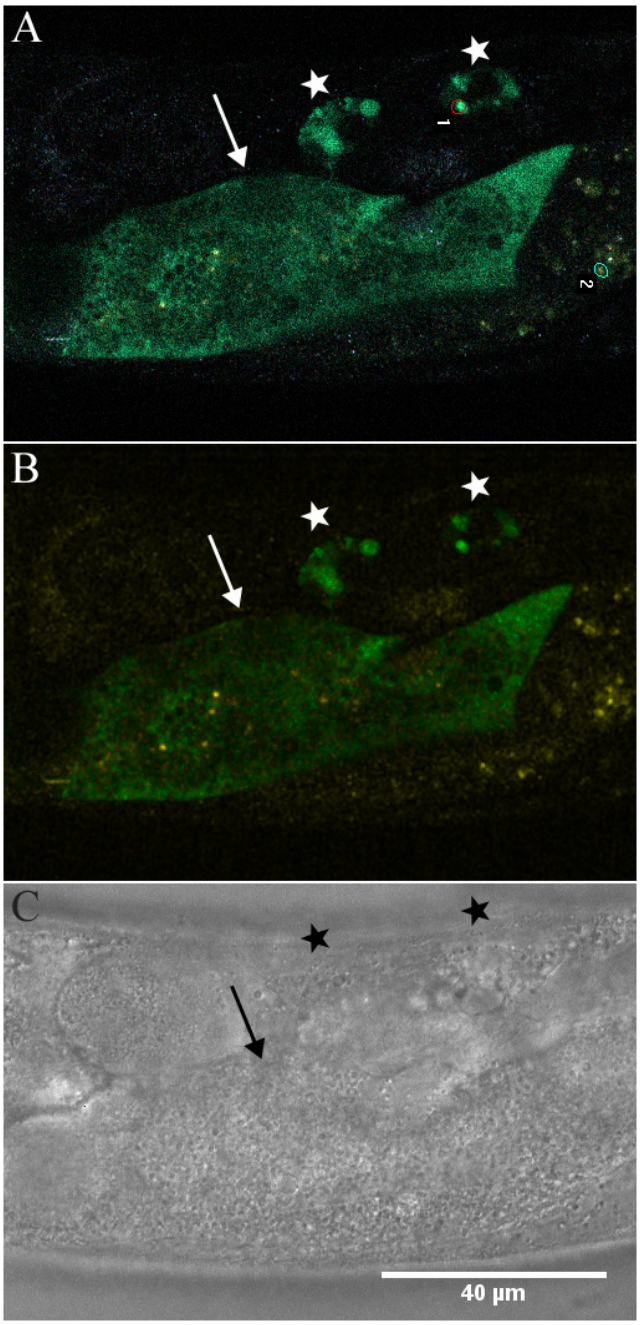


Figure 20: GFP expression of the transgenic worm carrying pJS8 construct limited to coelomocytes (asterisks) and intestinal cells (arrow). Differentiation of the specific GFP signal from the intestinal autofluorescence was based on the acquirement of spectral confocal image and its subsequent mathematical unmixing according to reference spectra (see section 4.1.6). (A) The original spectral image with defined regions of interests used for linear unmixing: (1) GFP signal (small red circle) and (2) intestinal autofluorescence (blue circle). (B) Unmixed image demonstrated green GFP and yellow intestinal autofluorescence signals. (C) Corresponding DIC image. Scale bar represents 40 µm.

5.1.2.6.2 *Aagr-2*

Transcriptional extrachromosomal GFP construct of *aagr-2* (pJS9) covered 3431bp upstream of *aagr-2* ATG (Figure 21). Positive signal was spatially limited to membrane-bounded vacuoles of coelomocytes and the diffuse GFP signal was observed also in the pseudocoelom. Signal in the pseudocoelom was highly suggestive of coelomocytic secretory origin (Figure 22). We found no other GFP positive tissue. Temporally the expression was observed in L4 and adult nematode stages, similarly to GFP expression of *aagr-1*.

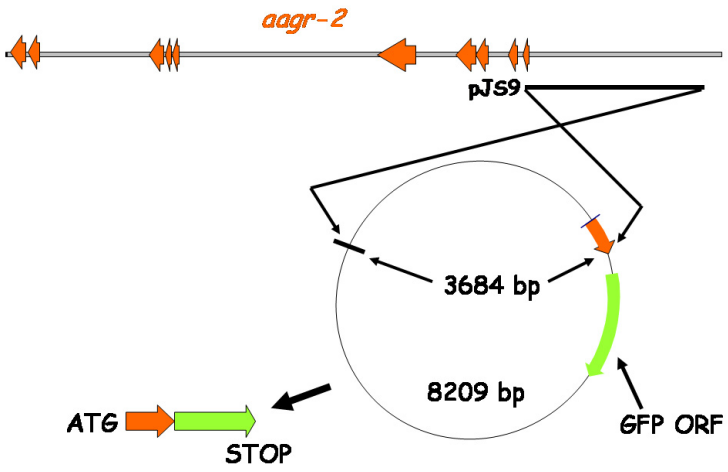


Figure 21: Schematic representation of *aagr-2* extrachromosomal GFP construct (pJS9) and its RNA transcript.

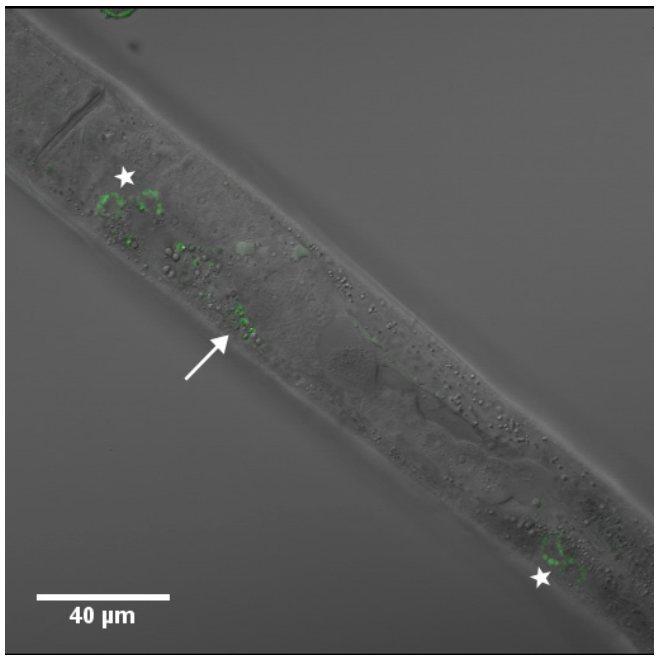


Figure 22: GFP expression of the transgenic worm carrying pJS9 construct limited to vesicular compartments of coelomocytes (asterisks). Arrow shows the diffuse signal in pseudocoelom. Scale bar represents 40 µm.

The coelomocytes expression pattern of *aagr-2, aagr-1* and *gana-1* was very similar and together suggest high content of glycohydrolytic enzymes in these cells.

5.1.2.7 Bioinformatic studies on *aagr-1-4*

This part of the thesis was performed in collaboration with Jakub Sikora[1], Filip Majer[1] and Karel Jelínek[2] ([1]Charles University in Prague, 1st Faculty of Medicine, Institute of Inherited Metabolic Disorders; [2]Charles University in Prague, Faculty of Science, Departement of Physical and Macromolecular Chemistry).

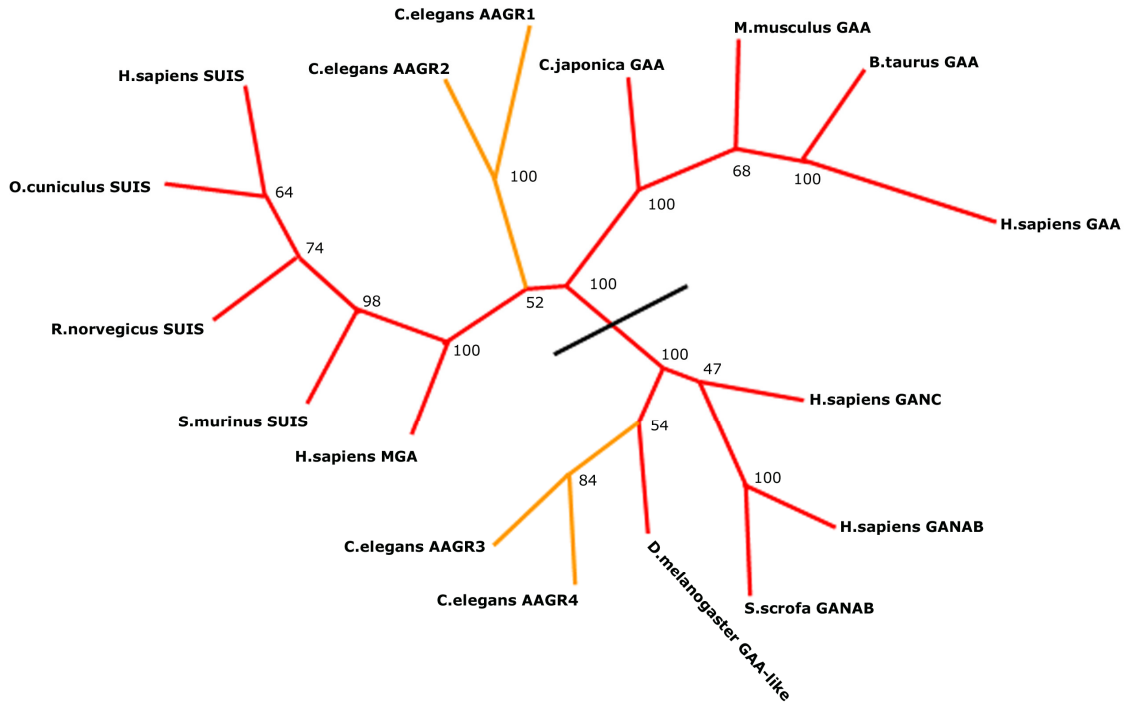


Figure 23: Phylogenetic tree of AAGR-1-4 and other selected GH31 proteins constructed using maximum likelihood method. Bootstrap values at the nodes evaluate the tree (maximum 100). Black line divides enzymes with the acidic pH optimum (including AAGR-1 and -2) from the ones with neutral pH optimum (including AAGR-3 and -4) (for details see appended submitted manuscript).

Multiple alignment of amino acid sequences of selected eukaryotic proteins from GH31 family and AAGRs demonstrated conservation of primary protein sequence within the extent of GH31 module [288]. The multiple alignment and unrooted phylogenetic tree (Figure 23) divided protein sequences into two discrete clades. Proteins forming the first clade were confirmed or predicted to have acid pH optima of enzymatic activity (GAA, SUIS, MGA) and included AAGR-1 and AAGR-2 and proteins from the second clade have neutral pH optima of enzymatic activity (GANC, GANAB) and included AAGR-3 and AAGR-4. The

immediate surroundings of the active site nucleophile (WiDMnE) was the most conserved region in the aligment. All bioinformatic studies (multiple alignment, phylogenetic analysis and homology modeling) were based on verified sequences of *aagr*s.

Homology modeling was based on multiple alignment of *E. coli* YicI [219], *S. solfataricus* [224], *H. Sapiens* NtMGA [225] and all four AAGRs (Figure 24). The best template for all AAGRs was 3D structure of NtMGA [225] with inserted inhibitor acarbose.

The most conserved part of all AAGRs proteins was the catalytic domain. The sequence identity of the catalytic domain between the template NtMGA and AAGR-1 and -2 was higher when compared to AAGR-3 and -4. AAGR-1 and -2 possess both structural inserts (proximal 1 and distal 2) in the catalytic domain which are present in NtMGA and MalA proteins (Figure 24 and 25). In addition, both inserts in AAGR-1 and -2 contain unique additional insertions (for details see appended submitted manuscript). On the contrary, AAGR-3 and -4 do not have distal structural insert 2 and unique additional insertions of AAGR-1 and -2. Moreover, the entire N-terminal domain of AAGR-3 and -4 shares low level of identity with all templates because of their higher similarity with GANAB and GANC proteins (for details see appended submitted manuscript).

We observed only few substitutions in the active site of AAGR-1-4. The most important difference between templates and model molecules was the substitution tyrosine/tryptophan (Y/W) in the active site which may discriminate between proteins which are inhibited/uninhibited by acarbose. The residues Y299 of NtMGA aligns to Y184 in MalA and structurally corresponds to Y304 in AAGR-2 (proteins inhibited by acarbose) but is substituted by W316, W361 and W353 in AAGR-1, -3 and -4 (proteins insensitive to acarbose) (Figure 24 and 25). These results further confirm AAGR-1 as the true ortholog of human acid α-glucosidase. For details about homology modeling and docking see appended submitted manuscript.

To conclude, the use of rationally designed [289] or other described GH31 substrates/inhibitors [290] and methods like site directed mutagenesis could further elucidate the enzyme specificity and physiological functions of the nematode´s AAGRs. Thereafter, it is plausible to speculate about the potentional misinterpretations of GAA activity assay with acarbose inhibition in GSD type II diagnostics caused by pathogenic substitutions in the presumed GAA active site residues (for details see appended submitted manuscript).

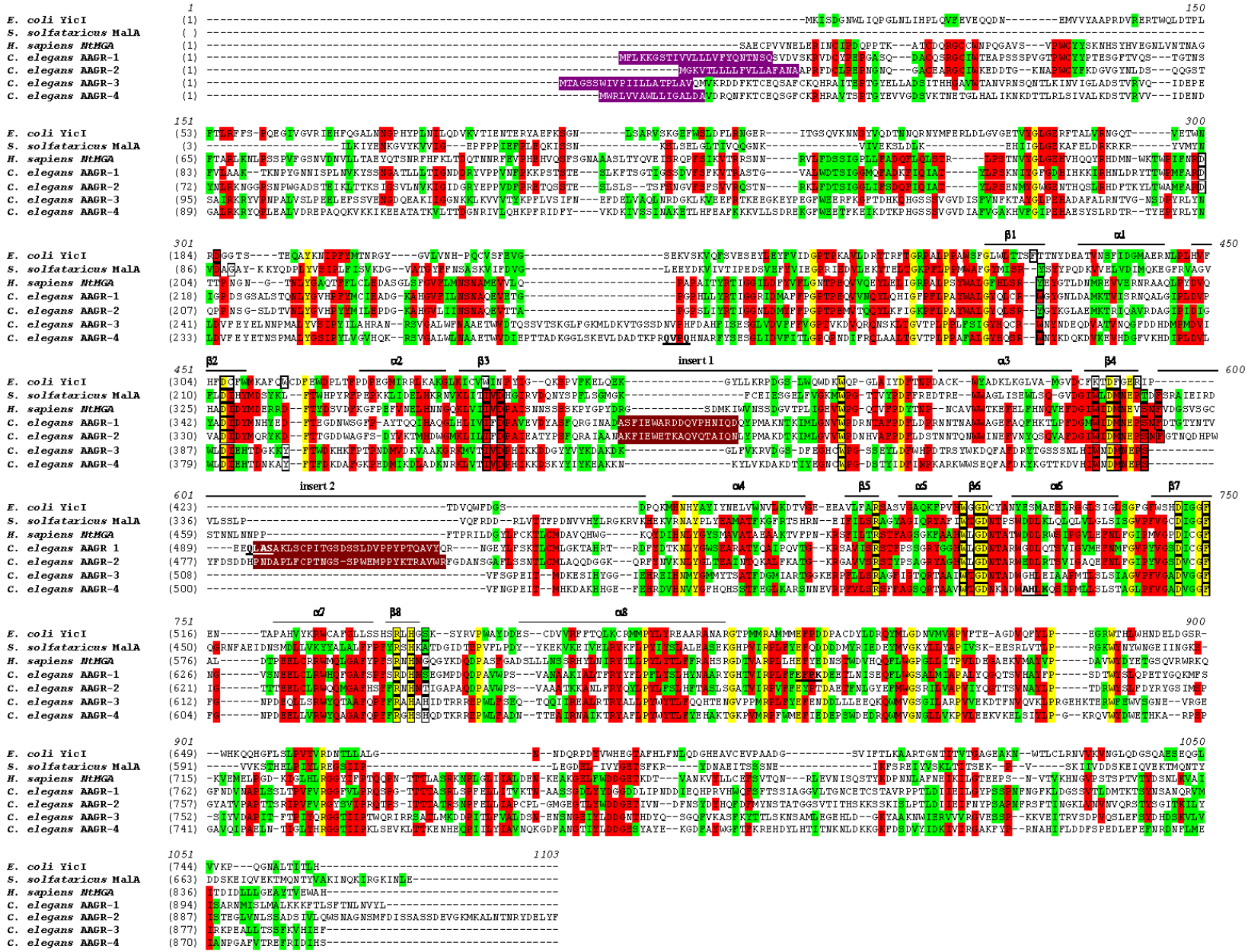

E. coli YicI
S. solfataricus MalA
H. sapiens NtMGA
C. elegans AAGR-1
C. elegans AAGR-2
C. elegans AAGR-3
C. elegans AAGR-4
β1
α1
β2
α2
β3
insert 1
α3
β4
insert 2
α4
β5
α5
β6
α6
β7
α7
β8
α8

Figure 24 (previous page)**: Multiple alignment of *E. coli* YicI [219], *S. solfataricus* [224], *H. Sapiens* NtMGA [225] and all four *C. elegans* AAGRs.** Identical residues are in yellow, conserved residues are in red and blocks of similar amino acids are in green color. The extent of structural elements $(\beta\alpha)_8$ and structural inserts in catalytic domain is shown above the alignment. Side chains of amino acids residues in the boxes correspond to those whose side chains lie with the diameter approximately 5 A° of the docked acarvosine. Sequences in brown (AAGR-1 and -2) correspond to insertions within the structural inserts. Residues in violet correspond to predicted signal peptids. Underlined sequences in AAGR-1 and -4 mark the extent of deletions in *ok2317* and *ok1423*, respectively (compare with small insets in Figure 25A).

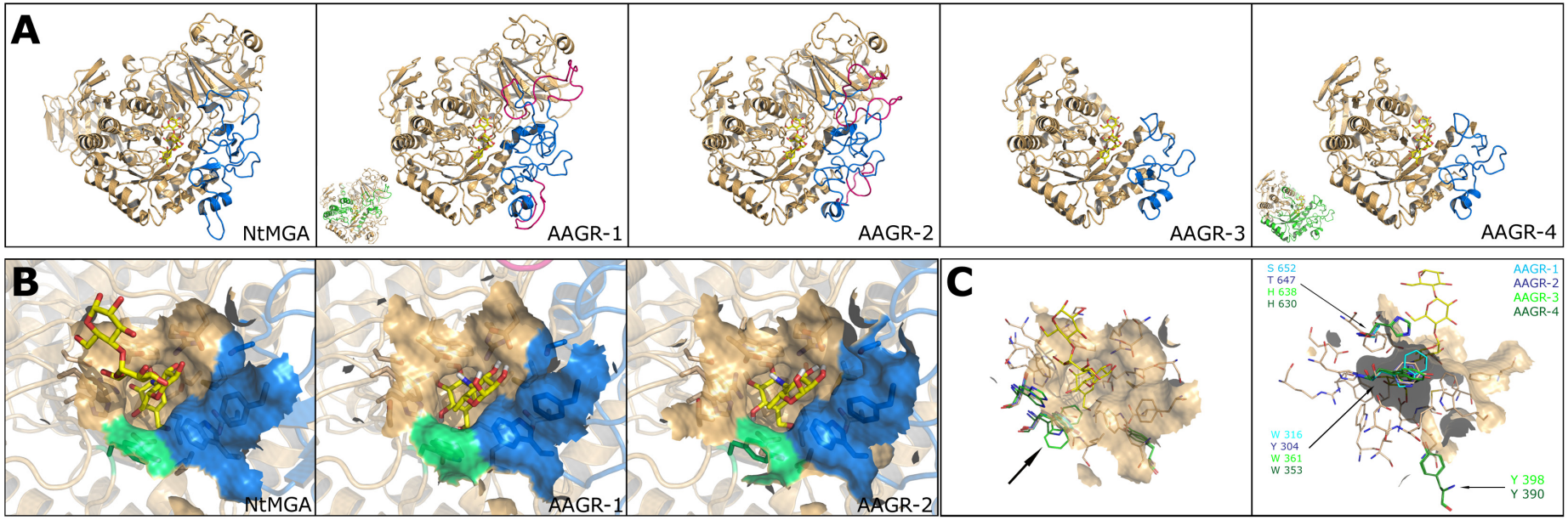


Figure 25: 3D models of AAGR-1-4 proteins. (**A**) 3D structure of the best template NtMGA [225] and structural models of all AAGRs in ribbon representation with secondary structure elements. Structural inserts 1 and 2 are in blue and insertions in AAGR-1 and -2 are in purple red. Acarbose (yellow) was inserted by superimposition from NtMGA structure. N-terminal domains of AAGR-3 and -4 are not shown because of their low identity with templates. The extent of deletions (RB1790 and RB1307 strains) in AAGR-1 and -4 models are shown in green on the small inset images. (**B**) Binding active site pocket of NtMGA, AAGR-1 and AAGR-2 represented as solvent accessible surface. Contributions of both inserts are colored in blue, the substitution Y/W is in green and docked acarvosine is in yellow. (**C**) Differences in active site of NtMGA and AAGR models. Side chains of amino acids in active site conserved in all models and NtMGA are shown in light orange for template only. Non-equivalent residues are coloured according to the annotation in the picture. An arrow indicates a viewing direction for the side view image.

5.1.3. Characterisation of human Acetyl-CoA: α-glucosaminide N-acetyl-transferase on the cellular level (appended publication 2 and supplementary material to this publication)

Based on the results of linkage analyses that narrowed the candidate region for human heparin acetyl-coenzyme A:α-glucosaminide N-acetyltransferase (HGSNAT, N-acetyltransferase) to a 2.6-cM interval between *D8S1051* and *D8S1831*, we identified the TMEM76 gene. TMEM76 was located within the candidate region as the gene that encodes lysosomal HGSNAT and whose deficit causes mucopolysaccharidosis IIIC (MPSIIIC, or Sanfilippo C syndrome) (for details see appended publication 2). Concurrently with publication of our results another group reported the same gene found by proteomic analyses [236].

5.1.3.1 Sequence verification of predicted gene

We verified the sequence of the predicted gene by sequencing of the full-length cDNA (4.5 kb) which is composed of two polyadenylation signals and 1992 bp long coding sequence containing 18 exons. The cDNAs acquired from various tissues were amplified in overlapping PCR fragments. We found that HGSNAT is ubiquitously expressed in all evaluated human tissues with the highest expression in leukocytes, placenta, heart, lung and liver. These results were consistent with results acquired by Northern-blotting (see appended publication 2). In addition to the full-lenght transcript we amplified two shorter ones by RT-PCR. We identified one shorter alternative transcript with spliced out exons 9 and 10 resulting in an in-frame deletion of 64 amino acids. We suppose that this transcript does not encode an active enzyme because it was detected in two patients with MPSIIIC who had almost complete loss of HGSNAT activity. The second shorter transcript lacked exons 3, 9 and 10 and led to an in-frame deletion of 107 amino acids.

5.1.3.2 Basic bioinformatic analysis

An integrated bioinformatic search reveal that HGSNAT encodes a protein composed of 663 amino acids (73 kDa) with N-terminal signal peptide, 11 predicted transmembrane domains and 4 potentional N-glycosylation sites (Figure 26). The signal peptid prediction done by SignalP server [254] predicted presence of N-terminal signal sequence with high probability (0.996). The most probable signal sequence cleavage site was predicted between

residues 47-48 or 58-59. Cleavage position is consistent with length of predicted signal peptid. Prediction of transmembrane domains in HGSNAT sequence truncated of predicted signal peptide sequence was performed by TMHMM server [254] and proposed 11 transmembrane domains with high probability (0.900-1.000) (see Figure 26). NetNGlyc server [254] was used for prediction of N-glycosylation sites in HGSNAT. Prediction results revealed four potentional N-glycosylated sites in positions N94, N142, N162 and N461 from the first methionin. For schematic representation see Figure 26.

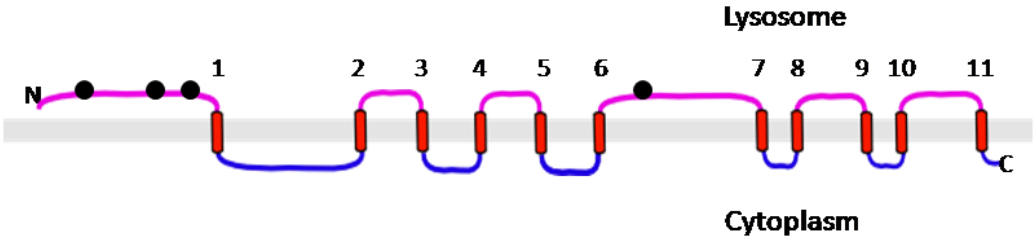


Figure 26: Schematic representation of HGSNAT protein in lysosomal membrane. HGSNAT contains 11 predicted transmembrane domains (red cylinders) and four potentional N-glycosylation sites (black circles).

Surprisingly, although it is present in a variety of mammalian genomes, the sequence homology searches demonstrated that HGSNAT does not have a structural similarity neither to any known prokaryotic or eukaryotic N-acetyltransferase nor to other lysosomal protein. HGSNAT shares homology with conserved uncharacterized family of bacterial membrane proteins COG 4299 (for details see appended publication 2). Based on these findings we were not surprised that *C. elegans* lacks HGSNAT ortholog in Wormpep database and also that the measurements of HGSNAT activity in *C. elegans* N2 homogenates revealed insignificant values. Based on these results we decided not to study HGSNAT in *C. elegans*.

5.1.3.3 Cellular distribution of HGSNAT (supplementary material to the appended publication 2)

We have obtained an affinity purified rabbit polyclonal antibody against a peptide epitope from one of the HGSNAT extramembrane domains which we have further used for characterisation of cellular distribution of HGSNAT in normal human fibroblasts and fibroblasts acquired from patients with selected lysosomal storage disorders. The specifity of the antibody was verified by Western blotting and by measuring of dependence of HGSNAT activity on concentration of the antibody (Figure 27). The activity of HGSNAT was inhibited

by increasing concentration of antibody in contrast to activity of control enzyme β-hexosaminidase that did not change during experiments.

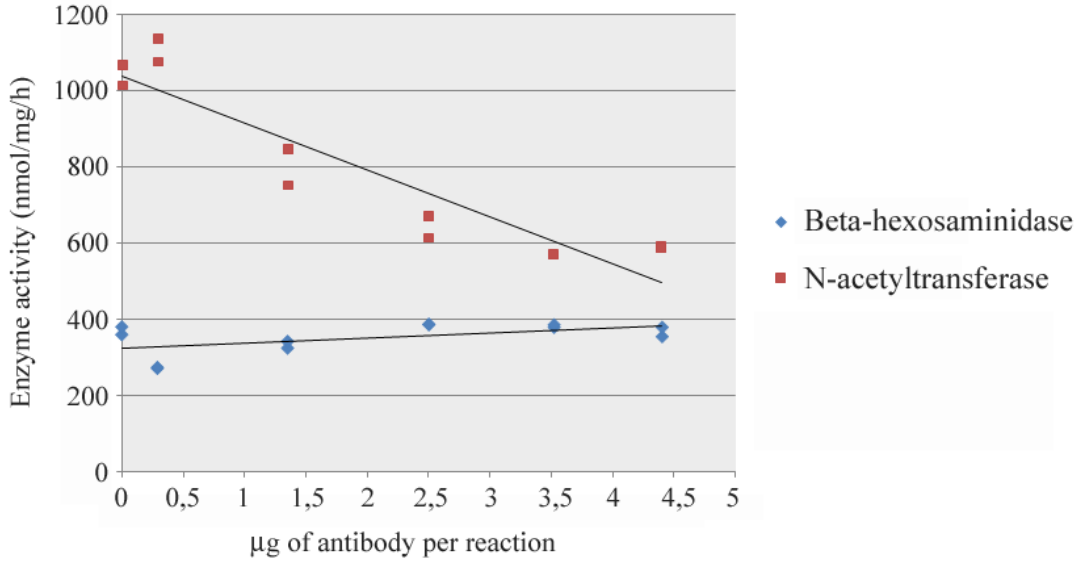


Figure 27: Graph of dependence of HGSNAT and β-hexosaminidase activities on concentration of antibody against HGSNAT. Activity of HGSNAT is inhibited by increasing concentration of antibody. Activity of control enzyme β-hexosaminidase remains constant. The curves represent linear regression.

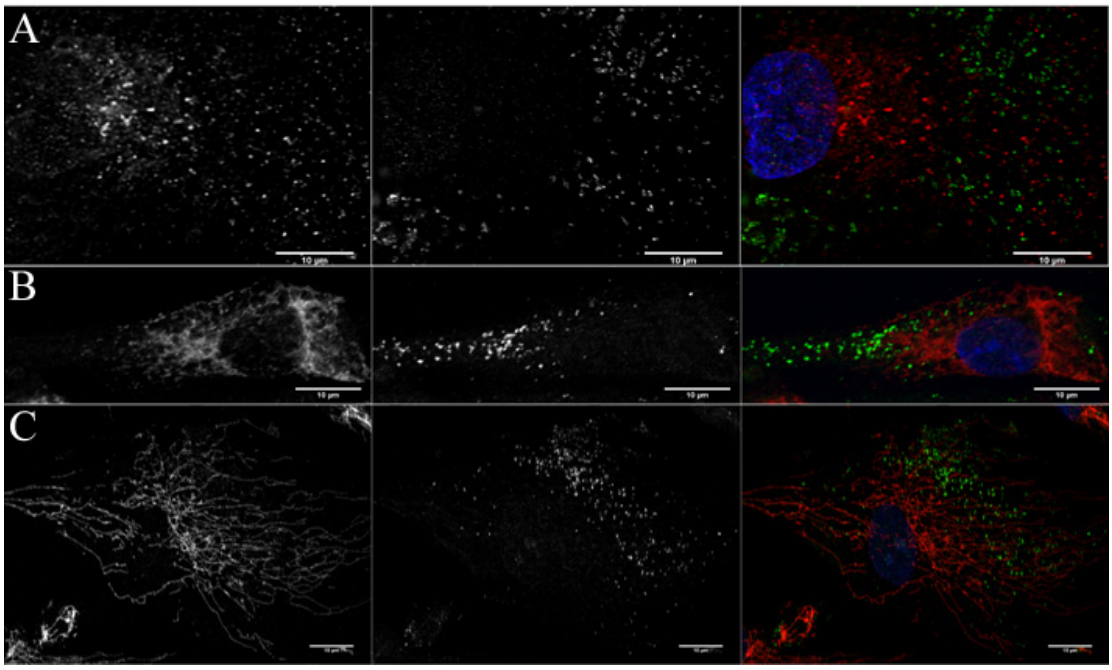


Figure 28: Subcellular organellar distribution of HGSNAT. Double immunolabeling of HGSNAT with (A) external Golgi apparatus 58K protein, (B) endoplasmic reticulum resident protein – protein disulfide isomerase (PDI) and (C) inner mitochondrial membrane marker – subunit I of cytochrome oxidase. HGSNAT signal is shown in green, the other detected epitopes are in red. Nuclei were counter stained with DAPI (in blue). Scale bars represent 10 µm.

Immunofluorescence microscopy showed that in control cultured human fibroblasts the anti-HGSNAT antibody stained a subpopulation (approximately 25- 50% of LAMP2-positive vesicles) of lysosomes. The signal did not systematically co-localize with any of the other organellar markers: PDI (endoplasmic reticulum), Golgi 58K (Golgi apparatus), COXI (mitochondria) (Figure 28).

The systematic co-localization studies with EEA1 (early endosomal compartment), mannoso-6-phosphate receptor (M6PR, late endosomal compartment), cathepsin D (lysosomal luminal protein) and LAMP 2 (lysosomal associated membrane protein) (Figures 29 and 30) showed that the anti-HGSNAT antibody is not detected neither in early nor late endosomes but stained only discrete regions of the lysosomal membrane probably on its luminal side.

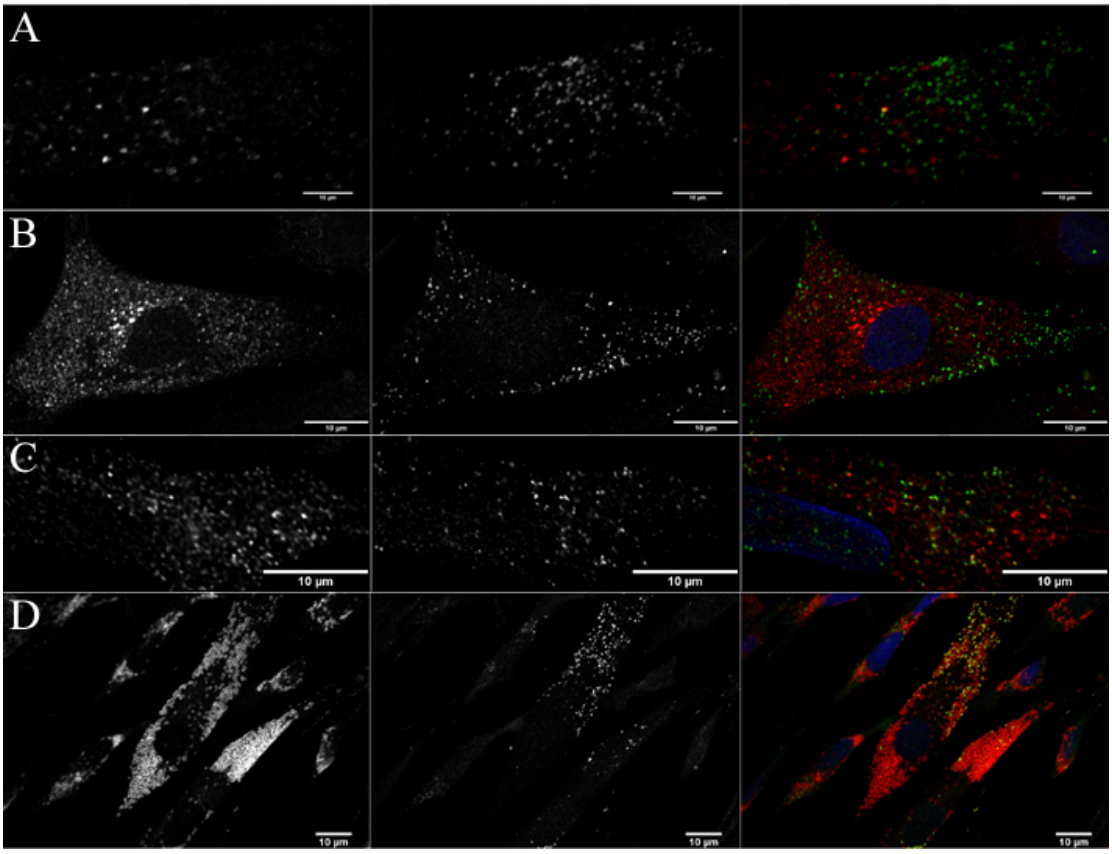


Figure 29: Distribution of HGSNAT throughout the endosomal-lysosomal system. Double immunolabeling of HGSNAT with (A) early endosomal antigen 1 (EEA1) (early endosomes), (B) mannose-6-phosphate receptor (M6PR) (late endosomes), (C) cathepsin D (lysosomal luminal protein) and (D) lysosomal associated membrane protein 2 (LAMP 2) (lysosomal membrane). HGSNAT signal is shown in green, the other detected epitopes are in red. Nuclei were counter stained with DAPI (in blue). Scale bars represent 10 µm.

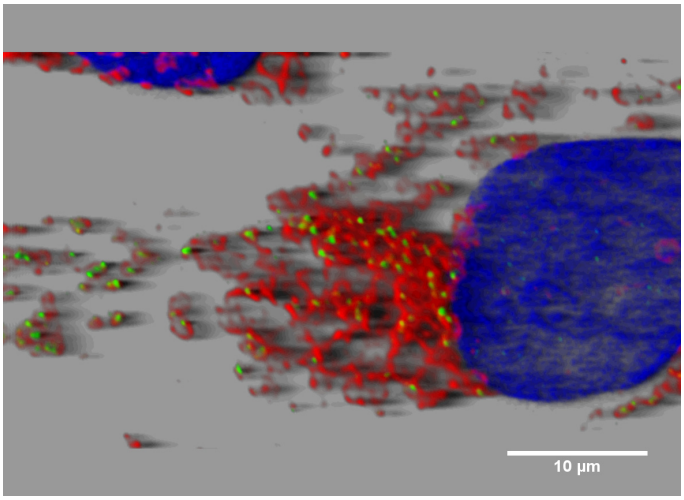


Figure 30: Detail view of HGSNAT distribution in the lysosomal membrane. HGSNAT signal is shown in green, LAMP 2 signal in red and nuclei (detected by DAPI) in blue. Image was acquired by 3D rendering of initial confocal Z-stacks in Imaris Personal software. Scale bar represents 10 µm.

Even thought the images suggest intraluminal orientation of the peptide epitope in HGSNAT protein, this observation should be perceived carefully, because the transmembrane orientation may depend on the protein´s functional status. The immunoelectron detection of HGSNAT (performed in collaboration with RNDr. Dusan Cmarko, Ph.D. (Charles University in Prague, 1st Faculty of Medicine, Institute of Cell Biology and Pathology)) in enriched lysosomal franctions further confirmed the localization in discrete regions of lysosomal membranes (Figure 31). The HGSNAT signals were accumulated on several, luminal located sites of single vesicular structures and suggest the possibility that lysosomal membrane could contain multiple membrane microdomains. These results corresponded well with the microdomain localization of HGSNAT previously proposed by biochemical methods [242].

In spite of apparent microdomain localization, HGSNAT did not co-localize with tested protein and lipid microdomain markers such as flotillin 1, GM1 ganglioside or globotriaosylceramide (Gb3) (Figure 32). Filipin staining (Figure 32C) for cholesterol was also not prominent in HGSNAT-positive regions, which makes HGSNAT presence in cholesterol microdomains unlikely. HGSNAT, however, partially co-localized with sortilin (Figure 33), a receptor in the alternative pathway for transport of soluble or membrane-associated lysosomal proteins into lysosomes [134, 136, 138]. Based on the result we hypothesize that lysosomal sorting of the HGSNAT can be sortilin dependent.

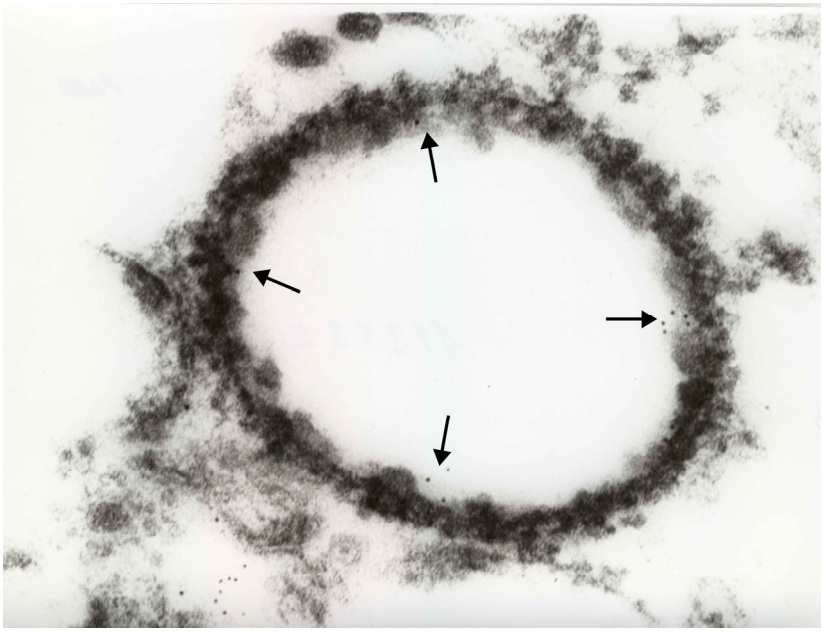


Figure 31: Immunoelectron detection of HGSNAT (arrows) in enriched lysosomal fraction. Indirect immunogold (6 nm) technique was used for detection. The magnification was 60 000 x.

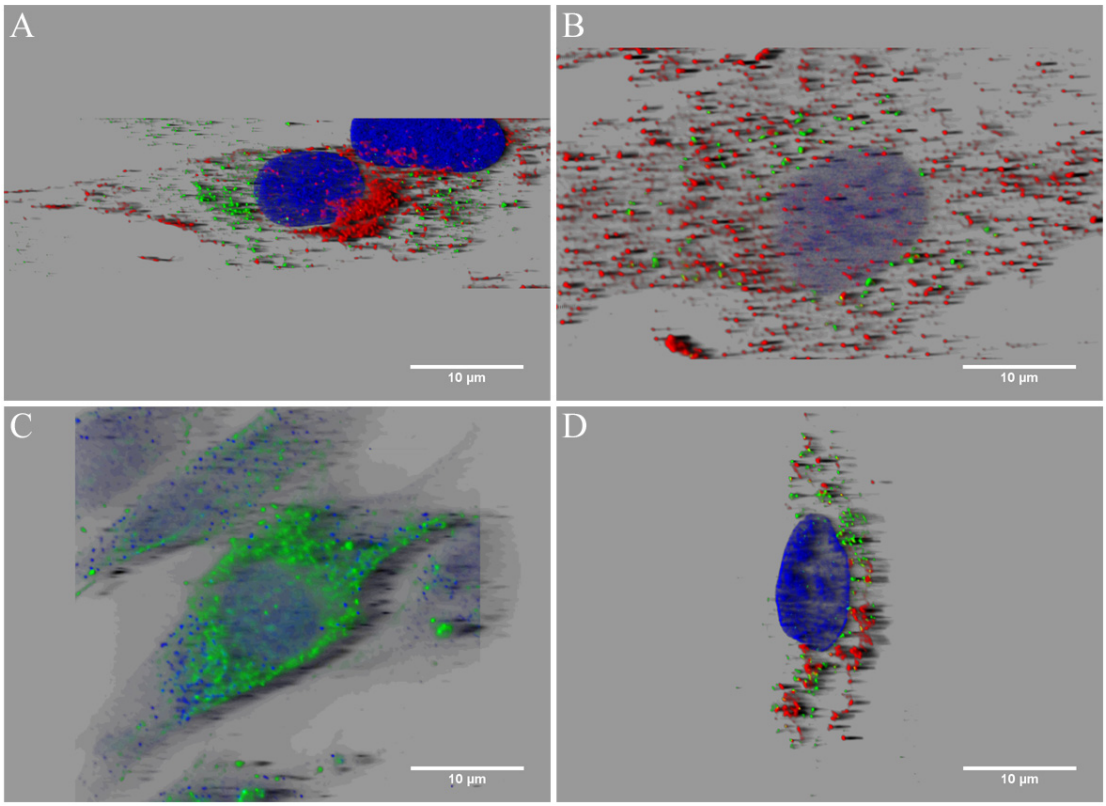


Figure 32: Distribution of HGSNAT in relation to some microdomain markers and selected membraneous lipidic moieties. Double labeling of HGSNAT and (A) GM1 ganglioside detected by choleratoxin B subunit binding, (B) globotriaosyl ceramide (Gb3 antibody), (C) cholesterol (filipin binding) (non-confocal image) and (D) membrane microdomain marker Flotilin. HGSNAT signal is always shown in green, filipin signal in blue and other detected macromolecules in red. Nuclei in images (A), (B) and (D) were counter stained with DAPI and are shown in blue. Images were acquired by 3D rendering of initial Z-stacks in Imaris Personal software. Scale bars represent 10 µm.

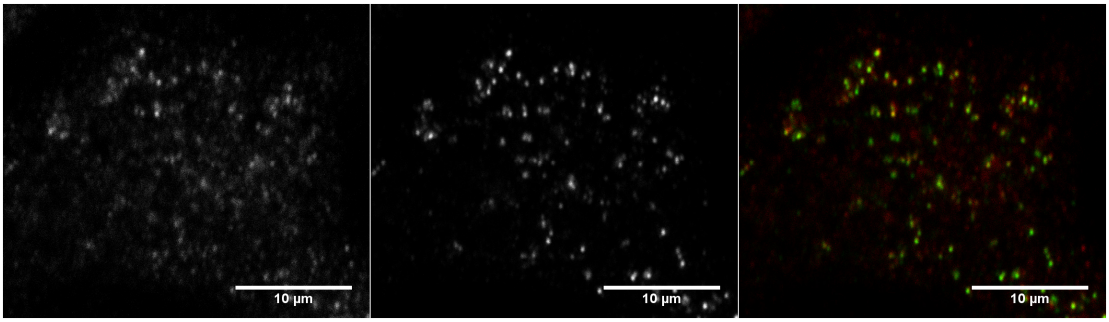


Figure 33: Double immunolabeling of HGSNAT (in green) and sortilin (in red). Scale bar represents 10 µm.

We studied the cellular distribution of HGSNAT in selected lysosomal storage disorders (Figure 34). The transport of HGSNAT apparently does not depend on MPR as it reached lysosomes in I-cell disease cells (N-Acetylglucosamine-1-phosphotransferase deficiency) [132]. This finding was not surprising because MPRs are receptors for delivering of soluble lysosomal proteins to LE. The subcellular distribution of HGSNAT also does not seem to be affected by deficiencies in other proteoglycan hydrolytic lysosomal pathways as demonstrated in selected mucopolysaccharidoses (I, II, IIIA and IIIC) cell lines. The expansion of the lysosomal system is evident especially in I-cell disease affected cells.

We were further interested in distribution of HGSNAT compared with its substrate – heparan sulphate (Figure 35). Surprisingly we were not able to demonstrate systematic co-localization of HGSNAT and heparan sulphate (HS) in any kind of evaluated cell lines. The antibody (Seikagaku Corporation) was prepared against the liposome-incorporated membrane heparan sulphate proteoglycan from human fetal lung fibroblasts and probably recognizes predominantly the whole heparan sulphate´s molecules. However, the antibody not necessarily recognises partially degraded forms of HS, which is present in lysosomes and also stored in MPS affected cell lines. So we explain this finding by the detection limits of the antibody and indirect immunofluorescence techniques. In addition, we have not yet quantified the overall content of heparan sulphate in any of the images.

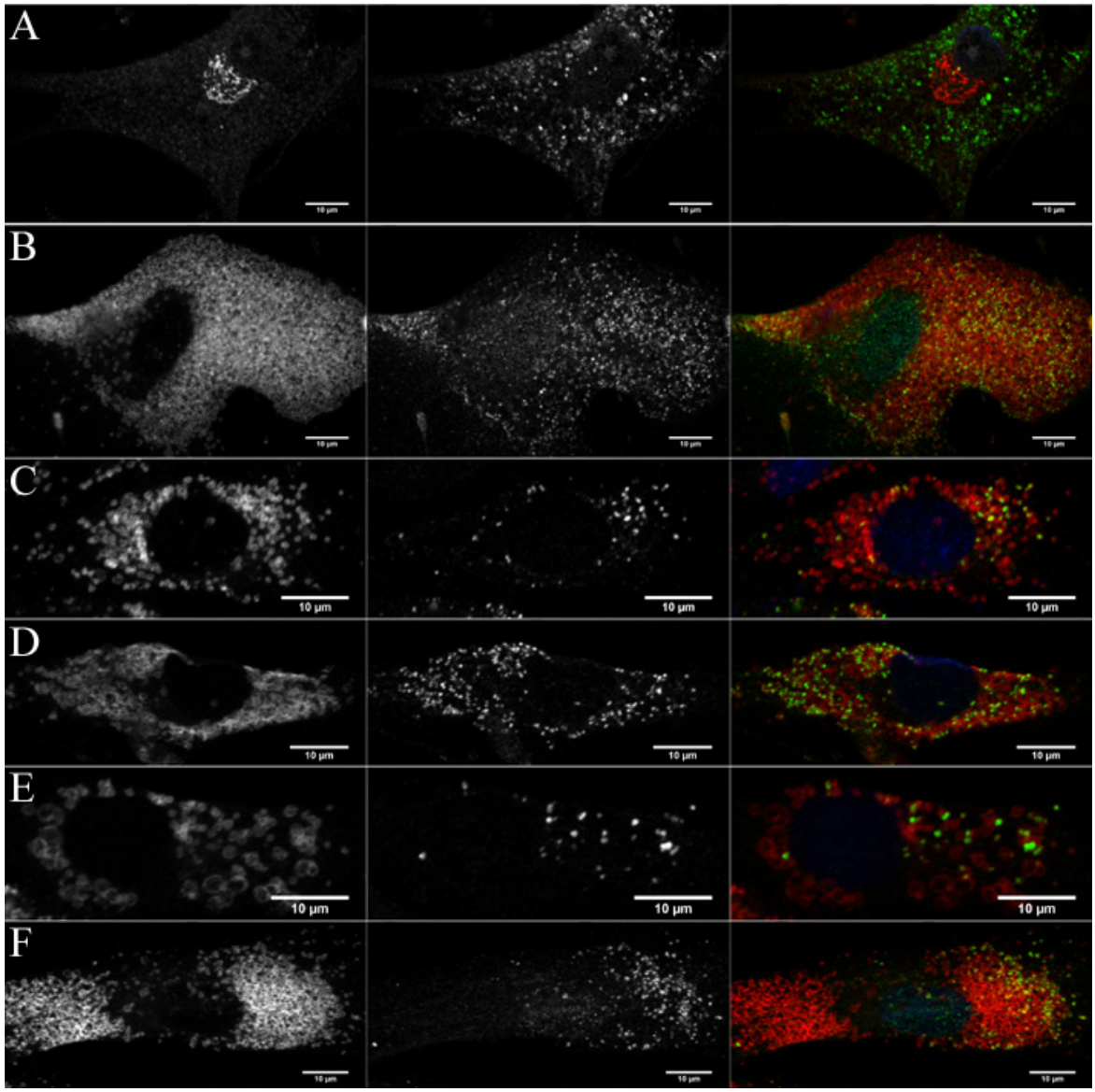


Figure 34: Cellular distribution of HGSNAT in selected lysosomal storage disorders. (A) I-cell disease (deficiency of N-acetylglucosamine-1-phosphotransferase) - double immunolabeling of HGSNAT with M6PR. Parallel immunolabeling of HGSNAT and LAMP 2 in (B) I-cell disease, (C) mucopolysaccharidosis type I (MPS I), (D) mucopolysaccharidosis type II (MPS II), (E) mucopolysaccharidosis type IIIA (MPS IIIA) and (F) mucopolysaccharidosis type IIIC (MPSIIIC). HGSNAT signal is shown in green, the other detected epitopes are in red. Nuclei were counter stained with DAPI (in blue). Scale bars represent 10 µm.

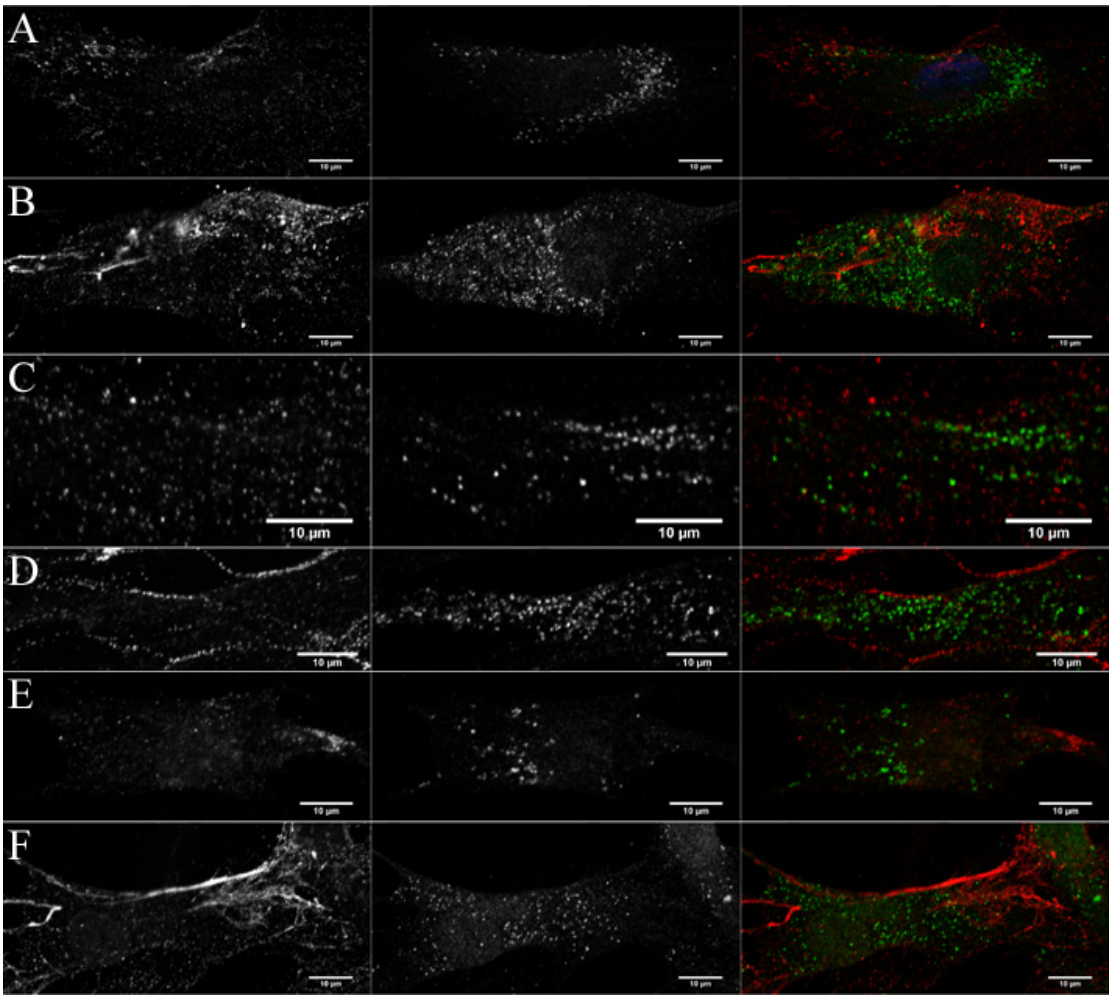


Figure 35: Cellular distribution of HGSNAT in relation to its substrate heparan sulphate in selected lysosomal storage disorders. (A) control fibroblasts (for comparison), (B) I-cell disease, (C) MPS I, (D) MPS II, (E) MPS IIIA and (F) MPSIIIC deficient fibroblasts. HGSNAT signal is shown in green, heparan sulphate is in red. Scale bars represent 10 µm.

The supplementary material to the appended publication 2 provides evidence of the presence of microdomains in lysosomal membrane on microscopy level and confirms the microdomain localization of HGSNAT in lysosomal membrane. In addition, we found out that the microdomains containing HGSNAT are not formed by any tested protein and lipid microdomain components, such as flotillin 1, GM1 ganglioside, globotriaosylceramide and cholesterol. These results will be an important part of the next publication about HGSNAT (*manuscript in preparation*).

6. Conclusions

6.1 General conclusions

The principal theme of this thesis was the study of selected deficits of lysosomal glycans metabolism - Fabry, Schindler, Pompe and Mucopolysaccharidosis IIIC diseases. The majority of this work was focused on simple metazoan model organism *Caenorhabditis elegans* and its use in modeling of lysosomal enzymopathies. The most important step was the identification of true nematode´s orthologs of selected human hydrolases and their characterisation. The main approach in modeling of diseases was RNA-mediated interference of relevant orthologous genes. We found that RNAi did not provide sufficient expression knock-down of hydrolases, which is necessary for development of lysosomal storage. Based on these results we propose that deletion mutants of relevant lysosomal hydrolases with zero activity may serve as valuable models for studying human LSDs. Even though we did not manage to generate strains with lysosomal storage, study of *C. elegans* orthologs of lysosomal glycosidases provided important information about these proteins and their phylogeny which could be applied in further studies of these diseases. In addition, as was shown in many other studies [81], *C. elegans* is a valuable model organism for studying human pathology states even in cases when the nematode´s mutant phenotype does not fully replicate human disease pathology.

The last part of this thesis dealt with the characterisation of lysosomal membrane protein HGSNAT mutated in mucopolysaccharidosis IIIC, which catalyzes a key step in degradation of glycosaminoglycan heparan sulphate. We found that HGSNAT does not have any *C. elegans* ortholog. The study of the cellular distribution of HGSNAT showed that HGSNAT is situated in discrete regions of lysosomal membranes and is present only in a subpopulation of lysosomal cellular pool. Our study gave evidence of presence of lysosomal membrane microdomains.

6.2 Individual conclusions

We found only one *C. elegans* ortholog (GANA-1) of both human α-GAL A and α-NAGA proteins. Biochemical, RNAi, GFP expression and bioinformatic analyses confirmed our assumption that GANA-1 has both α-GAL and α-NAGA activities and is localized in acidic cellular compartment. In addition, phylogenetic and homology modeling analyses confirm the hypothesis that α-Gal A and α-Naga genes arose by duplication from a

hypothetical ancestral gene and that *gana-1* probably developed from this hypothetical ancestral gene before the duplication event. We propose that a deletion mutant of *gana-1* may provide a suitable model for studying Fabry and Schindler diseases in *C. elegans*.

We found four potential *C. elegans* orthologs (*aagr-1-4*) of human acid α-glucosidase. We determined that AAGR-1 and -2 have predominant acidic and AAGR-3 and -4 neutral GAA activity based on results of biochemical, RNAi, GFP expression, bioinformatic and mutant analyses. We demonstrated that AAGR-2 is the prevalent contributor to the total acidic and AAGR-3 to the total neutral GAA activity and that expression of the acidic proteins AAGR-1 and -2 is localized to a subset of endocytically active cells. Our results shown that AAGR-1 is the least acarbose sensitive acidic AAGR which was further supported by the analysis of *aagr-1* deletion mutant. Based on these outcomes we consider AAGR-1 likely true ortholog of human acid α-glucosidase. However, deletion mutant of *aagr-1*, which we had at our disposal, does not clearly replicate GSD type II (Pompe) lysosomal storage phenotype. That is apparently due to existence of the second enzyme AAGR-2 with acidic GAA activity which sufficiently compensates the deficient AAGR-1 activity. We speculate that double knockout of both *C. elegans* gene *aagr-1* and *-2* could be a suitable model for GSD type II.

We identified the gene coding HGSNAT based on the genetic linkage analysis. Sequence verification confirmed that cDNA of HGSNAT is 4.5 kb long and contains 1992 bp long coding sequence composed of 18 exons. HGSNAT is ubiquitously expressed in various human tissues. Bioinformatic analyses reveal that HGSNAT contains N-terminal signal peptide, 11 predicted transmembrane domains and 4 potential N-glycosylation sites. Interestingly, the sequence homology searches indicate that human HGSNAT does not share homology with any known prokaryotic or eukaryotic N-acetyltransferase (including *C. elegans*) except its mammalian orthologs. Immunofluorescence microscopy analyses demonstrated that HGSNAT is present only in a fraction of lysosomes and is localised in microdomains of lysosomal membranes. The partial co-localisation of HGSNAT and sortilin suggests sortilin dependent lysosomal sorting of HGSNAT. The cellular distribution of HGSNAT does not seem to be influenced by lysosomal storage.

7. Abbreviation

aagr	acid α-glucosidases related
AP	adaptor protein
ASM	acid sphingomyelinase
BBB	blood – brain barrier
bp	base pairs
BSA	bovine serum albumin
C. briggsae	*Caenorhabditis briggsae*
C. elegans	*Caenorhabditis elegans*
CD-MPR	cation – dependent mannose-6-phosphate receptor
cDNA	complementary DNA
CGC	*Caenorhabditis* Genetics Center
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CI-MPR	cation - independent mannose-6-phosphate receptor
CMA	chaperone – mediated autophagy
CNS	central nervous system
Co A	coenzyme A
CON A	concanamycin A
DAPI	diamidinophenyl idole
dauers	dauer larvae
D-Gal	D-galactose
D-GalNAc	N-acetyl-D-galactosamine
DGNJ	deoxygalactonojirimicine
DIC	differential interference contrast
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOPE	Discrete Optimized Protein Energy
DRM	detergent resistant membrane
dsRNA	double-stranded RNA
E. coli	*Escherichia coli*
e.g.	exempli gratia
EE	early endosome
EEA	early endosome antigen
EET	enzyme enhancement therapy
ER	endoplasmic reticulum
ERAD	endoplasmatic reticulum – associated degradation
ERT	enzyme replacement therapy
EST	expressed sequence tag
EtDO-P4	1-ethylendioxyphenyl-2-palmitoylamino-3-pyrrolidinopropanol
FCS	fetal calf serum
Fig.	Figure
FITC	fluorescein isothiocyanate
Ga_2Cer	galabiosylceramide
GAA	acid α-glucosidase
GAG	glycosaminoglycan
GANA-1	α-GAL and α-NAGA *C. elegans* ortholog
GANAB	glucosidase II alpha subunit
GANC	neutral α-glucosidase C
Gb 3	globotriaosylceramide
Gb_3Cer	globotriaosylceramide
gDNA	genomic DNA
GFP	green fluorescent protein
GGA	monomeric adaptor protein
GH31	glycosyl hydrolases family 31
GSD II	glycogen storage disease type II
GSL	glycosphingolipids
HGSNAT	heparin acetyl-coenzyme A:α-glucosaminide *N*-acetyltransferase

HS	heparan sulphate
Hsc70	heat shock cognate protein 70
ICD	I-cell disease
IMD	inherited metabolic diseases
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase
L1-4	larval stages 1-4
LAMP	lysosomal associated membrane protein
LAP	lysosomal acid phosphatase
LE	late endosome
LIMP	lysosomal integral membrane protein
LMP	lysosomal membrane protein
LRO	lysosome – related organelles
LSD	lysosomal storage diseases
Ly-hsc70	lysosomal heat shock cognate protein 70
Ma	mouse
MALS	macroautophagy/lysosomal system
M6P	mannose-6-phosphate
Mb	megabase
MGA	maltase – glucoamylase
min	minute
MLII	mucolipidosis II
MPR	mannose-6-phosphate receptor
MPS	mucopolysaccharidoses
mRNA	messenger RNA
MU	methylumbelliferyl
NCL	neuronal ceroid lipofuscinosis
NCRR	National Center for Research Resources
NGM	nematode grow medium
NLS	nuclear localization sequence
NPC	Niemann-Pick type C
NtMGA	N-terminal subunit of human maltase-glucoamylase
ORF	open reading frame
OST	open reading frame sequence tag
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PSF	point spread function
Ra	rabbit
RNA	ribonucleic acid
RNAi	RNA - mediated interference
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
s	second
SAPLIP	saposin – like protein
SDS	sodium dodecyl sulfate
SL	splice leader
SNP	single nucleotide polymorphism
SRT	substrate reduction therapy
SUIS	sucrase – isomaltase
swPCR	single worm PCR
TGN	*trans* Golgi network
TMEM	transmembrane protein
TPP1	tripeptidyl peptidase I
UTR	untranslated region
α-GAL	α-galactosidase
α-Glu	α-glucosidase
α-NAGA	α-*N*-acetylgalactosaminidase
β-Hex	β-hexosaminidase

8. References

1. Ruvkun, G. and O. Hobert, *The taxonomy of developmental control in Caenorhabditis elegans.* Science, 1998. **282**(5396): p. 2033-41.
2. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. **77**(1): p. 71-94.
3. Riddle, D.L., et al., *C. elegans II*. 1997, New York: Cold Spring Harbor Laboratory Press.
4. Anderson, P., *Mutagenesis.* Methods Cell Biol, 1995. **48**: p. 31-58.
5. Jorgensen, E.M. and S.E. Mango, *The art and design of genetic screens: caenorhabditis elegans.* Nat Rev Genet, 2002. **3**(5): p. 356-69.
6. Wood, W.B., ed. *The nematode Caenorhabditis elegans*. 1988, Cold Spring Harbor Laboratory Press: New York, NY.
7. *Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.* Science, 1998. **282**(5396): p. 2012-8.
8. Kamath, R.S., et al., *Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans.* Genome Biol, 2000. **2**(1).
9. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
10. Piano, F., et al., *RNAi analysis of genes expressed in the ovary of Caenorhabditis elegans.* Curr Biol, 2000. **10**(24): p. 1619-1622.
11. Fraser, A.G., et al., *Functional genomic analysis of C. elegans chromosome I by systematic RNA interference.* Nature, 2000. **408**(6810): p. 325-30.
12. Ashrafi, K., et al., *Genome-wide RNAi analysis of Caenorhabditis elegans fat regulatory genes.* Nature, 2003. **421**(6920): p. 268-72.
13. Pothof, J., et al., *Identification of genes that protect the C. elegans genome against mutations by genome-wide RNAi.* Genes Dev, 2003. **17**(4): p. 443-8.
14. Kamath, R.S., et al., *Systematic functional analysis of the Caenorhabditis elegans genome using RNAi.* Nature, 2003. **421**(6920): p. 231-7.
15. Fire, A., S.W. Harrison, and D. Dixon, *A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans.* Gene, 1990. **93**(2): p. 189-98.
16. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.
17. Cassada, R.C. and R.L. Russell, *The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans.* Dev Biol, 1975. **46**(2): p. 326-42.
18. Hope, I.A., ed. *C. elegans: a practical approach*. The Practical Approach Series, ed. B.D. Hames. 1999, Oxford University Press: Oxford.
19. *Wormatlas*. [cited; Available from: http://www.wormatlas.org.
20. Sulston, J.E., et al., *The embryonic cell lineage of the nematode Caenorhabditis elegans.* Dev Biol, 1983. **100**(1): p. 64-119.
21. Deppe, U., Schierenberg, E., Cole, T., Krieg, C., Schmitt, D., Yoder, B., and von Ehrenstein, G., *Cell lineages of the embryo of the nematode Caenorhabditis elegans.* Proc. Natl. Acad. Sci., 1978. **75**: p. 376-380.
22. Ambros, V., *Control of developmental timing in Caenorhabditis elegans.* Curr Opin Genet Dev, 2000. **10**(4): p. 428-33.
23. Johnson, T.E., et al., *Arresting development arrests aging in the nematode Caenorhabditis elegans.* Mech Ageing Dev, 1984. **28**(1): p. 23-40.
24. Slack, F. and G. Ruvkun, *Temporal pattern formation by heterochronic genes.* Annu Rev Genet, 1997. **31**: p. 611-34.
25. *Caenorhabditis elegans WWW Server*. [cited; Available from: http://elegans.swmed.edu.
26. *WormBook*. [cited; Available from: http://www.wormbook.org.
27. *ORFeome project database WorfDB* [cited; Available from: http://worfdb.dfci.harvard.edu.
28. Vaglio, P., et al., *WorfDB: the Caenorhabditis elegans ORFeome Database.* Nucleic Acids Res, 2003. **31**(1): p. 237-40.
29. *WormBase*. [cited; Available from: http://www.wormbase.org.
30. Chen, N., et al., *WormBase as an integrated platform for the C. elegans ORFeome.* Genome Res, 2004. **14**(10B): p. 2155-61.
31. Chen, N., et al., *WormBase: a comprehensive data resource for Caenorhabditis biology and genomics.* Nucleic Acids Res, 2005. **33**(Database issue): p. D383-9.
32. Bieri, T., et al., *WormBase: new content and better access.* Nucleic Acids Res, 2007. **35**(Database issue): p. D506-10.

33. Rogers, A., et al., *WormBase 2007.* Nucleic Acids Res, 2007.
34. *WormBase Release WS188.* [cited; Available from: http://www.wormbase.org/wiki/index.php/WS188.
35. Mushegian, A.R., et al., *Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes.* Genome Res, 1998. **8**(6): p. 590-8.
36. de Voer, G., D. Peters, and P.E. Taschner, *Caenorhabditis elegans as a model for lysosomal storage disorders.* Biochim Biophys Acta, 2008. **1782**(7-8): p. 433-46.
37. Spieth, J. and D. Lawson, *Overview of gene structure.* WormBook, ed. The C. elegans Research Community, 2006: p. doi/10.1895/wormbook.1.65.1, http://www.wormbook.org.
38. Harrison, P.M., N. Echols, and M.B. Gerstein, *Digging for dead genes: an analysis of the characteristics of the pseudogene population in the Caenorhabditis elegans genome.* Nucleic Acids Res, 2001. **29**(3): p. 818-30.
39. Blumenthal, T., et al., *A global analysis of Caenorhabditis elegans operons.* Nature, 2002. **417**(6891): p. 851-4.
40. Blumenthal, T., *Trans-splicing and operons.* WormBook, 2005: p. 1-9.
41. Liu, Y., et al., *An uncapped RNA suggests a model for Caenorhabditis elegans polycistronic pre-mRNA processing.* Rna, 2003. **9**(6): p. 677-87.
42. Nimmo, R. and A. Woollard, *Widespread organisation of C. elegans genes into operons: fact or function?* Bioessays, 2002. **24**(11): p. 983-7.
43. Blumenthal, T. and K.S. Gleason, *Caenorhabditis elegans operons: form and function.* Nat Rev Genet, 2003. **4**(2): p. 112-20.
44. Sulston, J.E. and H.R. Horvitz, *Post-embryonic cell lineages of the nematode, Caenorhabditis elegans.* Dev Biol, 1977. **56**(1): p. 110-56.
45. Francis, G.R. and R.H. Waterston, *Muscle organization in Caenorhabditis elegans: localization of proteins implicated in thin filament attachment and I-band organization.* J Cell Biol, 1985. **101**(4): p. 1532-49.
46. Hedgecock, E.M. and J.G. White, *Polyploid tissues in the nematode Caenorhabditis elegans.* Dev Biol, 1985. **107**(1): p. 128-33.
47. Ward, S. and J.S. Carrel, *Fertilization and sperm competition in the nematode Caenorhabditis elegans.* Dev Biol, 1979. **73**(2): p. 304-21.
48. Garcia, L.R., P. Mehta, and P.W. Sternberg, *Regulation of distinct muscle behaviors controls the C. elegans male's copulatory spicules during mating.* Cell, 2001. **107**(6): p. 777-88.
49. Liu, K.S. and P.W. Sternberg, *Sensory regulation of male mating behavior in Caenorhabditis elegans.* Neuron, 1995. **14**(1): p. 79-89.
50. Fares, H. and I. Greenwald, *Genetic analysis of endocytosis in Caenorhabditis elegans: coelomocyte uptake defective mutants.* Genetics, 2001. **159**(1): p. 133-45.
51. Fares, H. and B. Grant, *Deciphering endocytosis in Caenorhabditis elegans.* Traffic, 2002. **3**(1): p. 11-9.
52. Schroeder, L.K., et al., *Function of the Caenorhabditis elegans ABC transporter PGP-2 in the biogenesis of a lysosome-related fat storage organelle.* Mol Biol Cell, 2007. **18**(3): p. 995-1008.
53. Hermann, G.J., et al., *Genetic analysis of lysosomal trafficking in Caenorhabditis elegans.* Mol Biol Cell, 2005. **16**(7): p. 3273-88.
54. Artal-Sanz, M., et al., *Lysosomal biogenesis and function is critical for necrotic cell death in Caenorhabditis elegans.* J Cell Biol, 2006. **173**(2): p. 231-9.
55. Chen, C.C., et al., *RAB-10 is required for endocytic recycling in the Caenorhabditis elegans intestine.* Mol Biol Cell, 2006. **17**(3): p. 1286-97.
56. Sato, M., et al., *Caenorhabditis elegans RME-6 is a novel regulator of RAB-5 at the clathrin-coated pit.* Nat Cell Biol, 2005. **7**(6): p. 559-69.
57. Treusch, S., et al., *Caenorhabditis elegans functional orthologue of human protein h-mucolipin-1 is required for lysosome biogenesis.* Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4483-8.
58. Hersh, B.M., E. Hartwieg, and H.R. Horvitz, *The Caenorhabditis elegans mucolipin-like gene cup-5 is essential for viability and regulates lysosomes in multiple cell types.* Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4355-60.
59. Nicot, A.S., et al., *The phosphoinositide kinase PIKfyve/Fab1p regulates terminal lysosome maturation in Caenorhabditis elegans.* Mol Biol Cell, 2006. **17**(7): p. 3062-74.
60. Larsen, M.K., et al., *MAA-1, a novel acyl-CoA-binding protein involved in endosomal vesicle transport in Caenorhabditis elegans.* Mol Biol Cell, 2006. **17**(10): p. 4318-29.
61. Kostich, M., A. Fire, and D.M. Fambrough, *Identification and molecular-genetic characterization of a LAMP/CD68-like protein from Caenorhabditis elegans.* J Cell Sci, 2000. **113 (Pt 14)**: p. 2595-606.

62. Sebastiano, M., M. D'Alessio, and P. Bazzicalupo, *Beta-glucuronidase mutants of the nematode Caenorhabditis elegans.* Genetics, 1986. **112**(3): p. 459-68.
63. Jacobson, L.A., et al., *Identification of a putative structural gene for cathepsin D in Caenorhabditis elegans.* Genetics, 1988. **119**(2): p. 355-63.
64. Bolanowski, M.A., L.A. Jacobson, and R.L. Russell, *Quantitative measures of aging in the nematode Caenorhabditis elegans: II. Lysosomal hydrolases as markers of senescence.* Mech Ageing Dev, 1983. **21**(3-4): p. 295-319.
65. Gerdt, S., et al., *Isolation and structural analysis of three neutral glycosphingolipids from a mixed population of Caenorhabditis elegans (Nematoda:Rhabditida).* Glycobiology, 1997. **7**(2): p. 265-75.
66. Bennett, E.P., et al., *Genomic organization and chromosomal localization of three members of the UDP-N-acetylgalactosamine: polypeptide N-acetylgalactosaminyltransferase family.* Glycobiology, 1998. **8**(6): p. 547-55.
67. Bulik, D.A., et al., *sqv-3, -7, and -8, a set of genes affecting morphogenesis in Caenorhabditis elegans, encode enzymes required for glycosaminoglycan biosynthesis.* Proc Natl Acad Sci U S A, 2000. **97**(20): p. 10838-43.
68. Hwang, H.Y., et al., *The Caenorhabditis elegans genes sqv-2 and sqv-6, which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase.* J Biol Chem, 2003. **278**(14): p. 11735-8.
69. Berninsone, P.M., *Carbohydrates and glycosylation.* WormBook, 2006: p. 1-22.
70. Matyash, V., et al., *Distribution and transport of cholesterol in Caenorhabditis elegans.* Mol Biol Cell, 2001. **12**(6): p. 1725-36.
71. Choi, B.K., D.J. Chitwood, and Y.K. Paik, *Proteomic changes during disturbance of cholesterol metabolism by azacoprostane treatment in Caenorhabditis elegans.* Mol Cell Proteomics, 2003. **2**(10): p. 1086-95.
72. Sym, M., M. Basson, and C. Johnson, *A model for niemann-pick type C disease in the nematode Caenorhabditis elegans.* Curr Biol, 2000. **10**(9): p. 527-30.
73. *The Nobel Prize in Physiology or Medicin 2006.* [cited; Available from: http://nobelprize.org/nobel_prizes/medicine/laureates/2006/adv.html.
74. Tomari, Y. and P.D. Zamore, *Perspective: machines for RNAi.* Genes Dev, 2005. **19**(5): p. 517-29.
75. Kim, D.H. and J.J. Rossi, *Strategies for silencing human disease using RNA interference.* Nat Rev Genet, 2007. **8**(3): p. 173-84.
76. Grishok, A., H. Tabara, and C.C. Mello, *Genetic requirements for inheritance of RNAi in C. elegans.* Science, 2000. **287**(5462): p. 2494-7.
77. Timmons, L., D.L. Court, and A. Fire, *Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans.* Gene, 2001. **263**(1-2): p. 103-12.
78. Kuwabara, P.E. and N. O'Neil, *The use of functional genomics in C. elegans for studying human development and disease.* J Inherit Metab Dis, 2001. **24**(2): p. 127-38.
79. Rugarli, E.I., et al., *The Kallmann syndrome gene homolog in C. elegans is involved in epidermal morphogenesis and neurite branching.* Development, 2002. **129**(5): p. 1283-94.
80. Lee, M.H., et al., *The gene expression and deficiency phenotypes of Cockayne syndrome B protein in Caenorhabditis elegans.* FEBS Lett, 2002. **522**(1-3): p. 47-51.
81. Kaletta, T. and M.O. Hengartner, *Finding function in novel targets: C. elegans as a model organism.* Nat Rev Drug Discov, 2006. **5**(5): p. 387-98.
82. Niwa, R., et al., *The expression of the Alzheimer's amyloid precursor protein-like gene is regulated by developmental timing microRNAs and their targets in Caenorhabditis elegans.* Dev Biol, 2008. **315**(2): p. 418-25.
83. de Voer, G., et al., *Deletion of the Caenorhabditis elegans homologues of the CLN3 gene, involved in human juvenile neuronal ceroid lipofuscinosis, causes a mild progeric phenotype.* J Inherit Metab Dis, 2005. **28**(6): p. 1065-80.
84. Lin, X., M.O. Hengartner, and R. Kolesnick, *Caenorhabditis elegans contains two distinct acid sphingomyelinases.* J Biol Chem, 1998. **273**(23): p. 14374-9.
85. Fares, H. and I. Greenwald, *Regulation of endocytosis by CUP-5, the Caenorhabditis elegans mucolipin-1 homolog.* Nat Genet, 2001. **28**(1): p. 64-8.
86. Schaheen, L., H. Dang, and H. Fares, *Basis of lethality in C. elegans lacking CUP-5, the Mucolipidosis Type IV orthologue.* Dev Biol, 2006. **293**(2): p. 382-91.
87. De Duve, C., et al., *Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue.* Biochem J, 1955. **60**(4): p. 604-17.
88. Beaufay, H., C. De Duve, and A.B. Novikoff, *Electron microscopy of lysosomerich fractions from rat liver.* J Biophys Biochem Cytol, 1956. **2**(4, Suppl): p. 179-84.
89. Weisz, O.A., *Organelle acidification and disease.* Traffic, 2003. **4**(2): p. 57-64.

90. Demaurex, N., *pH Homeostasis of cellular organelles.* News Physiol Sci, 2002. **17**: p. 1-5.
91. Schroder, B., et al., *Integral and Associated Lysosomal Membrane Proteins.* Traffic, 2007.
92. Gough, N.R. and D.M. Fambrough, *Different steady state subcellular distributions of the three splice variants of lysosome-associated membrane protein LAMP-2 are determined largely by the COOH-terminal amino acid residue.* J Cell Biol, 1997. **137**(5): p. 1161-9.
93. Isaac, E.L., et al., *Regulation of the lysosome-associated membrane protein in a sucrose model of lysosomal storage.* Exp Cell Res, 2000. **254**(2): p. 204-9.
94. Janvier, K. and J.S. Bonifacino, *Role of the endocytic machinery in the sorting of lysosome-associated membrane proteins.* Mol Biol Cell, 2005. **16**(9): p. 4231-42.
95. Barriocanal, J.G., et al., *Biosynthesis, glycosylation, movement through the Golgi system, and transport to lysosomes by an N-linked carbohydrate-independent mechanism of three lysosomal integral membrane proteins.* J Biol Chem, 1986. **261**(35): p. 16755-63.
96. Vega, M.A., et al., *Cloning, sequencing, and expression of a cDNA encoding rat LIMP II, a novel 74-kDa lysosomal membrane protein related to the surface adhesion protein CD36.* J Biol Chem, 1991. **266**(25): p. 16818-24.
97. Eskelinen, E.L., Y. Tanaka, and P. Saftig, *At the acidic edge: emerging functions for lysosomal membrane proteins.* Trends Cell Biol, 2003. **13**(3): p. 137-45.
98. Andrews, N.W., *Regulated secretion of conventional lysosomes.* Trends Cell Biol, 2000. **10**(8): p. 316-21.
99. Cuervo, A.M. and J.F. Dice, *Lysosomes, a meeting point of proteins, chaperones, and proteases.* J Mol Med, 1998. **76**(1): p. 6-12.
100. Dell'Angelica, E.C., et al., *Lysosome-related organelles.* Faseb J, 2000. **14**(10): p. 1265-78.
101. Ungewickell, E.J. and L. Hinrichsen, *Endocytosis: clathrin-mediated membrane budding.* Curr Opin Cell Biol, 2007. **19**(4): p. 417-25.
102. Lajoie, P. and I.R. Nabi, *Regulation of raft-dependent endocytosis.* J Cell Mol Med, 2007. **11**(4): p. 644-53.
103. Klionsky, D.J., A.M. Cuervo, and P.O. Seglen, *Methods for monitoring autophagy from yeast to human.* Autophagy, 2007. **3**(3): p. 181-206.
104. Klionsky, D.J., et al., *Guidelines for the use and interpretation of assays for monitoring autophagy in higher eukaryotes.* Autophagy, 2008. **4**(2): p. 151-75.
105. Cuervo, A.M., *Autophagy: many paths to the same end.* Mol Cell Biochem, 2004. **263**(1-2): p. 55-72.
106. Levine, B. and D.J. Klionsky, *Development by self-digestion: molecular mechanisms and biological functions of autophagy.* Dev Cell, 2004. **6**(4): p. 463-77.
107. Overbye, A., M. Fengsrud, and P.O. Seglen, *Proteomic analysis of membrane-associated proteins from rat liver autophagosomes.* Autophagy, 2007. **3**(4): p. 300-22.
108. Winchester, B., *Lysosomal metabolism of glycoproteins.* Glycobiology, 2005. **15**(6): p. 1R-15R.
109. Kaushik, S. and A.M. Cuervo, *Autophagy as a cell-repair mechanism: activation of chaperone-mediated autophagy during oxidative stress.* Mol Aspects Med, 2006. **27**(5-6): p. 444-54.
110. Massey, A.C., et al., *Consequences of the selective blockage of chaperone-mediated autophagy.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5805-10.
111. Cuervo, A.M., et al., *Selective degradation of annexins by chaperone-mediated autophagy.* J Biol Chem, 2000. **275**(43): p. 33329-35.
112. Agarraberes, F.A. and J.F. Dice, *A molecular chaperone complex at the lysosomal membrane is required for protein translocation.* J Cell Sci, 2001. **114**(Pt 13): p. 2491-9.
113. Majeski, A.E. and J.F. Dice, *Mechanisms of chaperone-mediated autophagy.* Int J Biochem Cell Biol, 2004. **36**(12): p. 2435-44.
114. Cuervo, A.M., et al., *Cathepsin A regulates chaperone-mediated autophagy through cleavage of the lysosomal receptor.* Embo J, 2003. **22**(1): p. 47-59.
115. Kaushik, S., A.C. Massey, and A.M. Cuervo, *Lysosome membrane lipid microdomains: novel regulators of chaperone-mediated autophagy.* Embo J, 2006. **25**(17): p. 3921-33.
116. Stuart, L.M. and R.A. Ezekowitz, *Phagocytosis and comparative innate immunity: learning on the fly.* Nat Rev Immunol, 2008. **8**(2): p. 131-41.
117. Sleat, D.E., H. Zheng, and P. Lobel, *The human urine mannose 6-phosphate glycoproteome.* Biochim Biophys Acta, 2007. **1774**(3): p. 368-72.
118. Lazzarino, D.A. and C.A. Gabel, *Biosynthesis of the mannose 6-phosphate recognition marker in transport-impaired mouse lymphoma cells. Demonstration of a two-step phosphorylation.* J Biol Chem, 1988. **263**(21): p. 10118-26.
119. Tikkanen, R., et al., *Several cooperating binding sites mediate the interaction of a lysosomal enzyme with phosphotransferase.* Embo J, 1997. **16**(22): p. 6684-93.

120. Tong, P.Y. and S. Kornfeld, *Ligand interactions of the cation-dependent mannose 6-phosphate receptor. Comparison with the cation-independent mannose 6-phosphate receptor.* J Biol Chem, 1989. **264**(14): p. 7970-5.
121. Tong, P.Y., W. Gregory, and S. Kornfeld, *Ligand interactions of the cation-independent mannose 6-phosphate receptor. The stoichiometry of mannose 6-phosphate binding.* J Biol Chem, 1989. **264**(14): p. 7962-9.
122. Dahms, N.M., P. Lobel, and S. Kornfeld, *Mannose 6-phosphate receptors and lysosomal enzyme targeting.* J Biol Chem, 1989. **264**(21): p. 12115-8.
123. Hirst, J., et al., *The role of cargo proteins in GGA recruitment.* Traffic, 2007. **8**(5): p. 594-604.
124. McComick, P., et al., *Palmitoylation Is Required for the Retrograde Transport of the Lysosomal Sorting Receptors*, in *ASCB Annual Meeting*. 2007: Washington, DC.
125. Canfield, W.M., et al., *Localization of the signal for rapid internalization of the bovine cation-independent mannose 6-phosphate/insulin-like growth factor-II receptor to amino acids 24-29 of the cytoplasmic tail.* J Biol Chem, 1991. **266**(9): p. 5682-8.
126. Johnson, K.F. and S. Kornfeld, *The cytoplasmic tail of the mannose 6-phosphate/insulin-like growth factor-II receptor has two signals for lysosomal enzyme sorting in the Golgi.* J Cell Biol, 1992. **119**(2): p. 249-57.
127. Johnson, K.F. and S. Kornfeld, *A His-Leu-Leu sequence near the carboxyl terminus of the cytoplasmic domain of the cation-dependent mannose 6-phosphate receptor is necessary for the lysosomal enzyme sorting function.* J Biol Chem, 1992. **267**(24): p. 17110-5.
128. Nair, P., B.E. Schaub, and J. Rohrer, *Characterization of the endosomal sorting signal of the cation-dependent mannose 6-phosphate receptor.* J Biol Chem, 2003. **278**(27): p. 24753-8.
129. Linke, M., V. Herzog, and K. Brix, *Trafficking of lysosomal cathepsin B-green fluorescent protein to the surface of thyroid epithelial cells involves the endosomal/lysosomal compartment.* J Cell Sci, 2002. **115**(Pt 24): p. 4877-89.
130. Hirschhorn, R.a.R., A.J.J., *Glycogen Storage Disease Type II: Acid α-Glucosidase (Acid Maltase) Deficiency*, in *The Metabolic and Molecular Bases of Inherited Disease*, C.R. Scriver, Beaudet, A. L., Sly, W. S, and Valle, D, Editor. 2001, McGraw-Hill: New York. p. 3389-3420.
131. Peters, C. and K. von Figura, *Biogenesis of lysosomal membranes.* FEBS Lett, 1994. **346**(1): p. 108-14.
132. Kornfeld, S.a.S., W. S., *I-Cell Disease and Pseudo-Hurler Polydystrofy: Disorders of Lysosomal Emzyme Phosphorylation* in *The Metabolic and Molecular Bases of Inherited Disease*, C.R. Scriver, Beaudet, A. L., Sly, W. S, and Valle, D, Editor. 2001, McGraw-Hill: New York. p. 3469-3482.
133. Dittmer, F., et al., *Alternative mechanisms for trafficking of lysosomal enzymes in mannose 6-phosphate receptor-deficient mice are cell type-specific.* J Cell Sci, 1999. **112 (Pt 10)**: p. 1591-7.
134. Lefrancois, S., et al., *The lysosomal trafficking of sphingolipid activator proteins (SAPs) is mediated by sortilin.* Embo J, 2003. **22**(24): p. 6430-7.
135. Lefrancois, S., et al., *Inactivation of sortilin (a novel lysosomal sorting receptor) by dominant negative competition and RNA interference.* Biol Proced Online, 2005. **7**: p. 17-25.
136. Ni, X. and C.R. Morales, *The lysosomal trafficking of acid sphingomyelinase is mediated by sortilin and mannose 6-phosphate receptor.* Traffic, 2006. **7**(7): p. 889-902.
137. Hassan, A.J., et al., *The trafficking of prosaposin (SGP-1) and GM2AP to the lysosomes of TM4 Sertoli cells is mediated by sortilin and monomeric adaptor proteins.* Mol Reprod Dev, 2004. **68**(4): p. 476-83.
138. Canuel, M., et al., *Sortilin mediates the lysosomal targeting of cathepsins D and H.* Biochem Biophys Res Commun, 2008. **373**(2): p. 292-7.
139. Reczek, D., et al., *LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of beta-glucocerebrosidase.* Cell, 2007. **131**(4): p. 770-83.
140. Obermuller, S., et al., *The tyrosine motifs of Lamp 1 and LAP determine their direct and indirect targetting to lysosomes.* J Cell Sci, 2002. **115**(Pt 1): p. 185-94.
141. Honing, S., et al., *The tyrosine-based lysosomal targeting signal in lamp-1 mediates sorting into Golgi-derived clathrin-coated vesicles.* Embo J, 1996. **15**(19): p. 5230-9.
142. Karlsson, K. and S.R. Carlsson, *Sorting of lysosomal membrane glycoproteins lamp-1 and lamp-2 into vesicles distinct from mannose 6-phosphate receptor/gamma-adaptin vesicles at the trans-Golgi network.* J Biol Chem, 1998. **273**(30): p. 18966-73.
143. Peden, A.A., et al., *Localization of the AP-3 adaptor complex defines a novel endosomal exit site for lysosomal membrane proteins.* J Cell Biol, 2004. **164**(7): p. 1065-76.
144. Kametaka, S., et al., *Canonical interaction of cyclin G associated kinase with adaptor protein 1 regulates lysosomal enzyme sorting.* Mol Biol Cell, 2007. **18**(8): p. 2991-3001.
145. Nielsen, M.S., et al., *The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein.* Embo J, 2001. **20**(9): p. 2180-90.

146. Puertollano, R., et al., *Sorting of mannose 6-phosphate receptors mediated by the GGAs.* Science, 2001. **292**(5522): p. 1712-6.
147. Meikle, P.J., et al., *Prevalence of lysosomal storage disorders.* Jama, 1999. **281**(3): p. 249-54.
148. *OMIM - Online Mendelian Inheritance in Man*. [cited; Available from: http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim.
149. Schulz, A., et al., *Impaired cell adhesion and apoptosis in a novel CLN9 Batten disease variant.* Ann Neurol, 2004. **56**(3): p. 342-50.
150. Steinfeld, R., et al., *Cathepsin D deficiency is associated with a human neurodegenerative disorder.* Am J Hum Genet, 2006. **78**(6): p. 988-98.
151. Siintola, E., et al., *The novel neuronal ceroid lipofuscinosis gene MFSD8 encodes a putative lysosomal transporter.* Am J Hum Genet, 2007. **81**(1): p. 136-46.
152. Carstea, E.D., et al., *Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis.* Science, 1997. **277**(5323): p. 228-31.
153. Morris, J.A., et al., *The genomic organization and polymorphism analysis of the human Niemann-Pick C1 gene.* Biochem Biophys Res Commun, 1999. **261**(2): p. 493-8.
154. Naureckiene, S., et al., *Identification of HE1 as the second gene of Niemann-Pick C disease.* Science, 2000. **290**(5500): p. 2298-301.
155. Kudo, M., et al., *The alpha- and beta-subunits of the human UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase [corrected] are encoded by a single cDNA.* J Biol Chem, 2005. **280**(43): p. 36141-9.
156. Kudo, M., M.S. Brem, and W.M. Canfield, *Mucolipidosis II (I-cell disease) and mucolipidosis IIIA (classical pseudo-hurler polydystrophy) are caused by mutations in the GlcNAc-phosphotransferase alpha / beta -subunits precursor gene.* Am J Hum Genet, 2006. **78**(3): p. 451-63.
157. Bassi, M.T., et al., *Cloning of the gene encoding a novel integral membrane protein, mucolipidin-and identification of the two major founder mutations causing mucolipidosis type IV.* Am J Hum Genet, 2000. **67**(5): p. 1110-20.
158. Nishino, I., et al., *Primary LAMP-2 deficiency causes X-linked vacuolar cardiomyopathy and myopathy (Danon disease).* Nature, 2000. **406**(6798): p. 906-10.
159. Desnick, R.J., *Enzyme replacement and enhancement therapies for lysosomal diseases.* J Inherit Metab Dis, 2004. **27**(3): p. 385-410.
160. Grabowski, G.A., et al., *Enzyme therapy in type 1 Gaucher disease: comparative efficacy of mannose-terminated glucocerebrosidase from natural and recombinant sources.* Ann Intern Med, 1995. **122**(1): p. 33-9.
161. Mignani, R. and L. Cagnoli, *Enzyme replacement therapy in Fabry's disease: recent advances and clinical applications.* J Nephrol, 2004. **17**(3): p. 354-63.
162. Clarke, J.T. and R.M. Iwanochko, *Enzyme replacement therapy of Fabry disease.* Mol Neurobiol, 2005. **32**(1): p. 43-50.
163. Wraith, J.E., et al., *Enzyme replacement therapy for mucopolysaccharidosis I: a randomized, double-blinded, placebo-controlled, multinational study of recombinant human alpha-L-iduronidase (laronidase).* J Pediatr, 2004. **144**(5): p. 581-8.
164. Rossi, M., et al., *Long-term enzyme replacement therapy for pompe disease with recombinant human alpha-glucosidase derived from chinese hamster ovary cells.* J Child Neurol, 2007. **22**(5): p. 565-73.
165. Koeberl, D.D., P.S. Kishnani, and Y.T. Chen, *Glycogen storage disease types I and II: treatment updates.* J Inherit Metab Dis, 2007. **30**(2): p. 159-64.
166. Sakuraba, H., et al., *Molecular pathologies of and enzyme replacement therapies for lysosomal diseases.* CNS Neurol Disord Drug Targets, 2006. **5**(4): p. 401-13.
167. Brady, R.O. and R. Schiffmann, *Enzyme-replacement therapy for metabolic storage disorders.* Lancet Neurol, 2004. **3**(12): p. 752-6.
168. Hoogerbrugge, P.M., et al., *Allogeneic bone marrow transplantation for lysosomal storage diseases. The European Group for Bone Marrow Transplantation.* Lancet, 1995. **345**(8962): p. 1398-402.
169. Barranger, J.M. and E.A. Novelli, *Gene therapy for lysosomal storage disorders.* Expert Opin Biol Ther, 2001. **1**(5): p. 857-67.
170. Zheng, Y., et al., *Treatment of the mouse model of mucopolysaccharidosis I with retrovirally transduced bone marrow.* Mol Genet Metab, 2003. **79**(4): p. 233-44.
171. Cheng, S.H. and A.E. Smith, *Gene therapy progress and prospects: gene therapy of lysosomal storage disorders.* Gene Ther, 2003. **10**(16): p. 1275-81.
172. Eto, Y., et al., *Treatment of lysosomal storage disorders: cell therapy and gene therapy.* J Inherit Metab Dis, 2004. **27**(3): p. 411-5.

173. Xu, F., et al., *Glycogen storage in multiple muscles of old GSD-II mice can be rapidly cleared after a single intravenous injection with a modified adenoviral vector expressing hGAA.* J Gene Med, 2005. **7**(2): p. 171-8.
174. Aerts, J.M., et al., *Biochemistry of glycosphingolipid storage disorders: implications for therapeutic intervention.* Philos Trans R Soc Lond B Biol Sci, 2003. **358**(1433): p. 905-14.
175. Elstein, D., et al., *Sustained therapeutic effects of oral miglustat (Zavesca, N-butyldeoxynojirimycin, OGT 918) in type I Gaucher disease.* J Inherit Metab Dis, 2004. **27**(6): p. 757-66.
176. Abe, A., et al., *Reduction of globotriaosylceramide in Fabry disease mice by substrate deprivation.* J Clin Invest, 2000. **105**(11): p. 1563-71.
177. Kolter, T. and M. Wendeler, *Chemical chaperones--a new concept in drug research.* Chembiochem, 2003. **4**(4): p. 260-4.
178. Yam, G.H., et al., *Pharmacological chaperone corrects lysosomal storage in Fabry disease caused by trafficking-incompetent variants.* Am J Physiol Cell Physiol, 2006. **290**(4): p. C1076-82.
179. Desnick, R.J., Ioannou, Y. A. and Eng, Ch. M., *α-Galactosidase A Deficiency: Fabry Disease*, in *The Metabolic and Molecular Bases of Inherited Disease*, C.R. Scriver, Beaudet, A. L., Sly, W. S, and Valle, D, Editor. 2001, McGraw-Hill: New York. p. 3733-3774.
180. Thompson, T.E. and T.W. Tillack, *Organization of glycosphingolipids in bilayers and plasma membranes of mammalian cells.* Annu Rev Biophys Biophys Chem, 1985. **14**: p. 361-86.
181. Schibanoff, J.M., S. Kamoshita, and J.S. O'Brien, *Tissue distribution of glycosphingolipids in a case of Fabry's disease.* J Lipid Res, 1969. **10**(5): p. 515-20.
182. Wherrett, J.R. and S.I. Hakomori, *Characterization of a blood group B glycolipid, accumulating in the pancreas of a patient with Fabry's disease.* J Biol Chem, 1973. **248**(9): p. 3046-51.
183. Asfaw, B., et al., *Defects in degradation of blood group A and B glycosphingolipids in Schindler and Fabry diseases.* J Lipid Res, 2002. **43**(7): p. 1096-104.
184. Bishop, D.F., R. Kornreich, and R.J. Desnick, *Structural organization of the human alpha-galactosidase A gene: further evidence for the absence of a 3' untranslated region.* Proc Natl Acad Sci U S A, 1988. **85**(11): p. 3903-7.
185. Chow, J.C. and C.J. Brown, *Forming facultative heterochromatin: silencing of an X chromosome in mammalian females.* Cell Mol Life Sci, 2003. **60**(12): p. 2586-603.
186. Bishop, D.F., et al., *Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the mature enzyme.* Proc Natl Acad Sci U S A, 1986. **83**(13): p. 4859-63.
187. Novo, F.J., et al., *Editing of human alpha-galactosidase RNA resulting in a pyrimidine to purine conversion.* Nucleic Acids Res, 1995. **23**(14): p. 2636-40.
188. Blom, D., et al., *Recombinant enzyme therapy for Fabry disease: absence of editing of human alpha-galactosidase A mRNA.* Am J Hum Genet, 2003. **72**(1): p. 23-31.
189. Garman, S.C. and D.N. Garboczi, *Structural basis of Fabry disease.* Mol Genet Metab, 2002. **77**(1-2): p. 3-11.
190. Henrissat, B. and G. Davies, *Structural and sequence-based classification of glycoside hydrolases.* Curr Opin Struct Biol, 1997. **7**(5): p. 637-44.
191. Henrissat, B. and G.J. Davies, *Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics.* Plant Physiol, 2000. **124**(4): p. 1515-9.
192. Murali, R., et al., *Crystallization and preliminary X-ray analysis of human alpha-galactosidase A complex.* J Mol Biol, 1994. **239**(4): p. 578-80.
193. Garman, S.C. and D.N. Garboczi, *The molecular defect leading to Fabry disease: structure of human alpha-galactosidase.* J Mol Biol, 2004. **337**(2): p. 319-35.
194. Fujimoto, Z., et al., *Crystal structure of rice alpha-galactosidase complexed with D-galactose.* J Biol Chem, 2003. **278**(22): p. 20313-8.
195. Fujimoto, Z., et al., *Crystallization and preliminary X-ray crystallographic studies of alpha-galactosidase I from Mortierella vinacea.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 12): p. 2289-91.
196. Askari, H., et al., *Cellular and tissue localization of globotriaosylceramide in Fabry disease.* Virchows Arch, 2007. **451**(4): p. 823-34.
197. Ashton-Prolla, P., et al., *Fabry disease: twenty-two novel mutations in the alpha-galactosidase A gene and genotype/phenotype correlations in severely and mildly affected hemizygotes and heterozygotes.* J Investig Med, 2000. **48**(4): p. 227-35.
198. Dobrovolny, R., et al., *Relationship between X-inactivation and clinical involvement in Fabry heterozygotes. Eleven novel mutations in the alpha-galactosidase A gene in the Czech and Slovak population.* J Mol Med, 2005. **83**(8): p. 647-54.
199. Mayes, J.S., et al., *Differential assay for lysosomal alpha-galactosidases in human tissues and its application to Fabry's disease.* Clin Chim Acta, 1981. **112**(2): p. 247-51.

200. Racchi, O., et al., *X chromosome inactivation patterns in normal females.* Blood Cells Mol Dis, 1998. **24**(4): p. 439-47.
201. Schiffmann, R., et al., *Long-term therapy with agalsidase alfa for Fabry disease: safety and effects on renal function in a home infusion setting.* Nephrol Dial Transplant, 2006. **21**(2): p. 345-54.
202. Banikazemi, M., et al., *Agalsidase-beta therapy for advanced Fabry disease: a randomized trial.* Ann Intern Med, 2007. **146**(2): p. 77-86.
203. Keslova-Veselikova, J., et al., *Replacement of alpha-galactosidase A in Fabry disease: effect on fibroblast cultures compared with biopsied tissues of treated patients.* Virchows Arch, 2008. **452**(6): p. 651-65.
204. Desnick, R.J.a.S., D, *α-N-Acetylgalactosaminidase Deficiency: Schindler Disease*, in *The Metabolic and Molecular Bases of Inherited Disease*, C.R. Scriver, Beaudet, A. L., Sly, W. S, and Valle, D, Editor. 2001, McGraw-Hill: New York. p. 3483-3505.
205. Kusiak, J.W., J.M. Quirk, and R.O. Brady, *Purification and properties of the two major isozymes of alpha-galactosidase from human placenta.* J Biol Chem, 1978. **253**(1): p. 184-90.
206. Schram, A.W., et al., *Enzymological properties and immunological characterization of alpha-galactosidase isoenzymes from normal and Fabry human liver.* Biochim Biophys Acta, 1977. **482**(1): p. 125-37.
207. Schram, A.W., M.N. Hamers, and J.M. Tager, *The identity of alpha-galactosidase B from human liver.* Biochim Biophys Acta, 1977. **482**(1): p. 138-44.
208. Wang, A.M. and R.J. Desnick, *Structural organization and complete sequence of the human alpha-N-acetylgalactosaminidase gene: homology with the alpha-galactosidase A gene provides evidence for evolution from a common ancestral gene.* Genomics, 1991. **10**(1): p. 133-42.
209. Wang, A.M., et al., *Murine alpha-N-acetylgalactosaminidase: isolation and expression of a full-length cDNA and genomic organization: further evidence of an alpha-galactosidase gene family.* Mol Genet Metab, 1998. **65**(2): p. 165-73.
210. Wang, A.M., D.F. Bishop, and R.J. Desnick, *Human alpha-N-acetylgalactosaminidase-molecular cloning, nucleotide sequence, and expression of a full-length cDNA. Homology with human alpha-galactosidase A suggests evolution from a common ancestral gene.* J Biol Chem, 1990. **265**(35): p. 21859-66.
211. Ohta, M., et al., *Human alpha-N-acetylgalactosaminidase: site occupancy and structure of N-linked oligosaccharides.* Glycobiology, 2000. **10**(3): p. 251-61.
212. Garman, S.C., et al., *The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases.* Structure, 2002. **10**(3): p. 425-34.
213. Kotoulas, O.B., S.A. Kalamidas, and D.J. Kondomerkos, *Glycogen autophagy.* Microsc Res Tech, 2004. **64**(1): p. 10-20.
214. Hoefsloot, L.H., et al., *Characterization of the human lysosomal alpha-glucosidase gene.* Biochem J, 1990. **272**(2): p. 493-7.
215. Martiniuk, F., et al., *Isolation of a cDNA for human acid alpha-glucosidase and detection of genetic heterogeneity for mRNA in three alpha-glucosidase-deficient patients.* Proc Natl Acad Sci U S A, 1986. **83**(24): p. 9641-4.
216. Kroos, M., et al., *Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating.* Hum Mutat, 2008. **29**(6): p. E13-26.
217. Hermans, M.M., et al., *Human lysosomal alpha-glucosidase: functional characterization of the glycosylation sites.* Biochem J, 1993. **289 (Pt 3)**: p. 681-6.
218. Hoefsloot, L.H., et al., *Primary structure and processing of lysosomal alpha-glucosidase; homology with the intestinal sucrase-isomaltase complex.* Embo J, 1988. **7**(6): p. 1697-704.
219. Lovering, A.L., et al., *Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate.* J Biol Chem, 2005. **280**(3): p. 2105-15.
220. Frandsen, T.P. and B. Svensson, *Plant alpha-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin.* Plant Mol Biol, 1998. **37**(1): p. 1-13.
221. Lee, S.S., S. He, and S.G. Withers, *Identification of the catalytic nucleophile of the Family 31 alpha-glucosidase from Aspergillus niger via trapping of a 5-fluoroglycosyl-enzyme intermediate.* Biochem J, 2001. **359**(Pt 2): p. 381-6.
222. Angelov, A., M. Putyrski, and W. Liebl, *Molecular and biochemical characterization of alpha-glucosidase and alpha-mannosidase and their clustered genes from the thermoacidophilic archaeon Picrophilus torridus.* J Bacteriol, 2006. **188**(20): p. 7123-31.
223. Hermans, M.M., et al., *Human lysosomal alpha-glucosidase. Characterization of the catalytic site.* J Biol Chem, 1991. **266**(21): p. 13507-12.

224. Ernst, H.A., et al., *Structure of the Sulfolobus solfataricus alpha-glucosidase: implications for domain conservation and substrate recognition in GH31.* J Mol Biol, 2006. **358**(4): p. 1106-24.
225. Sim, L., et al., *Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity.* J Mol Biol, 2008. **375**(3): p. 782-92.
226. Okumiya, T., et al., *A new diagnostic assay for glycogen storage disease type II in mixed leukocytes.* Mol Genet Metab, 2006. **88**(1): p. 22-8.
227. Pauly, D.F., et al., *Complete correction of acid alpha-glucosidase deficiency in Pompe disease fibroblasts in vitro, and lysosomally targeted expression in neonatal rat cardiac and skeletal muscle.* Gene Ther, 1998. **5**(4): p. 473-80.
228. Mah, C., et al., *Sustained correction of glycogen storage disease type II using adeno-associated virus serotype 1 vectors.* Gene Ther, 2005. **12**(18): p. 1405-9.
229. Holt, S.J. and D.L. Riddle, *SAGE surveys C. elegans carbohydrate metabolism: evidence for an anaerobic shift in the long-lived dauer larva.* Mech Ageing Dev, 2003. **124**(7): p. 779-800.
230. Hanover, J.A., et al., *A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11266-71.
231. Neufeld, E.F.a.M., J., *The Mucopolysaccharidoses*, in *The Metabolic and Molecular Bases of Inherited Disease*, C.R. Scriver, Beaudet, A. L., Sly, W. S, and Valle, D, Editor. 2001, McGraw-Hill: New York. p. 3421-3452.
232. Wight, T.N., M.G. Kinsella, and E.E. Qwarnstrom, *The role of proteoglycans in cell adhesion, migration and proliferation.* Curr Opin Cell Biol, 1992. **4**(5): p. 793-801.
233. Esko, J.D. and U. Lindahl, *Molecular diversity of heparan sulfate.* J Clin Invest, 2001. **108**(2): p. 169-73.
234. Ausseil, J., et al., *Localisation of a gene for mucopolysaccharidosis IIIC to the pericentromeric region of chromosome 8.* J Med Genet, 2004. **41**(12): p. 941-5.
235. Ausseil, J., et al., *An acetylated 120-kDa lysosomal transmembrane protein is absent from mucopolysaccharidosis IIIC fibroblasts: a candidate molecule for MPS IIIC.* Mol Genet Metab, 2006. **87**(1): p. 22-31.
236. Fan, X., et al., *Identification of the gene encoding the enzyme deficient in mucopolysaccharidosis IIIC (Sanfilippo disease type C).* Am J Hum Genet, 2006. **79**(4): p. 738-44.
237. Bame, K.J. and L.H. Rome, *Acetyl coenzyme A: alpha-glucosaminide N-acetyltransferase. Evidence for a transmembrane acetylation mechanism.* J Biol Chem, 1985. **260**(20): p. 11293-9.
238. Bame, K.J. and L.H. Rome, *Genetic evidence for transmembrane acetylation by lysosomes.* Science, 1986. **233**(4768): p. 1087-9.
239. Bartsocas, C., et al., *Sanfilippo type C disease: clinical findings in four patients with a new variant of mucopolysaccharidosis III.* Eur J Pediatr, 1979. **130**(4): p. 251-8.
240. Voznyi Ya, V., et al., *A fluorimetric enzyme assay for the diagnosis of Sanfilippo disease C (MPS III C).* J Inherit Metab Dis, 1993. **16**(2): p. 465-72.
241. He, W., et al., *Prenatal diagnosis of Sanfilippo disease type C using a simple fluorometric enzyme assay.* Prenat Diagn, 1994. **14**(1): p. 17-22.
242. Taute, A., et al., *Presence of detergent-resistant microdomains in lysosomal membranes.* Biochem Biophys Res Commun, 2002. **298**(1): p. 5-9.
243. Pike, L.J., *Lipid rafts: heterogeneity on the high seas.* Biochem J, 2004. **378**(Pt 2): p. 281-92.
244. Bacia, K., et al., *Fluorescence correlation spectroscopy relates rafts in model and native membranes.* Biophys J, 2004. **87**(2): p. 1034-43.
245. Mielenz, D., et al., *Lipid rafts associate with intracellular B cell receptors and exhibit a B cell stage-specific protein composition.* J Immunol, 2005. **174**(6): p. 3508-17.
246. Bagshaw, R.D., D.J. Mahuran, and J.W. Callahan, *A proteomic analysis of lysosomal integral membrane proteins reveals the diverse composition of the organelle.* Mol Cell Proteomics, 2005. **4**(2): p. 133-43.
247. Sobo, K., et al., *Diversity of raft-like domains in late endosomes.* PLoS ONE, 2007. **2**(4): p. e391.
248. Ikonen, E., *Roles of lipid rafts in membrane transport.* Curr Opin Cell Biol, 2001. **13**(4): p. 470-7.
249. Sengupta, P., B. Baird, and D. Holowka, *Lipid rafts, fluid/fluid phase separation, and their relevance to plasma membrane structure and function.* Semin Cell Dev Biol, 2007. **18**(5): p. 583-90.
250. Brown, D.A., *Lipid rafts, detergent-resistant membranes, and raft targeting signals.* Physiology (Bethesda), 2006. **21**: p. 430-9.
251. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.* Anal Biochem, 1987. **162**(1): p. 156-9.
252. Strauss, W.M., *Preparation of Genomic DNA from Mammalian Tissue*, in *Current Protocols in Molecular Biology*, K. Struhl, Editor. 2000, John Wiley & Sons, Inc.: New York. p. 2.2.

253. Kramer, M.F.a.D., M.C., *The Polymerase Chain Reaction*, in *Current Protocols in Molecular Biology*, F.M. Ausubel, Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., Editor. 1999, John Wiley & Sons, Inc.: New York. p. 15.
254. *Center for Biological Sequence Analysis* [cited; Available from: http://www.cbs.dtu.dk/.
255. Hartree, E.F., *Determination of protein: a modification of the Lowry method that gives a linear photometric response.* Anal Biochem, 1972. **48**(2): p. 422-7.
256. Zimmermann, T., J. Rietdorf, and R. Pepperkok, *Spectral imaging and its applications in live cell microscopy.* FEBS Lett, 2003. **546**(1): p. 87-92.
257. Simmer, F., et al., *Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi.* Curr Biol, 2002. **12**(15): p. 1317-9.
258. *C. elegans Genetics Center* [cited; Available from: http://www.cbs.umn.edu/CGC/.
259. *C.elegans Gene Knockout Consortium* [cited; Available from: http://celeganskoconsortium.omrf.org/.
260. *WormBase - Blast/Blat Search*. [cited; Available from: http://www.wormbase.org/db/searches/blast_blat.
261. *GenBank*. [cited; Available from: http://www.ncbi.nlm.nih.gov/Genbank/.
262. Wenger DA, W.C., *Screening for Lysosomal Disorders.*, in *Techniques in Diagnostic Human Biochemical Genetics: A Laboratory Manual.*, H. FA, Editor. 1991, Wiley-Liss, Inc: New York. p. 587-617.
263. Mello, C.C., et al., *Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences.* Embo J, 1991. **10**(12): p. 3959-70.
264. Poole, B. and S. Ohkuma, *Effect of weak bases on the intralysosomal pH in mouse peritoneal macrophages.* J Cell Biol, 1981. **90**(3): p. 665-9.
265. Nonet, M.L., et al., *Synaptic function is impaired but not eliminated in C. elegans mutants lacking synaptotagmin.* Cell, 1993. **73**(7): p. 1291-305.
266. Gallagher, S., Winston, S.E., Fuller, S.A., Hurrell, J.G.R., *Analysis of Proteins.*, in *Current Protocols in Molecular Biology*, F.M. Ausubel, Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., Editor. 1997, John Wiley & Sons, Inc.: New York. p. 10.8.
267. Hodgkin, J., *Conventional genetics*, in *C.elegans: A Practical Approach*, I.A. Hope, Editor. 1999, Oxford University Press: Oxford. p. 245-269.
268. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
269. Felsenstein, J., *PHYLIP - Phylogeny Inference Package (Version 3.2).* Cladictics, 1988. **5**: p. 164-166.
270. *Swiss-Model Server*. [cited; Available from: http://swissmodel.expasy.org/.
271. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-23.
272. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.
273. Laskowski, R.A., et al., *Procheck - a Program to Check the Stereochemical Quality of Protein Structures* Journal od Applied Crystallography, 1993. **26**: p. 283-291.
274. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W375-83.
275. Holm, L. and J. Park, *DaliLite workbench for protein structure comparison.* Bioinformatics, 2000. **16**(6): p. 566-7.
276. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
277. *Amaxa Server*. [cited; Available from: http://www.amaxa.com/.
278. *WormBase - Freeze Version WS100*. [cited; Available from: http://ws100.wormbase.org/.
279. Suzuki, K., *Enzymatic diagnosis of sphingolipidoses.* Methods Enzymol, 1987. **138**: p. 727-62.
280. *WormBase - Freeze Version WS110*. [cited; Available from: http://ws110.wormbase.org/.
281. *WormBase - Freeze Version WS180* [cited; Available from: http://ws180.wormbase.org/.
282. Conzelmann, E. and K. Sandhoff, *Glycolipid and glycoprotein degradation.* Adv Enzymol Relat Areas Mol Biol, 1987. **60**: p. 89-216.
283. Conzelmann, E. and K. Sandhoff, *Partial enzyme deficiencies: residual activities and the development of neurological disorders.* Dev Neurosci, 1983. **6**(1): p. 58-71.
284. Kneen, M., et al., *Green fluorescent protein as a noninvasive intracellular pH indicator.* Biophys J, 1998. **74**(3): p. 1591-9.
285. Stinchcomb, D.T., et al., *Extrachromosomal DNA transformation of Caenorhabditis elegans.* Mol Cell Biol, 1985. **5**(12): p. 3484-96.

286. Clokey, G.V. and L.A. Jacobson, *The autofluorescent "lipofuscin granules" in the intestinal cells of Caenorhabditis elegans are secondary lysosomes.* Mech Ageing Dev, 1986. **35**(1): p. 79-94.
287. Hwang, H.Y. and H.R. Horvitz, *The SQV-1 UDP-glucuronic acid decarboxylase and the SQV-7 nucleotide-sugar transporter may act in the Golgi apparatus to affect Caenorhabditis elegans vulval morphogenesis and embryonic development.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14218-23.
288. Naumov, D.G., *[Structure and evolution of mammalian maltase-glucoamylase and sucrase-isomaltase genes]*. Mol Biol (Mosk), 2007. **41**(6): p. 1056-68.
289. Majer, F., et al., *Structure-based specificity mapping of secreted aspartic proteases of Candida parapsilosis, Candida albicans, and Candida tropicalis using peptidomimetic inhibitors and homology modeling.* Biol Chem, 2006. **387**(9): p. 1247-54.
290. Nasi, R., et al., *New chain-extended analogues of salacinol and blintol and their glycosidase inhibitory activities. Mapping the active-site requirements of human maltase glucoamylase.* J Org Chem, 2007. **72**(1): p. 180-6.

9. List of publications and selected presentations

Impact Factor (IF) values and citation report correspond to ISI Web of Knowledge (as of August 2008).
Number of citations: **15**; h-index: **2**.

9.1 Publications related to the thesis

Characterization of gana-1, a Caenorhabditis elegans gene encoding a single ortholog of vertebrate alpha-galactosidase and alpha-N-acetylgalactosaminidase.
Hujová J.*,Sikora J.*,Dobrovolný R.*,Poupětová H.,Ledvinová J.,Kostrouchová M.,Hřebíček M.
BMC Cell Biol 2005 Jan 27; 6(1):5. **(IF2007 = 3.092)**
*equal contributors
cited by:
de Voer G, et al. *Biophysic Acta – Molec Basis of Dis.* 2008; 1782(7-8): 433-446
Kaletta T, et al. *Nat Rev Drug Disc.* 2006; 5(5): 387-398

Mutations in TMEM76 Cause Mucopolysaccharidosis IIIC (Sanfilippo C Syndrome).
Hrebicek M., Mrazova L., Seyrantepe V., Durand S., Roslin N.M., Noskova L., Hartmannova H., Ivanek R., Cizkova A., Poupetova H., Sikora J., **Urinovska J.,** Stranecky V., Zeman J., Lepage P., Roquis D.,Verner A., Ausseil J., Beesley C.E., Maire I., Poorthuis B.J., van de Kamp J., van Diggelen O.P., Wevers R.A., Hudson T.J., Fujiwara T.M., Majewski J., Morgan K., Kmoch S., Pshezhetsky A.V.
Am J Hum Genet. 2006 Nov;79(5):807-819. Epub 2006 Sep 8. **(IF2007 = 11.092)**
cited by:
Countinho MJ, et al. *Clin Genet.* 2008; 74(2): 194-195
Valstar MJ, et al. *J Inherit Metab Dis.* 2008; 31(2): 240-252
Cizkova A, et al. *BMC Genomics.* 2008; 9: 38
Ruijter GJG, et al. *Mol Genet and Metab.* 2008; 93(2): 104-111
Schroder B, et al. *Traffic.* 2007; 8(12): 1676-1686
Sleat DE, et al. *Proteom Clin Aplic.* 2007; 1(9): 1134-1146

Members of glycoside hydrolases family 31 in *Caenorhabditis elegans*: Search for the nematode's acid α-glucosidase.
Uřinovská J.*, Sikora J.*, Poupětová H., Majer F., Hlavatá J., Jelínek J., Kostrouchová M., Hřebíček M.
Submitted manuscript to BMC Cell Biology
*equal contributors

9.2 Other publications

Eight novel ABCD1 gene mutations and three polymorphisms in patients with X-linked adrenoleukodystrophy: The first polymorphism causing an amino acid exchange.
Dvorakova L.,Storkanova G.,Unterrainer G.,**Hujova J.,**Kmoch S.,Zeman J.,Hrebicek M.,Berger J.
Hum Mutat 2001; 18(1):52-60. **(IF2007 = 6.273)**
cited by:
Berger J, et al. *Biochim et Biophysic Acta – Mol Cell Res.* 2006; 1763 (12): 1721-1723
Pan H, et al. *Pediatric Neurology.* 2005; 33(2): 114-120
Coll MJ, et al. *Clin Genet.* 2005; 67(5): 418-424
Paintlia AS, et al. *Neurobiol of Dis.* 2003; 14(3): 425-439
Guimaraes CP, et al. *Mol Genet and Metab.* 2002; 76(1): 62-67
Lida A, et al. *J Hum Genet.* 2002; 47(6): 285-310
Kemp S, et al. *Hum Mutat.* 2001; 18(6): 499-515

9.3 Published abstracts

XVII. biochemický sjezd, Prague, Czech republic, September 2000
Dvořáková L., Štorkánová G., Hřebíček M., Zeman J., **Hujová J.**, Unterrainer G., Berger J.
Mutation analysis and identification of carriers in X-linked adrenoleukodystrophy (X-ALD) families.The first polymorphism causing an amino acid exchange within the ALDP-gene.
Chem Listy. 94, 511 (2000)

VIII. International Congress of Inborn Errors of Metabolism Cambridge, UK, September 2000
Dvořáková L., Štorkánová G., Hřebíček M., Zeman J., **Hujová J.**, Unterrainer G. and Berger J.
The first polymorphism causing an amino acid exchange within the ALDP-gene. Eight novel mutations and three polymorphisms in patients with X-linked adrenoleukodystrophy.
J Inherit Metab Dis. 2000, 23 Suppl.1, 254

European Human Genetics Conference, May 2002
Dvořáková L., **Hujova J.**, Dobrovolny R., Stolnaja L., Tietzeova E, Hrebicek M., Kumsta M, Sebesta I.
Highly skewed X-inactivation pattern in a female with unique presentation of hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency.
Eur J Hum Genet. 2002, 10, Suppl.1., 239

European Human Genetics Conference, Birminhgham, UK, May 2003
Sikora J., **Hujova J.**, Dobrovolny R., Holanova D., Asfaw B., Poupetova H., Ledvinova J., Kostrouchova M., Hrebicek M.
Caenorhabditis elegans as a model for lysosomal glycosidase deficiencies.
Eur J Hum Genet. 2003, 11, Suppl.1.

53rd Annual Meeting of The American Society of Human Genetics, Los Angeles, USA, November 2003
Hrebicek M., Sikora J., **Hujova J.**, Holanova D., Poupetova H., Asfaw B., Ledvinova J., Kostrouchova M.
Evaluation of *Caenorhabditis elegans* model for lysosomal diseases caused by deficits of glycosidases.
Am J Hum Genet. 2003, 73: 348-348

9.4 Other selected presentations

18. pracovní dny Dědičné Metabolické Poruchy, Slušovice, Czech republic, May 2003
Sikora J., **Hujova J.**, Dobrovolny R., Holanova D., Asfaw B., Poupetova H., Ledvinova J., Kostrouchova M., Hrebicek M.
Caenorhabditis elegans as a model for lysosomal glycosidase deficiencies.
First price for the scientists under 30 years of age.

14th ESGLD Workshop, Poděbrady, Czech Republic, September 2003
Hujova J., Dobrovolny R., Sikora J., Holanova D., Asfaw B., Poupetova H., Ledvinova J., Kostrouchova M., Hrebicek M.
Caenorhabditis elegans as a model for Fabry and Schindler diseases.

14th ESGLD Workshop, Poděbrady, Czech Republic, September 2003
Sikora J., **Hujova J.**, Dobrovolny R., Holanova D., Asfaw B., Poupetova H., Ledvinova J., Kostrouchova M., Hrebicek M.
Caenorhabditis elegans as a model for acid α-glucosidase deficiency.

European Worm Meeting , Casino Kursaal, Interlaken, Switzerland, May 2004
Hujova J., Dobrovolny R., Sikora J., Holanova D., Asfaw B., Poupetova H., Ledvinova J., Kostrouchova M., Hrebicek M.
Caenorhabditis elegans has only one α-galactosidase/ α -N-acetylgalaktosaminidase ortholog

5th students´ scientific conference 1. LF UK, Prague, Czech Republic, May 2004
Sikora J., **Hujova J.**, Dobrovolny R.
Caenorhabditis elegans has only one α-galactosidase and α -N-acetylgalaktosaminidase ortholog
First price.

15th ESGLD Workshop, Oslo, Norway, September 2005
Uřinovská J., Sikora J., Dobrovolný R., Poupetova H., Ledvinova J., Kostrouchova M., Hřebiček M.
Characterization of *gana-1*, a *Ceanoharbditis elegans* gene encoding a single ortholog of vertebrate α-galactosidase and α-N-acetylgalactosaminidase

16th ESGLD Workshop, Perugia, Italy, September 2007
Uřinovská J., Sikora J., Poupetova H., Hlavata J., Kostrouchova M., Hřebiček M.
Glycosyl hydrolases family 31 members in *Caenorhabditis elegans*: Search for worm´s lysosomal acid α-glucosidase.

47th ASCB Annual Meeting, Washington, DC, December 2007
Uřinovská J., Svobodová E., Cmarko D., Hřebíček M., Sikora J.
Microscopic evidence for HGSNAT-containing microdomains in lysosomal membrane.

10. Appendix

10.1. Tables of PCR primers and tables of single and double immunolabeling.

gene	forward primer (5´ → 3´)	reverse primer (5´ → 3´)	restriction sites (tag)	purpose of PCR
gana-1	GGTTTAATTACCCAAGTTTGAG	ATCCTGATTAATTTTAATTGC		**verification of sequence**
gana-1	GGTTTTAACCCAGTTACTCAAG	ATCCTGATTAATTTTAATTGC		verification of sequence
gana-1	GGTTTAATTACCCAAGTTTGAG	TCAATTTGCTTGAGGTACATA		verification of sequence
gana-1	CTTAAGTTTGGAATTTATGAA	AATACGACTCACTATAG		verification of sequence
gana-1	GGTTTAATTACCCAAGTTTGAG	TCAATTTGCTTGAGGTACATA		**RNA interference**
gana-1	GTGAGAGTGGGGAGATAGAA	TCAATTTGCTTGAGGTACATA		**transgenic GFP epression** (external primers)
gana-1	ACATGCATGCAACTTTCACAGGAACAC AAC	CGACGTCGACAATTGAACTCTATTGGTT CTCAA	SphI/SalI	transgenic GFP epression (internal primers)
gana-1	GGTTTAATTACCCAAGTTTGAG	TCAATTTGCTTGAGGTACATA		**mammalian expression** (external primers)
gana-1	GAGCTCCACCATGAGATTGCTCCTT	GCCCGGGCAAAATTGAACTCTATTG	SacI/SrfI (FLAG C terminal)	mammalian expression (internal primers)
gana-1	AAGATCTTGAGATTGCTCCTTCC	GTCGACTAAAATTGAACTCTATTGG	BglII/SalI (FLAG N terminal)	mammalian expression (internal primers)
gana-1	AAGATCTTGAGATTGCTCCTTCC	GACGTCGACAAATTGAACTCTATTGG	BglII/SalI (FLAG N/c-myc C terminal)	mammalian expression (internal primers)
gana-1	GAGCTCCACCATGAGATTGCTCCTT	GGATCCATGAGAATCTCTTTCCATT	SacI/BamHI (FLAG internal I)	mammalian expression (internal primers,1st part)
gana-1	AAGATCTTGGTTGCAAATAAAACA	GTCGACTAAAATTGAACTCTATTGG	BglII/SalI (FLAG internal I)	mammalian expression (internal primers,2nd part)
gana-1	GAGCTCCACCATGAGATTGCTCCTT	GGATCCGTGTTTGTCTTGCATTT	SacI/BamHI (FLAG internal II)	mammalian expression (internal primers,1st part)
gana-1	AAGATCTCAACTCATGGTCCAGGAAA	GTCGACTAAAATTGAACTCTATTGG	BglII/SalI (FLAG internal II)	mammalian expression (internal primers,2nd part)
gana-1	GAGCTCCACCATGAGATTGCTCCTT	GCCCGGGCCGGCAAGTTGAATTTCAAT	SacI/SrfI (FLAG internal III)	mammalian expression (internal primers,1st part)
gana-1	AAGATCTCAATTGGTCTCACTGATCCT	GTCGACTAAAATTGAACTCTATTGG	BglII/SalI (FLAG internal III)	mammalian expression (internal primers,2nd part)

Table 8: Primers for amplification of *gana-1* PCR products.

gene	forward primer (5´ → 3´)	reverse primer (5´ → 3´)	restriction sites	purpose of PCR
D2096.12	GGTTTAATTACCCAAGTTTGAG	GTACGTGTTTTGGAGGATTTTT		**verification of sequence**
D2096.12	GGTTTTAACCCAGTTACTCAAG	GTACGTGTTTTGGAGGATTTTT		verification of sequence
D2096.12	TCCCCCGGGATCAACAAAACGTTTCAAGT	ATTCTAATCGTCGGAAGTTC		verification of sequence
D2096.12	AAAAGAACGGAAAAAGAGAAC	TGGATACCACCAATCTTCAT		verification of sequence
D2096.12	AGTGACGGTGGCAGAAAT	AATACGACTCACTATAG		verification of sequence
aagr-1	GGTTTAATTACCCAAGTTTGAG	GAGTTGCTCCATTTGATGAAT		verification of sequence
aagr-1	GGTTTTAACCCAGTTACTCAAG	GAGTTGCTCCATTTGATGAAT		verification of sequence
aagr-1	TGCCCGTGACATTGGA	TTCGTTGATATACCGCTTGTG		verification of sequence
aagr-1	AATCACTGGATCAGACTCATC	GAAATACGCATGAACACTTG		verification of sequence
aagr-1	AATCACTGGATCAGACTCATC	AATACGACTCACTATAG		verification of sequence
D2096.12	TGCCCGTGACATTGGA	TTCGTTGATATACCGCTTGTG		**RNA interference**
aagr-1	AATCACTGGATCAGACTCATC	GAAATACGCATGAACACTTG		**RNA interference**
CEOP4284 operon	ATCTTGAGCCAATGTGAACAA	AGCATTTCTTCGGACAATAAA		**transgenic GFP epression** (external primers)
CEOP4284 operon	AAAACTGCAGCCCATCACCGTCATCTTGTA	CGCGGATCCTGTACGTGTTTTGGAGGATTTTT	PstI/BamHI (pJS6)	transgenic GFP expression (internal primers)
CEOP4284 operon	AAAACTGCAGGTAGAACTAACCCCAGATAAC	CGCGGATCCTGTACGTGTTTTGGAGGATTTTT	PstI/BamHI (pJS4)	transgenic GFP expression (internal primers)
CEOP4284 operon	AAAACTGCAGGACGATGAGAACGCAACAAG	CGCGGATCCTGTACGTGTTTTGGAGGATTTTT	PstI/BamHI (pJS1)	transgenic GFP expression (internal primers)
aagr-1	CCCGGGATCAACAAAACGTTTCAAGT	CTATTACCCAACAGAATCCGCCCGGG	SmaI/SmaI (pJS8)	**transgenic GFP expression**
aagr-1	GATTTTTGCTCGTAGTCCCG	CGGTGCAATCATAAGAGCAG		**deletion mutant RB 1790**
aagr-2	GGTTTAATTACCCAAGTTTGAG	ATGGATAAGTGGCTTCAATG		**verification of sequence**
aagr-2	GGTTTTAACCCAGTTACTCAAG	ATGGATAAGTGGCTTCAATG		verification of sequence
aagr-2	GATACAAGGATTTCACTACTG	GATGTTGTTCCTTGATAAAT		verification of sequence
aagr-2	TTCAATCTTGGATATGAAT	AATACGACTCACTATAG		verification of sequence
aagr-2	GGTTTAATTACCCAAGTTTGAG	ATGGATAAGTGGCTTCAATG		**RNA interference**
aagr-2	GCTCTGCCAAGAGATATTATT	AATTCCCTGAACCTTTAGTTA		**transgenic GFP epression** (external primers)
aagr-2	CGCGGATCCTAAACGGTTTTCAAACATCTT	CGCGGATCCACACCGTCTTTGAAGTAACAC	BamHI/BamHI (pJS9)	transgenic GFP expression (internal primers)

Table 9: Primers for amplification of *operon* CEOP4284, *aagr-1* and *aagr-2* PCR products.

gene	forward primer (5´ → 3´)	reverse primer (5´ → 3´)	restriction sites	purpose of PCR
aagr-3	GGTTTAATTACCCAAGTTTGAG	ATTATAATTCCAGCGACATTG		**verification of sequence**
aagr-3	GGTTTTAACCCAGTTACTCAAG	ATTATAATTCCAGCGACATTG		verification of sequence
aagr-3	AAACATGGGTTGACACTCAGT	AGAGCATAACGGGTTCTTAAA		verification of sequence
aagr-3	GTGCTCATGCTCATATTGA	AATACGACTCACTATAG		verification of sequence
aagr-3	GTGCTCATGCTCATATTGA	AATACGACTCACTATAG		**RNA interference**
aagr-4	GGTTTAATTACCCAAGTTTGAG	GCAGAATCCCGACTGTTC		**verification of sequence**
aagr-4	GGTTTTAACCCAGTTACTCAAG	GCAGAATCCCGACTGTTC		verification of sequence
aagr-4	GGTTTAATTACCCAAGTTTGAG	CTCAAACTCTCCGTGATGTT		verification of sequence
aagr-4	GGAAAGCCAGAAGATATGAT	TCCATCATCCAAATAAATTG		verification of sequence
aagr-4	GAGCTATTCCTTGAGAGACA	AATACGACTCACTATAG		verification of sequence
aagr-4	CTGATTGGAGCACTTGATG	GGTGAACTCCAACAAGATATG		**RNA interference**
aagr-4	CCGGCAAGTAGATTGAGAGC	CAATATCACGCAGGTTGTGG		**deletion mutant RB 1307**
F53F4.8	GGTTTAATTACCCAAGTTTGAG	CTCTGAAGAGGGAAGTGTTG		**pseudogene**
F53F4.8	GGTTTTAACCCAGTTACTCAAG	CTCTGAAGAGGGAAGTGTTG		pseudogene
F53F4.8	TATATGTATGGAAGTAGAA	AATACGACTCACTATAG		pseudogene

Table 10: Primers for amplification of *aagr-3*, *aagr -4* and F53F4.8 (pseudogene) PCR products.

gene	forward primer (5´ → 3´)	reverse primer (5´ → 3´)	additive	purpose of PCR
HGSNAT	GTGACTCAGGCGGCGGTGAC	TGACAGCAAGACCAATAAGGAATG	8%DMSO (GC rich)	**verification of cDNA sequence**
HGSNAT	ATGACGGGCGCGCGGGCGTC	TGACAGCAAGACCAATAAGGAATG	8%DMSO (GC rich)	verification of cDNA sequence
HGSNAT	CAGAGCTGAAGATGGATCAGGCT	TGACAGCAAGACCAATAAGGAATG		verification of cDNA sequence
HGSNAT	CCTGTGACCTGGCTGTGAACGA	GCAATCTTCCCCAGCAATCTGA		verification of cDNA sequence
HGSNAT	TCGTGTTCCCGTGGTTTGTATT	ATGCCCCCAGGACCAAGATAAC		verification of cDNA sequence
HGSNAT	GCTGGGCTTGACATTCCTCCTG	AGGTGCTCCTTGTGGGACTGGT		verification of cDNA sequence
HGSNAT	TGTTTGAGAACTACTTCCCCTT	GGCTTTCTGGAAACATTAAACT		verification of cDNA sequence
HGSNAT	CCTACTGCGGCTGCTATGAAGC	TCCCCGATTTCACACCTAAGAA		verification of cDNA sequence
HGSNAT	CAATGGAGGCAGCGTTCCTACT	GGTGGCAATGGACAGGTGACTT		verification of cDNA sequence
HGSNAT	TGAGTCTGGACCCTTGTTATCA	TTTGGCACCATGAGGAATACAG		verification of cDNA sequence
HGSNAT	CAGAGCTGAAGATGGATCAGGCT	ATGCCCCCAGGACCAAGATAAC		**Human Multiple Tissue cDNA Panel 1 (Clontech)**

Table 11: Primers for PCR amplification of HGSNAT cDNA.

antigen	Ab type	vendor	fixation	duration (′)	permeabilization	duration (′)	dilution	incubation (′)	temperature (°C)
2785-6	rabbit (Ra)		PFA 4%	10	no	no	1_200	overnight	4
2785-6	Ra		PFA 4%	10	Triton X 100, 0.1%	10	1_200	overnight	4
2785-6	Ra		MetOH	10	no	no	1_200	overnight	4
Complex IV (COX) subunit I	mouse (Ma) (1D6E1A8)	Mitosciences	MetOH	10	no	no	1_200	overnight	4
Protein Disulfide Isomerase (PDI)	Ma (1D3)	Stressgen	MetOH	10	no	no	1_800	60	37
peripheral Golgi membrane (Golgi 58 K)	Ma (58K-9)	Sigma-Aldrich	MetOH	10	no	no	1_800	overnight	4
Early Endosomal Antigen 1 (EEA1)	Ma	Abcam	MetOH	10	no	no	1_200	overnight	4
Mannose- 6 - phosphate receptor (M6PR)	Ma (2G11)	Abcam	MetOH	10	no	no	1_250	overnight	4
Cathepsin D	Ma (MCA2068)	Serotec	MetOH	10	no	no	1_50	overnight	4
LAMP2	Ma (H4B4)	Developmental Studies Hybridoma Bank	MetOH	10	no	no	1_500	overnight	4
Flotillin - 1	Ma	BD Biosciences	MetOH	10	no	no	1_800	overnight	4
choleratoxin B - Alexa Fluor 555		Molecular Probes	PFA 4%	10	no	no	1_800	60	37
GB3	Ma	Seikagaku	PFA 4%	10	Triton X 100, 0.1%	10	1_150	overnight	4
Filipin		Sigma-Aldrich	PFA 4%	10	no	no	500μg/μl	60	37
sortilin	goat	R&D Systems	MetOH	10	no	no	1_50	overnight	4
Heparan sulfate	Ma	Seikagaku	PFA 4%	10	Triton X 100, 0.1%	10	1_2000	60	37

Table 12: Table of single immunolabeling.

1st antigen	source animal	dilution	2nd antigen	source animal	dilution	fixation/duration/temperature/ permeabilization/duration/anti body incubation/temperature	1st secondary antibody	2nd secondary antibody
2785-6	Ra	1_200	**COXIV/I**	Ma	1_200	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**PDI**	Ma	1_800	MetOH/10/overnight/60/4/37	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**Golgi 58K**	Ma	1_800	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**EEA1**	Ma	1_200	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**M6PR**	Ma	1_250	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**Cathepsin D**	Ma	1_25	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**LAMP2**	Ma	1_500	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**Flotillin**	Ma	1_800	MetOH/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**choleratoxin B Alexa Fluor555**		1_800	PFA/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	
2785-6	Ra	1_200	**Gb3**	Ma	1_150	PFA/10/Triton/10/overnight/4	goat anti-rabbit IgG Alexa Fluor 488	donkey anti-mouse IgG Alexa Fluor 555
2785-6	Ra	1_200	**Filipin**		500μg/μl	PFA/10/Triton/10/overnight/4/60	goat anti-rabbit IgG Alexa Fluor 488	
2785-6	Ra	1_200	**sortilin**	goat	1_50	MetOH/10/overnight/4	donkey anti-rabbit IgG Alexa Fluor 488	donkey anti-goat IgG Alexa Fluor 568 labeled
2785-6	Ra	1_200	**Heparan sulfate**	Ma	1_2000	PFA/10/Triton/10/60˚/37	goat anti-rabbit IgG Alexa Fluor 488	goat anti-mouse IgM Alexa Fluor 555

Table 13: Table of double immunolabeling.

10.2. Appended publications related to the thesis.

Appended publication 1:

Characterization of gana-1, a Caenorhabditis elegans gene encoding a single ortholog of vertebrate alpha-galactosidase and alpha-N-acetylgalactosaminidase.
Hujová J*,Sikora J*,Dobrovolný R*,Poupětová H,Ledvinová J,Kostrouchová M,Hřebíček M. *BMC Cell Biol* 2005 Jan 27; 6(1):5.
*equal contributors

Appended publication 2:

Mutations in TMEM76 Cause Mucopolysaccharidosis IIIC (Sanfilippo C Syndrome).
Hrebicek M., Mrazova L., Seyrantepe V., Durand S., Roslin N.M., Noskova L., Hartmannova H., Ivanek R., Cizkova A., Poupetova H., Sikora J., **Urinovska J.,** Stranecky V., Zeman J., Lepage P., Roquis D.,Verner A., Ausseil J., Beesley C.E., Maire I.,Poorthuis B.J.,van de Kamp J.,van Diggelen OP., Wevers RA., Hudson TJ., Fujiwara TM., Majewski J.,Morgan K., Kmoch S., Pshezhetsky AV.
Am J Hum Genet. 2006 Nov;79(5):807-819. Epub 2006 Sep 8.

Appended submitted manuscript:

Members of glycoside hydrolases family 31 in *Caenorhabditis elegans*: Search for the nematode's acid α-glucosidase.
Uřinovská J.*, Sikora J.*, Poupětová H., Majer F., Hlavatá J., Jelínek J., Kostrouchová M., Hřebíček M.
Submitted manuscript to BMC Cell Biology
*equal contributors

BMC Cell Biology

BioMed Central

Research article

Open Access

Characterization of *gana-1*, a Caenorhabditis elegans gene encoding a single ortholog of vertebrate α-galactosidase and α-N-acetylgalactosaminidase

Jana Hujová†, Jakub Sikora†, Robert Dobrovolný†, Helena Poupětová, Jana Ledvinová, Marta Kostrouchová and Martin Hřebíček*

Address: Institute of Inherited Metabolic Disorders, Charles University, 1st Medical Faculty, Prague, Czech Republic

Email: Jana Hujová - jhujo@lf1.cuni.cz; Jakub Sikora - jakub.sikora@lf1.cuni.cz; Robert Dobrovolný - rdobr@lf1.cuni.cz; Helena Poupětová - helena.poupetova@lf1.cuni.cz; Jana Ledvinová - jledvin@beba.cesnet.cz; Marta Kostrouchová - marta.kostrouchova@lf1.cuni.cz; Martin Hřebíček* - mhreb@lf1.cuni.cz

* Corresponding author †Equal contributors

Published: 27 January 2005

BMC Cell Biology 2005, **6**:5 doi:10.1186/1471-2121-6-5

Received: 13 September 2004
Accepted: 27 January 2005

This article is available from: http://www.biomedcentral.com/1471-2121/6/5

Abstract

Background: Human α-galactosidase A (α-GAL) and α-N-acetylgalactosaminidase (α-NAGA) are presumed to share a common ancestor. Deficiencies of these enzymes cause two well-characterized human lysosomal storage disorders (LSD) – Fabry (α-GAL deficiency) and Schindler (α-NAGA deficiency) diseases. *Caenorhabditis elegans* was previously shown to be a relevant model organism for several late endosomal/lysosomal membrane proteins associated with LSDs. The aim of this study was to identify and characterize *C. elegans* orthologs to both human lysosomal luminal proteins α-GAL and α-NAGA.

Results: BlastP searches for orthologs of human α-GAL and α-NAGA revealed a single *C. elegans* gene (R07B7.11) with homology to both human genes (α-galactosidase and α-N-acetylgalactosaminidase) – *gana-1*. We cloned and sequenced the complete *gana-1* cDNA and elucidated the gene organization.

Phylogenetic analyses and homology modeling of GANA-1 based on the 3D structure of chicken α-NAGA, rice α-GAL and human α-GAL suggest a close evolutionary relationship of GANA-1 to both human α-GAL and α-NAGA.

Both α-GAL and α-NAGA enzymatic activities were detected in *C. elegans* mixed culture homogenates. However, α-GAL activity on an artificial substrate was completely inhibited by the α-NAGA inhibitor, N-acetyl-D-galactosamine.

A GANA-1::GFP fusion protein expressed from a transgene, containing the complete *gana-1* coding region and 3 kb of its hypothetical promoter, was not detectable under the standard laboratory conditions. The GFP signal was observed solely in a vesicular compartment of coelomocytes of the animals treated with Concanamycin A (CON A) or NH_4Cl, agents that increase the pH of the cellular acidic compartment.

Immunofluorescence detection of the fusion protein using polyclonal anti-GFP antibody showed a broader and coarsely granular cytoplasmic expression pattern in body wall muscle cells, intestinal cells, and a vesicular compartment of coelomocytes.

Inhibition of *gana-1* by RNA interference resulted in a decrease of both α-GAL and α-NAGA activities measured in mixed stage culture homogenates but did not cause any obvious phenotype.

Conclusions: GANA-1 is a single *C. elegans* ortholog of both human α-GAL and α-NAGA proteins. Phylogenetic, homology modeling, biochemical and GFP expression analyses support the hypothesis that GANA-1 has dual enzymatic activity and is localized in an acidic cellular compartment.

Background

Humans have two enzymes with α-galactosidase activity and an acidic pH optimum, α-N-acetylgalactosaminidase (α-NAGA) (previously called α-galactosidase B) and α-galactosidase A (α-GAL). Hereditary deficiency of each of the hydrolases causes a distinct lysosomal storage disorder in humans, Schindler and Fabry diseases, respectively [1,2].

Early studies suggested that both human enzymes were glycoforms with similar substrate specifities. Purified enzymes had similar physical properties, including subunit molecular mass (~46 kDa), homodimeric structure, and amino acid sequences. However, additional studies showed kinetic, structural, and immunologic differences proving that α-GAL and α-NAGA were products of two different genes [3,4]. The two genes differed in the number of exons (7 and 9, respectively) and also in the number, placement, and orientation of Alu repeats. Exons 2 – 7 of the α-NAGA gene showed high similarity to the first six exons of α-GAL gene. Because of the remarkable amino acid identity (49%) and similarity (63%) between the two genes and the similar intron placement, Wang [5] and co-workers suggested that a duplication event occurred during the evolution of both enzymes.

Both enzymes belong to the glycoside hydrolase family 27 clan D [6]. Glycoside hydrolase family 27 clan D orthologs have been identified in a broad spectrum of prokaryotes and eukaryotes, including plants. Members of the family have a highly similar active site and share the same reaction mechanism. The structures of chicken α-NAGA, human α-GAL and rice α-GAL have been determined by X-ray crystallography [7-9]. Chicken and human enzymes have a homodimeric quarternary structure whereas rice α-GAL acts as a monomer. The monomer units are composed of two distinct domains. Domain I contains the active site and adopts a $(\beta/\alpha)_8$ barrel structure, a domain fold observed commonly in glycosidases. Domain II has eight antiparallel β strands, packed into two β sheets in a β sandwich fold containing a Greek key motif [8].

The physiological importance of both enzymes is evidenced by the severe presentation of α-NAGA and α-GAL deficiencies in humans [1,2].

Our recent study on degradation of blood group A and B glycolipids in Fabry cells indicated a high residual activity in Fabry cells toward natural substrate glycolipid B-6-2 [10] although α-galactosidase activity was completely absent when measured in vitro by routine procedures using artificial substrates. We proposed that another enzyme, different from α-GAL, contributes in vivo to hydrolysis of α-galactosides. We suggested α-NAGA as the most likely candidate. Human α-NAGA is known to accept α-galactosides albeit with a high K_m [11,12]. Its activity must be inhibited when measuring α-GAL in tissues with high α-NAGA activity [13].

We investigated the phylogenesis of a single *C. elegans* α-GAL and α-NAGA ortholog (*gana-1*) to both human genes. We present evidence suggesting that this gene has indeed evolved from the α-GAL/α-NAGA ancestral gene before the duplication event which resulted in separate α-NAGA and α-GAL genes in higher metazoans. We further performed structural analysis of the GANA-1 3D model acquired by homology modeling. We determined the spatial and temporal expression of the gene in transgenic worms using a translational reporter and examined the effect of RNA interference (RNAi) as a first step in the possible use of *C. elegans* as a model organism for Schindler and Fabry diseases.

Results and discussion

cDNA amplification and sequencing

The complete *C. elegans* genome [14,15] contains only one open reading frame, designated *gana-1* (R07B7.11) that has sequence similar to human genes encoding α-GAL and α-NAGA. Similar results were obtained from searching the available *C. briggsae* genome sequence [15]. The *gana-1* gene consists of 5 exons and 4 introns and is annotated as an ortholog of human α-NAGA. Several EST clones for this open reading frame (ORF) have been reported and open-reading-frame sequence tag (OST) is present in the Worfdb database [16]. Available public database data are in agreement with our findings.

We verified the gene structure by sequencing the PCR products from genomic DNA and cDNA (Figure 1). The analyzed region spanned the entire coding region and the 3' and 5' untranslated regions (UTR). The 5' UTR SL1 element suggests that the gene is either the only gene transcribed from the promoter or is the first gene in an operon including *gana-1* and the two predicted downstream genes R07B7.12 and R07B7.13. Although this region has not been reported as an operon, the physical distances between this cluster of three genes are suggestive of an operon [17-19]. No alternative splicing was found using RT/PCR, a feature similar to both human and mouse orthologs. RNA editing was reported in the 3' UTR of human α-GAL, a finding that another group was unable to confirm [20]. We noted no signs of RNA editing in clones derived from the *gana-1* cDNA.

Phylogenetic studies

We aligned GANA-1 protein sequence with other melibiase family members (Figure 2) and constructed a phylogenetic tree (Figure 3). The alignment showed a striking similarity of GANA-1 to all other included sequences.

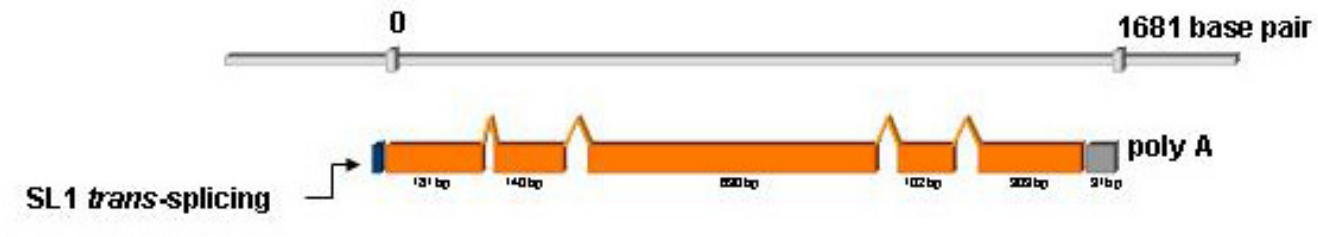


Figure 1
***gana-1* gene structure.** Schematic representation of *gana-1* gene structure. The length of genomic DNA from start to stop codons is 1681 bp. The spliced cDNA consists of 1356 bp + 91 bp of 3' UTR.

GANA-1 had the highest sequence similarity with *Anopheles gambiae* GAL (46%), the lowest similarity was observed with *Streptomyces avertimilis* GAL (22%). The results of our phylogenetic analysis are in accordance with generally accepted evolutionary concepts. The analysis identified four main clades: animal NAGAs, animal GALs, plant/lower organisms GALs and the clade containing sequences of *Drosophila melanogaster*, *Anopheles gambiae* and *Caenorhabditis elegans*. The branch including *C. elegans* is positioned between higher animal GALs and NAGAs and plant/lower organisms GALs. This position in the tree infers the evolutionary ancestrality of *gana-1* gene to both animal GALs and animal NAGAs. However, this conclusion is not in complete agreement with the presence of pairs of genes in *Drosophila* and *Anopheles* genomes annotated as α-GALs and α-NAGAs. The presence of these genes in the *Caenorhabditis/Drosophila/Anopheles* branch (and not in the GAL and NAGA clades) could be due to low divergence of these sequences from a common ancestral gene or to independent gene duplication in the *Drosophila/Anopheles* ancestral organism. It is also important to note that the phylogenetic analysis by maximum parsimony algorithm placed the *Caenorhabditis/Drosophila/Anopheles* branch into the neighborhood of the animal NAGAs branch [8]. In this case the computational algorithm probably favored the lower number of necessary sequence changes (parsimony) between GANA-1 and NAGA clade sequences.

In our opinion, the phylogenetic analysis provides evidence that the GANA-1 evolved from a common ancestor of α-GAL and α-NAGA enzymes. However, the topology of the tree could also be explained by a loss of the second gene during the evolution of *C. elegans*. In this case the enzyme found in *C. elegans* would probably be the ortholog of human α-NAGA and the lost gene would likely be the ortholog of human α-GAL. The likelihood of these two hypotheses depends on functional divergence of duplicated gene products and their dispensability for organism's metabolic pathways [21].

Homology modeling

The best Squared Root of Mean Square Deviations value (RMSD), found between GANA-1 backbone atoms and the chicken α-NAGA template [7], was 0.52 Å. The structural model of the enzyme molecule has a two-domain structure (Figure 4A). Domain I, which contains the predicted active site, adopts a $(\beta/\alpha)_8$ barrel structure which represents a common motif in many glycoside hydrolases. Less conserved is domain II that has a β domain with β sandwich structure containing a Greek key motif. The active site pocket of GANA-1 is formed by the same twelve amino acids (W31, D76, D77, Y118, C126, K152, D154, C156, S186, A189, Y190 and R211) (Figure 4B) as in chicken α-NAGA. This finding infers their identical function in catalytic reactions as described for chicken α-NAGA [7]. D134 carboxyl starts the initial nucleophilic attack and D215 carboxylic oxygen serves as a subsequent donor and acceptor of the proton during the reaction cycle.

Residues forming the "N-acetyl recognition loop" in the chicken α-NAGA [8] (S172, A175, Y176) have the closest contact with the N-acetyl moiety of the ligand. These residues are completely conserved between human and chicken NAGAs, but in human α-GAL serin 172 is replaced by glutamine and alanine 175 is replaced by leucine. The replacements with bulkier residues apparently discriminate between α-GAL and α-NAGA substrates. While NAGAs can accommodate α-galactose and can have some α-GAL activity, GALs do not have α-NAGA activity

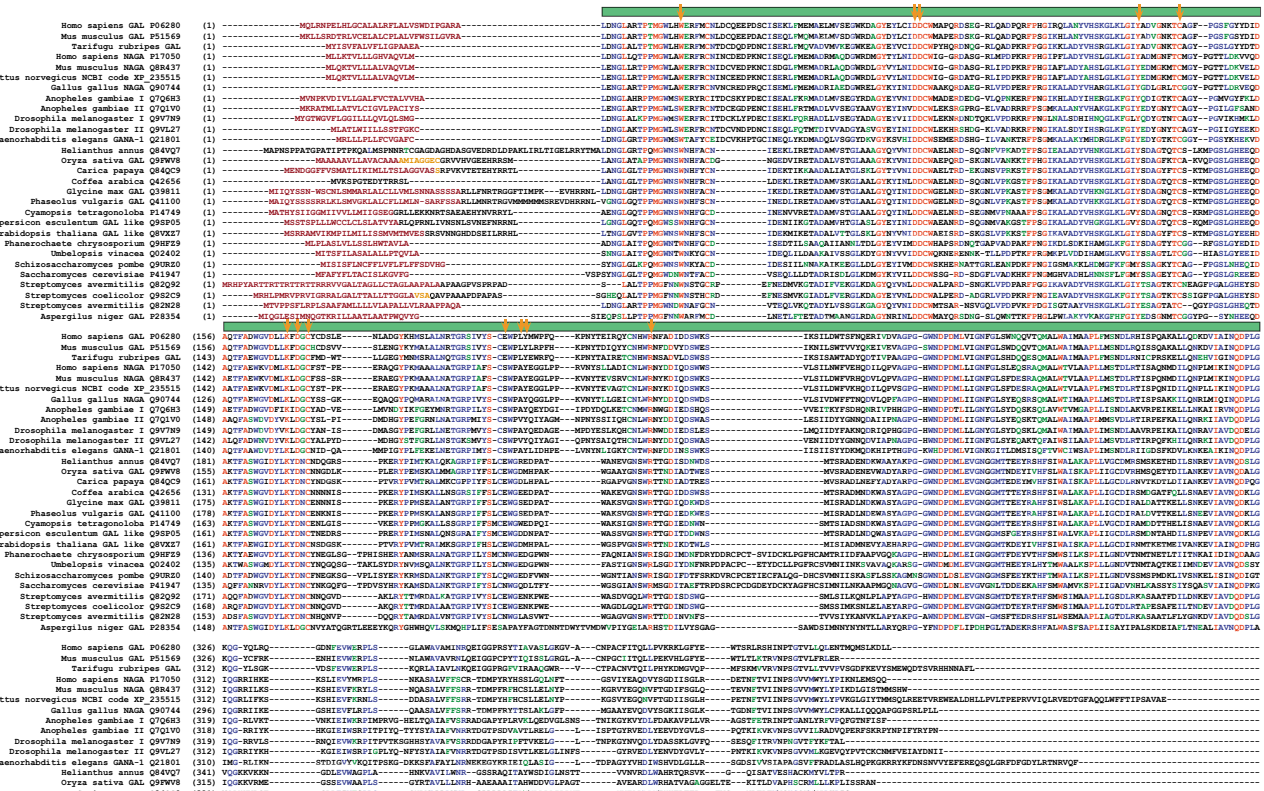


Figure 2

Multiple alignment. Multiple alignment of 29 sequences homologous with GANA-1. These sequences represent animal, plant and protozoan kingdoms. The SwissProt/TrEMBL codes are part of the sequence names. Predicted signal peptides are shown in brown letters. In cases where two signal sequence prediction algorithms gave different results the difference is marked by amber color. The residues forming active site pocket of GANA-1 are indicated by arrowheads above the alignment. The catalytic domain I is indicated by green band above the alignment.

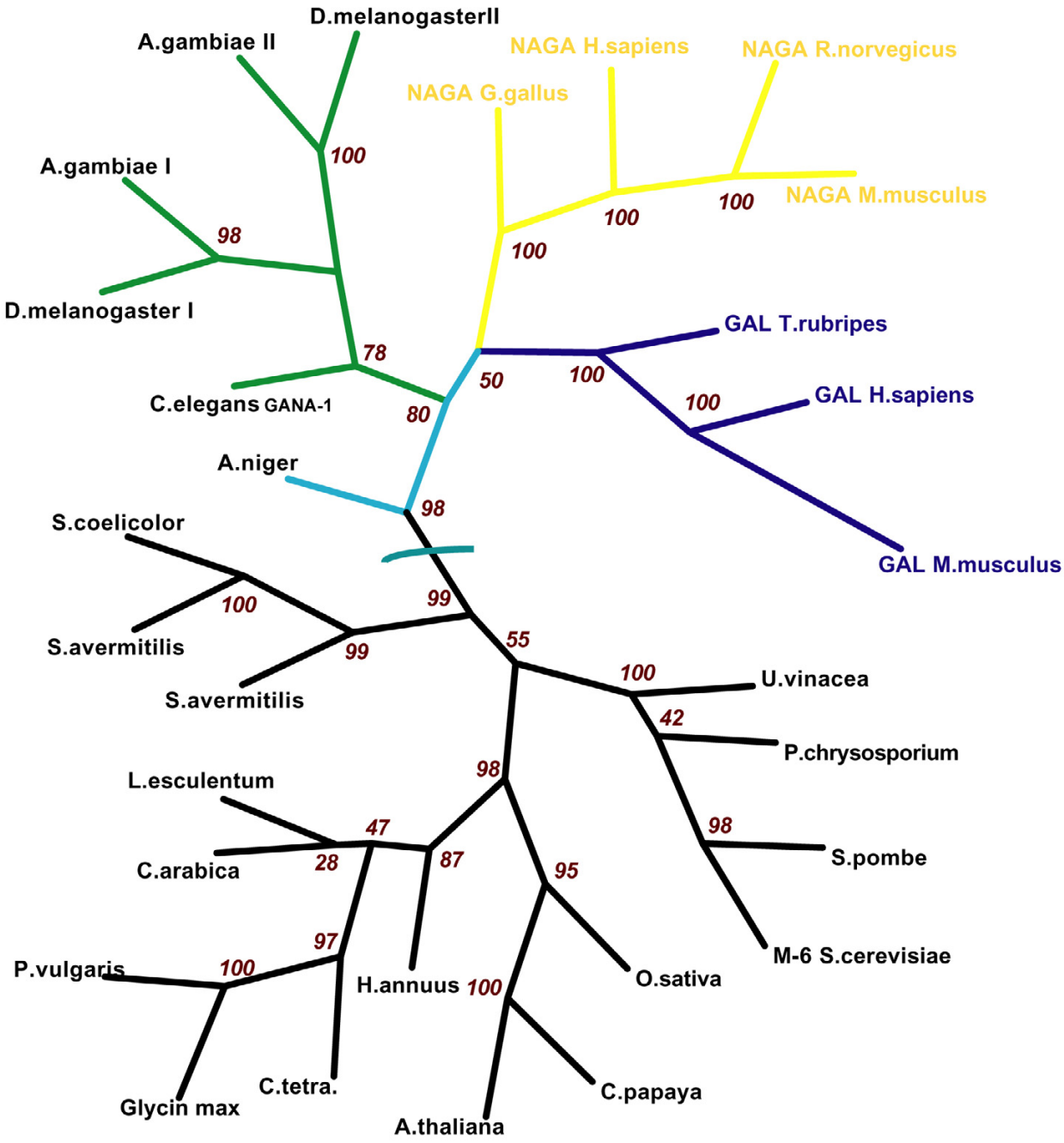


Figure 3
Cladogram of GANA-1 orthologs. Cladogram of GANA-1 orthologs. The numbers at the branch nodes represent bootstrap values.

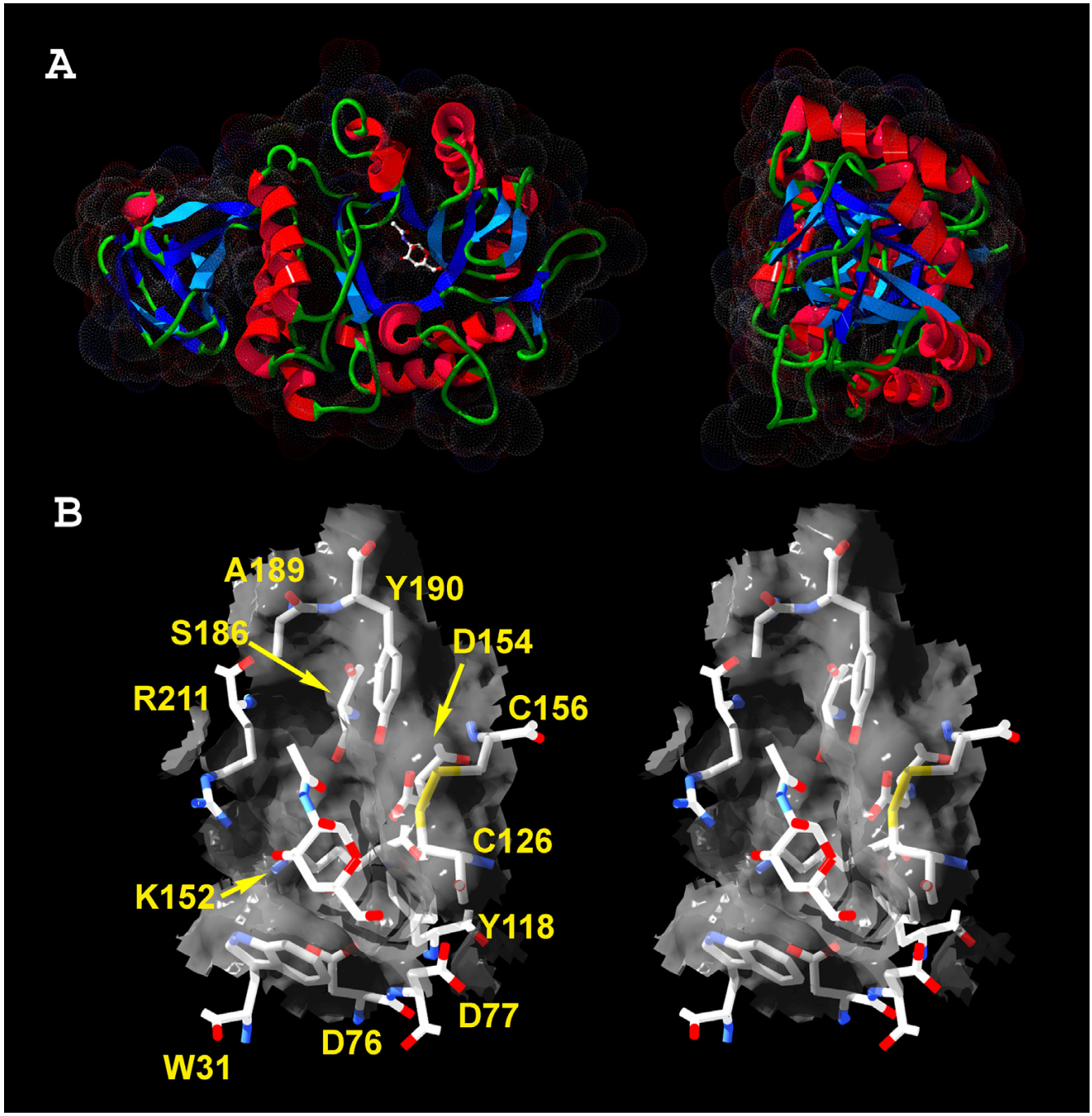


Figure 4
GANA-1 protein model. A) Ribbon representation of GANA-1 monomer model. A two-domain structure is apparent in the left picture. The N-acetyl-D-galactosamine (inhibitor) is placed into the active site. Dots represent VdW radii of surface atoms. B) Stereo picture of the active site pocket with N-acetyl-D-galactosamine (inhibitor) and amino acid labels. The viewing angles for stereo representation of the protein structure are ±2 degrees from the central axis.

and are not inhibited by N-acetylgalactosamine. The corresponding residues of GANA-1 in the NAGA recognition loop are S186 and A189 and are characteristic for NAGAs.

According to Garman [8] the key residue in the dimer interface in human α-GAL is F273. Residues corresponding to this position in other orthologs can serve as predictive markers of the protein quartenary structure. Phenylalanine or tyrosine is present in enzymes that act as homodimers while glycin indicates a monomeric structure [8]. The equivalent residue to human α-GAL F273 in GANA-1 is lysine at position 257 which is suggestive of homodimeric structure due to its sterical properties. The homology modeling showed that a groove opposing K257 is formed by residues T260, L261, D262, M263, I389, V390 and V391 of the other monomer unit of GANA-1. In the case of chicken α-NAGA these residues are equivalent to S246, Y247, E248, Q249, N375, P376 and S377 (for details see Additional file 1).

Biochemical studies

Standard bacteria/nematode separation protocol previously used by other authors [22,23] while evaluating lysosomal enzyme activites is based on sucrose flotation approach. We avoided standard sucrose flotation of worms because we could not exclude unpredictable artifacts caused by this compound, which is known to induce artificial lysosomal storage in eukaryotic cells and to alter lysosomal gene expression at concentrations significantly lower [24] than those used in flotation protocols.

We found both α-GAL and α-NAGA enzymatic activities in the homogenates from *C. elegans* N2 strain using 4-methylumbelliferyl (MU) substrates. The α-NAGA activity was dominant over the α-GAL activity. The activity of α-NAGA measured at 37°C was 430 $nmol.mg^{-1}.h^{-1}$ with MU-α-N-acetylgalactosaminide compared to the activity of α-GAL with MU-α-galactopyranoside of 43 $nmol.mg^{-1}.h^{-1}$ (about 10% of that of α-NAGA).

In the assay of α-GAL, the degradation of the MU-α-galactopyranoside was inhibited up to 95% in the presence of N-acetyl-D-galactosamine (D-GalNAc), whereas in the presence of D-galactose (D-Gal) the degradation of the same substrate was inhibited up to 75%. In the assay of α-NAGA, the degradation of the MU-α-N-acetylgalactosaminide was inhibited up to 97% by D-GalNAc and up to 90% by D-Gal. No inhibition of α-NAGA and α-GAL activity by D-glucose was observed.

According to published observations in human enzymes, D-GalNAc has no inhibitory effect on α-GAL activity. On the other hand, human α-NAGA activity is inhibited by both, D-GalNAc and D-Gal [25]. The model of GANA-1 predicts only one active site per monomer of the enzyme. If the enzyme had activity toward both substrates, MU-α-D-galactopyranoside and MU-α-N-acetylgalactosaminide, it is to be expected that D-GalNAc and D-Gal would inhibit both activities. The strong inhibitory effect of D-GalNAc on the α-GAL activity, which is not present in human α-GAL, supports the hypothesis that *C. elegans* has only one enzyme with both α-GAL and α-NAGA activities. Nevertheless, these experiments were not conducted with the pure enzyme and thus do not provide absolute proof of this hypothesis.

RNA interference

RNA interference assays directed against the whole coding region of *gana-1*, employing combination of microinjection and feeding approaches, did not reveal any abnormal morphological phenotypes. Nevertheless, measurement of α-GAL and α-NAGA activities in four different experiments showed a simultaneous decrease of both enzymatic activities in RNAi-treated worms (Table 1) as compared with control animals. In all RNAi experiments, both α-GAL and α-NAGA activities decreased proportionally, usually by tens of percent of activity of appropriate controls. The activity of the control enzyme (β-hexosaminidase) did not differ between the RNAi-treated nematodes and controls (data not shown). This finding supports the specificity of *gana-1* RNAi. The differences between individual experiments are not surprising due to the well-known variability in the efficiency of RNAi [26]. The results of RNAi experiments further support the hypothesis that GANA-1 has both enzymatic activities.

Both enzymatic activities were lower in RNAi-treated and control worms cultured on the bacterial strain HT115 [26] compared to a N2 strain cultured on the OP50 strain.

RNAi previously provided sufficient level of inhibition of structural lysosomal proteins for development of abnormal phenotypes in the worm [27,28]; however, it is apparently not efficient enough for lysosomal catalytic proteins.

Expression of gana-1

To study the expression of *gana-1* in *C. elegans*, we created transgenic worms with a reporter gene containing the entire coding region of *gana-1* C-terminally tagged with green fluorescent protein (GFP) under the control of a 3 kb region of the *gana-1* hypothetical promoter. The presence of the *gana-1::GFP* transgene in the worms was confirmed on the level of genomic DNA, cDNA and protein. However, no GFP signal was observed by fluorescence microscopy under the standard laboratory conditions. As Western blotting showed the presence of fusion protein of the expected size (data not shown), we assumed that the absence of the GFP signal was caused by a pH-dependent

Table 1: α-GAL and α-NAGA activities after *gana-1* RNAi. The table shows a proportional parallel decrease of both enzymatic activities (α-GAL and α-NAGA) after *gana-1* RNAi compared to controls.

experiment	sample	α-GAL	α-GAL (% of control)	α-NAGA	α-NAGA (% of control)	α-NAGA/α-GAL (% of control)
		nmol $mg^{-1}h^{-1}$		nmol $mg^{-1}h^{-1}$		
1	control	1.78	100	53.13	100	
	gana-1 RNAi	1.19	67	26.63	50	0.75
2	control	13.26	100	221.68	100	
	gana-1 RNAi	11.1	84	195.13	88	1.05
3	control	3.43	100	61.68	100	
	gana-1 RNAi	1.02	30	11.75	19	0.63
4	control	9.6	100	212.1	100	
	gana-1 RNAi	2.9	30	50.69	24	0.80

quenching of GFP fluorescence, which has neutral to alkaline optimum (pH 5.5–12) [29].

To study the tissue and intracellular distribution of the fusion protein, we resorted to immunofluorescence detection of the transgene product. Immunofluorescence detection of GFP fusion protein showed a specific and coarsely granular cytoplasmic pattern of fusion protein expression. This transgene product was limited to body wall muscle cells (30% of population) (Figure 5A, Additional file 2) or found in a broader tissue distribution that included body wall muscle cells, intestinal cells and coelomocytes (3–5% of population) (Figure 5B, Additional file 3).

This latter staining pattern is consistent with the GFP detection in NH_4Cl and concanamycin A (CON A) experiments (Figure 6) discussed below. The expression of the transgene was observed in about 30% of the population which corresponded to usual expression efficiency of extrachromosomal array transgenes [30,31]. The immunofluorescence staining protocol resulted in a significant decrease of inherent intestinal granular autofluorescence previously assigned to secondary lysosomes [32]. The decrease of autofluorescence intensity together with its poorly defined emission spectra hampered co-localization study.

To confirm that the absence of the GFP signal was due to the quenching of fluorescence by low pH in the acidic cellular compartment, we used two agents specifically alkalizing acidic cellular compartment [33,34] to enhance the GFP emission. Soaking of *gana-1::GFP* transgenic animals in NH_4Cl or CON A resulted in a distinct GFP signal in a vesicular compartment of endocytically active coelomocytes in a small proportion of worms (3–5% of population). The GFP signal intensity was dependent on the time of incubation and the concentration of the alkalizing agent used. The first visible GFP signal was observed after 8 hour incubation in 100 mM NH_4Cl and within 2 hours of incubation in 50 nM CON A. Lower concentrations of both NH_4Cl and CON A did not result in visible GFP signal even after prolonged incubation of up to 24 hours. The reappearance of the GFP signal after treatment of the worms with compounds increasing the acidic compartment pH indirectly confirms lysosomal localization of the fusion protein. The GFP signal in coelomocytes had the same coarsely granular pattern as that observed after immunostaining.

Limited access of alkalizing agents to the tissues can explain the differences between the results of immunofluorescence and alkalization studies.

Conclusions

Our findings showed that *gana-1* is the only ortholog of human α-NAGA and α-GAL in *C. elegans*. Based on phylogenetic and homology modeling analyses we speculate that GANA-1 most probably developed from a hypothetical ancestral gene before the duplication event which gave rise to separate α-NAGA and α-GAL genes.

We also speculate that *gana-1* gene product has both α-NAGA and α-GAL activities as detected in *C. elegans* homogenates. Importantly, both activities in the worm were inhibited by D-galactose and N-acetyl-D-galactosamine, which is a specific inhibitor of human α-NAGA and does not inhibit α-GAL.

The GANA-1::GFP fusion protein had a pattern of distribution that is compatible with lysosomal subcellular localization. The lysosomal localization of the fusion protein was also supported by pH sensitive fluorescence of GFP that was detectable only after alkalization of the acidic cellular compartment.

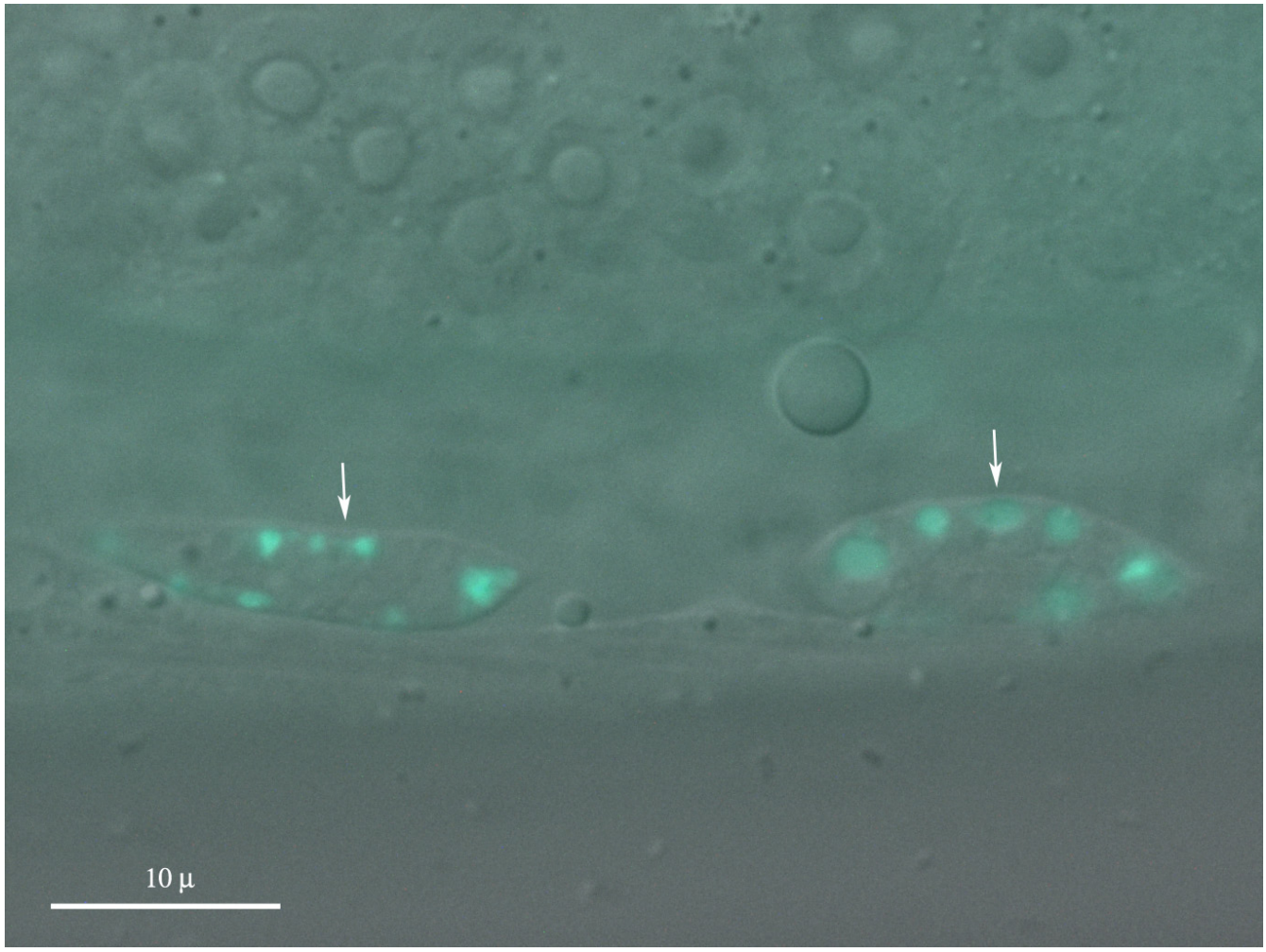


Figure 6
Alkalization of transgenic worms using CON A. Two coelomocytes showing a GFP signal in a membrane bound vesicular compartment (arrowheads) after 24 hour incubation in 50 nM CON A. DIC/fluorescence merged image.

Not suprisingly, RNAi of *gana-1* yielded no abnormal morphological phenotypes, most likely because it did not provide sufficient knockdown of enzymatic activities, necessary for development of lysosomal storage as observed in human pathology states. Nevertheless, *gana-1* RNAi resulted in a partial decrease of both enzymatic activities supporting the notion that this gene encodes both of them.

It is possible that a deletion allele of *gana-1* may provide more insight into the function of *gana-1* and efforts are underway to isolate such alleles. Deletion alleles of lysosomal hydrolases may serve as valuable models of human lysosomal storage disorders.

Methods

C. elegans *methods, strains and nomeclature*

The wild type Bristol N2 strain was used for all experiments and was handled under standard laboratory conditions as described previously [35]. Standard methods were used for DNA microinjection [36] and dsRNA synthesis and microinjection [37]. Nomenclature is in agreement with available Genetic Nomenclature for *Caenorhabditis elegans* [15] and has been approved prior to manuscript submission.

BLAST search

Wormbase (2002–2004 versions and freeze versions [15,38,39]) databases were repeatedly searched for

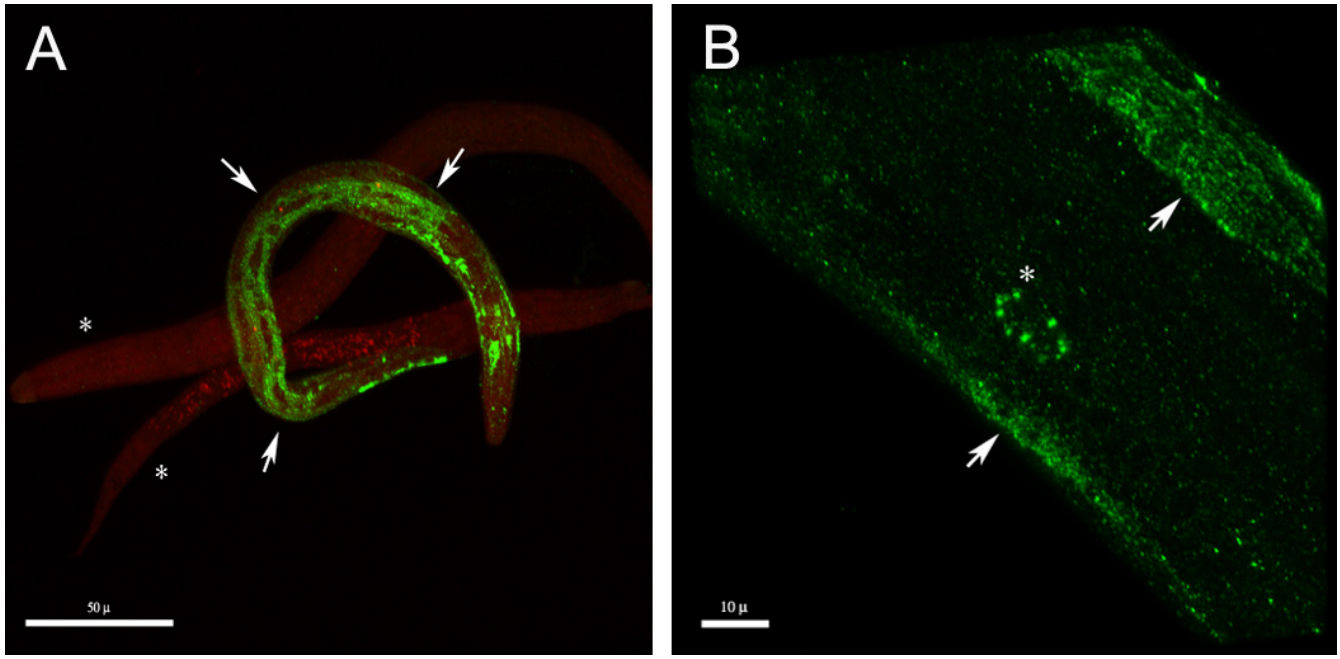


Figure 5
Immunofluorescence detection of GANA-1::GFP. A) A coarsely granular cytoplasmic distribution of immunopositivity (green) in body-wall muscle cells (arrowheads). Two non-transgenic worms are shown in the background (asterisks) for comparison. Nuclei are counterstained in red. B) Detailed view of two body wall muscle cells with coarsely granular cytoplasmic distribution of immunopositivity (arrowheads) and a coelomocyte (asterisk), both pictures were acquired by 3D rendering of initial confocal Z-stacks. Note: compare with figure 6.

human α-GAL and α-NAGA orthologs using the BLASTP [40] program set at default values. Amino acid sequences of human lysosomal α-GAL and α-NAGA (acc. no. NP_000160 and acc. no. NP_000253 [41]) were used as query sequences.

cDNA amplification and sequencing

Total RNA was isolated from mixed stages of N2 cultures [42] and reverse transcribed with an oligodT-T7 (5'-AATACGACTCACTATAG) primer and Superscript reverse transcriptase (Invitrogen). The entire coding region of R07B7.11 was PCR amplified in two overlapping PCR products, with intragenic primers designed according to available Wormbase [15] and Worfdb [16] data. SL1 primer (5'GGTTTAATTACCCAAGTTTGAG) and SL2 primer (5'GGTTTTAACCCAGTTACTCAAG) [17] together with gene specific primer (5'ATCCTGATTAATTTTAATTGC) were used to amplify 942 bp of the 5' part of the cDNA and to evaluate the mode of *trans* splicing; the 1142 bp fragment of the 3' end of cDNA was amplified with T7 primer and a gene specific primer (5'CTTAAGTTTGGAATTTATGAA). The dominant PCR products were cloned with TOPO TA cloning kit (Invitrogen) into the pCR 2.1 vector. Positive clones were sequenced using the Li-Cor automated fluorescent sequencer and sequences were aligned with R07B7 reference cosmid sequence in the AlignIR software (Li-Cor) to evaluate splicing boundaries and overall gene organization.

Multiple alignment and phylogenetic analyses

Confirmed or predicted amino acid sequences of melibiase family members [43] representing plant, unicellular, and animal kingdoms were aligned using ClustalW algorithm [44] and Blosum62 matrix. The SwissProt/TrEMBL [45] accesion code and source organism of the sequences are depicted in Figure 2. The sequence alignment was used for phylogenetic analysis with the software package PHYLIP [46]. The phylogenetic tree is based on 100 bootstraped input alignments and was constructed by maximum likelihood method with Jones-Taylor-Thornton matrix model [47]. Sequence identities between species were calculated without signal sequence in EMBOSS by Needleman-Wunsch global alignment algorithm with Blosum62 matrix, gap penalty – 10 and gap extension

penalty – 0.5 [8,48,49]. Signal peptides were predicted at the SignalP server [50] both by algorithms using neural networks and Hidden Markov Models. The results were compared to known signal sequences. The differences between signal peptides predicted by the algorithms are depicted in Figure 2.

The 3D model of GANA-1 is based on the X-ray structure of chicken α-NAGA, rice α-GAL and human α-GAL (PDB codes 1ktcA, 1uas and 1r47, respectively) [7-9,51]. The model was created using the automated homology modeling server SwissModel with structure refinement and model evaluation in the DeepView program [52]. The print quality figures (Figure 4) and animations (Additional file 1) were ray traced using PovRay software package [53].

Transgenic GFP expression

The entire coding region of the *gana-1* gene, including 3 kb of its 5'upstream sequence, was amplified from N2 genomic DNA through a nested PCR reaction using DyNAzyme EXT™ PCR system (Finnzymes) and two pairs of primers: the external pair (5'GTGAGAGTGGGGAGATAGAA and 5'TCAATTTGCTTGAGGTACATA) and the internal primers, with overhangs containing SphI and SalI restriction sites respectively (5'ACATGCATGCAACTTTCACAGGAACACAAC and 5'CGACGTCGACAATTGAACTCTATTGGTTCTCAA). The amplified DNA fragment (4709 bp) was cloned using TOPO-XL cloning kit (Invitrogen) into the pCR-XL-TOPO vector. The SphI and SalI *gana-1* restriction fragment was recloned into the GFP reporter vector pPD95.67 (supplied by A. Fire, Stanford University). The in-frame nature of the insert was confirmed by sequencing. The green fluorescent protein (GFP) fusion construct pJH3 (50 ng/μl) and pRF4 plasmid (50 ng/μl) used as the dominant marker were co-injected into the gonads of young adult N2 worms. Transgenic animals were screened for GFP signal. Nikon Eclipse E800 with C1 confocal module and 488 nm and 543 nm lasers and differential interference contrast (DIC) optics was used for specimen examination. EZ-C1 software (Nikon) was used for picture analysis and 3D rendering (Additional Files 1, 3).

Alkalization of acidic cell compartment

Mixed stage pJH3 and N2 (control) cultures were harvested from NGM OP50 plates and washed with water. Worms were pelleted by centrifugation (max. 1000 RPM, 2 min.) between the washes. Worms were treated with either one of two agents (NH_4Cl, concanamycin A – CON A) [33,34], that are known to specifically increase pH in the cellular acidic compartment. For the NH_4Cl method, animals were suspended in 0, 10, 25, 50, 75 and 100 mM aqueous solutions of NH_4Cl. Small aliquots of worms were examined after 30 min, 2, 4, 6, 8 and 24 hours. For CON A, animals were suspended in 0, 10, 20, 50, 100, 150, 200 nM solutions of CON A in aqueous media. Small aliquots of worms were examined after 1, 3, 6 and 24 hours.

Microscopical examination was performed as described above.

Immunofluorescence

The fixation and immunofluorescence staining procedures were based on the approaches of Nonet et al. [54]. Mixed stage N2 cultures were harvested from NGM OP50 plates and washed thoroughly in M9 buffer to remove intestinal bacteria. Worms were pelleted by centrifugation (1000 RPM, 2 min.) between the washes. Worms were fixed overnight in 4% paraformaldehyde in 100 mM sodium/potassium phosphate buffer. Afterwards the pellets were washed three times in 1 × PBS, and incubated in 1% Triton X-100, 100 mM Tris (pH 7.0), 1% β-Mercaptoethanol overnight at 37°C to reduce the cuticle. After 5 washes in 1 × PBS, the worms were incubated for 5 hours in 900 U/ml collagenase type IV (Sigma) diluted in Krebs-Ringer solution (119 mM NaCl, 25 mM $NaHCO_3$, 11.1 mM glucose, 1.6 mM $CaCl_2 \cdot H_2O$, 4.7 mM KCl, 1.2 mM KH_2PO_4, 1.2 mM $MgSO_4 \cdot 7H_2O$, pH 7.4). The reduction/digestion step was performed twice. Pellets were washed 3 times with 1 × PBS and stored for further processing in AbA buffer (1 × PBS, 0.1% Triton X-100, 1% BSA, 0.05% NaN_3). AbA buffer was used for antibody dilution. Primary antibody (polyclonal rabbit anti-GFP IgG (Abcam)) was diluted 1:500. Secondary antibody (goat anti-rabbit IgG Alexa Fluor 488 labeled (Molecular Probes)) was diluted 1:1000. Both incubations were performed overnight at room temperature, with AbB buffer (1 × PBS, 0.1% Triton X-100, 0.1% BSA, 0.05% NaN_3) washes in between.

Nuclei were counterstained with SYTOX orange (Molecular Probes) and the microscopic evaluation was performed as described above.

Western Blotting

Mixed stage pJH3 and N2 cultures were harvested from NGM OP50 plates. Worms were homogenized by sonication and the concentration of protein was measured by the Hartree method [55]. The proteins (equivalent of 25–50 μg of protein per lane) were separated by SDS-PAGE gradient gel (4% to 20% polyacrylamide) and transferred onto nitrocellulose membrane by semi-dry blotting. The membrane was treated according to a common Western blotting protocol with chemiluminiscence detection (SuperSignal, West Pico) [56]. Rabbit polyclonal anti-GFP IgG (Abcam, dilution 1:5000) was used as the primary antibody, the secondary antibody was goat anti-rabbit IgG/Px (Pierce, diluted 1:20 000).

RNA mediated interference

The PCR product containing the entire *gana-1* cDNA was cloned into pCRII-TOPO vector (Invitrogen) and L4440 double promoter vector for microinjection and feeding RNAi respectively. In-vitro transcription employing T7 and SP6 RNA polymerases (Promega) was used to generate antisense single stranded RNA molecules, which were annealed to generate double stranded RNA (dsRNA). dsRNA was microinjected into N2 worms which were fed on HT115 [26]*E. coli* strain carrying L4440 plasmid with *gana-1* insert. The F_1 and early F_2 progeny was screened for morphological phenotypes. N2 worms microinjected with water and fed on HT115 *E. coli* transformed with L4440 vector without insert were used as a control. 5–7 worms were microinjected both with dsRNA and water in each of 4 separate experiments, single worm progeny reaching 110–150 individuals.

Determination of α-GAL and α-NAGA and β-hexosaminidase activities

Prior to all activity measurements worms were washed from culture plates and repeatedly (6 times) washed and centrifuged in M9 buffer and finally pelleted by centrifugation. 4-methylumbelliferyl (MU)-α-D-N-acetylgalactosaminide (1 mM), 4-MU-α-D-galactopyranoside and 4-MU-β-D-glucopyranoside in theMcIlvaine buffer (0.1 M citrate/0.2 M phosphate buffer at acidic pH) were used as enzyme substrates. Reaction mixtures (sample and enzyme substrate) were incubated at 37°C and reactions were stopped by 600 μl of 0.2 M glycine/ NaOH buffer (pH 10.6) [13,57]. Fluorescence signal of the 4-methylumbelliferone was measured on the luminiscence spectrofotometer LS 50B (Perkin Elmer) (emission 365 nm and excitation 448 nm). Inhibitors (N-acetyl-D-galactosamine, D-galactose and D-glucose) were used in 0.1 M final concentration. All measurements were performed in doublets.

Authors' contributions

JH carried out molecular, RNAi, expression and biochemical analyses and wrote the first draft of the manuscript. JS participated in the design of all experiments and participated on the bioinformatic, molecular, RNAi and expression analyses and wrote the final version of the manuscript. RD performed phylogenetic analyses including homology modeling. HP participated on the biochemical analyses. MK and JL participated on the coordination of the project. MH conceived the project and provided fundraising. All authors read and approved the final version of the manuscript.

Additional material

Additional File 1

Structure of GANA-1 dimer. The color of the backbone represents differences of amino acids between GANA-1 and chicken NAGA. Blue color represents identical residues and orange stands for non-conservative changes. The colors from cyan to green represent different degrees of conservation. The surface of one monomer unit at the interface area is rendered with colors representing electrostatic potential. N-acetyl-D-galactosamine (inhibitor) is placed in the active site pocket of both monomer units (D-GalNAc arrowhead). K257 arrowhead depicts predicted dimerisation residue.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-6-5-S1.mpg]

Additional File 2

Immunofluorescence detection of GANA-1::GFP in muscle cells. 3D volume rendered and animated image corresponding to Figure 5A

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-6-5-S2.mpg]

Additional File 3

Immunofluorescence detection of GANA-1::GFP in muscle cells and coelomocytes. 3D volume rendered and animated image corresponding to Figure 5B

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-6-5-S3.mpg]

Acknowledgements

This work was supported by a grant 303/02/1324 from the Czech Science Foundation and partially also from grant VZ111100003 from Ministry of Education of the Czech Republic (institutional support).

We would like to thank Michael Krause, Ph.D. (NIDDK) and Zdeněk Kostrouch, Ph.D. (Institute of Inherited Metabolic Disorders) for critical reading of the manuscript and Eva Brožová, Kateřina Šimečková, Hana Prouzová and Lenka Brtvová (Institute of Inherited Metabolic Disorders) for technical advice and assistance.

References

1. Desnick RJ, Schindler D: **α-N-Acetylgalactosaminidase Deficiency: Schindler Disease.** In *The Metabolic & Molecular Bases of Inherited Disease Volume 3*. 8th edition. Edited by: Scriver CR, Beaudet AL, Sly WS and Valle D. New York, McGraw-Hill Companies, Inc.; 2001:3483-3506.
2. Desnick RJ, Ioannou YA, Eng CM: **α-Galactosidase A Deficiency: Fabry Disease.** In *The Metabolic & Molecular Bases of Inherited Disease Volume 3*. 8th edition. Edited by: Scriver CR, Beaudet AL, Sly WS and Valle D. New York, McGraw-Hill Companies, Inc.; 2001:3733-3774.
3. Dean KJ, Sung SS, Sweeley CC: **The identification of alpha-galactosidase B from human liver as an alpha-N-acetylgalactosaminidase.** *Biochem Biophys Res Commun* 1977, **77:**1411-1417.
4. Schram AW, Hamers MN, Brouwer-Kelder B, Donker-Koopman WE, Tager JM: **Enzymological properties and immunological characterization of alpha-galactosidase isoenzymes from normal and Fabry human liver.** *Biochim Biophys Acta* 1977, **482:**125-137.
5. Wang AM, Desnick RJ: **Structural organization and complete sequence of the human alpha-N-acetylgalactosaminidase**

gene: homology with the alpha-galactosidase A gene provides evidence for evolution from a common ancestral gene. *Genomics* 1991, **10:**133-142.
6. Henrissat B, Davies G: **Structural and sequence-based classification of glycoside hydrolases.** *Curr Opin Struct Biol* 1997, **7:**637-644.
7. Garman SC, Hannick L, Zhu A, Garboczi DN: **The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases.** *Structure (Camb)* 2002, **10:**425-434.
8. Garman SC, Garboczi DN: **The molecular defect leading to Fabry disease: structure of human alpha-galactosidase.** *J Mol Biol* 2004, **337:**319-335.
9. Fujimoto Z, Kaneko S, Momma M, Kobayashi H, Mizuno H: **Crystal structure of rice alpha-galactosidase complexed with D-galactose.** *J Biol Chem* 2003, **278:**20313-20318.
10. Asfaw B, Ledvinova J, Dobrovolny R, Bakker HD, Desnick RJ, van Diggelen OP, de Jong JG, Kanzaki T, Chabas A, Maire I, Conzelmann E, Schindler D: **Defects in degradation of blood group A and B glycosphingolipids in Schindler and Fabry diseases.** *J Lipid Res* 2002, **43:**1096-1104.
11. Dean KJ, Sweeley CC: **Studies on human liver alpha-galactosidases. II. Purification and enzymatic properties of alpha-galactosidase B (alpha-N-acetylgalactosaminidase).** *J Biol Chem* 1979, **254:**10001-10005.
12. Schram AW, Hamers MN, Tager JM: **The identity of alpha-galactosidase B from liver.** *Adv Exp Med Biol* 1978, **101:**525-529.
13. Mayes JS, Scheerer JB, Sifers RN, Donaldson ML: **Differential assay for lysosomal alpha-galactosidases in human tissues and its application to Fabry's disease.** *Clin Chim Acta* 1981, **112:**247-251.
14. **Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.** *Science* 1998, **282:**2012-2018.
15. **Wormbase** [http://www.wormbase.org/]
16. **The C. elegans ORFeome cloning project** [http://wormfdb.dfci.harvard.edu/]
17. Krause M: **Transcription and Translation.** In *Caenorhabditis elegans: Modern Biological Analysis of an Organism Volume 48*. 1st edition. Edited by: Epstein HF and Shakes DC. San Diego, Academic Press, Inc.; 1995:483-512.
18. Nimmo R, Woollard A: **Widespread organisation of C. elegans genes into operons: fact or function?** *Bioessays* 2002, **24:**983-987.
19. Blumenthal T, Gleason KS: **Caenorhabditis elegans operons: form and function.** *Nat Rev Genet* 2003, **4:**112-120.
20. Blom D, Speijer D, Linthorst GE, Donker-Koopman WG, Strijland A, Aerts JM: **Recombinant enzyme therapy for Fabry disease: absence of editing of human alpha-galactosidase A mRNA.** *Am J Hum Genet* 2003, **72:**23-31.
21. Nadeau JH, Sankoff D: **Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution.** *Genetics* 1997, **147:**1259-1266.
22. Sarkis GJ, Kurpiewski MR, Ashcom JD, Jen-Jacobson L, Jacobson LA: **Proteases of the nematode Caenorhabditis elegans.** *Arch Biochem Biophys* 1988, **261:**80-90.
23. Sebastiano M, D'Alessio M, Bazzicalupo P: **Beta-glucuronidase mutants of the nematode Caenorhabditis elegans.** *Genetics* 1986, **112:**459-468.
24. Karageorgos LE, Isaac EL, Brooks DA, Ravenscroft EM, Davey R, Hopwood JJ, Meikle PJ: **Lysosomal biogenesis in lysosomal storage disorders.** *Exp Cell Res* 1997, **234:**85-97.
25. Conzelmann E, Sandhoff K: **Glycolipid and glycoprotein degradation.** *Adv Enzymol Relat Areas Mol Biol* 1987, **60:**89-216.
26. Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J: **Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans.** *Genome Biol* 2000, **2:**.
27. Kostich M, Fire A, Fambrough DM: **Identification and molecular-genetic characterization of a LAMP/CD68-like protein from Caenorhabditis elegans.** *J Cell Sci* 2000, **113:**2595-2606.
28. Fares H, Greenwald I: **Regulation of endocytosis by CUP-5, the Caenorhabditis elegans mucolipin-1 homolog.** *Nat Genet* 2001, **28:**64-68.
29. Kneen M, Farinas J, Li Y, Verkman AS: **Green fluorescent protein as a noninvasive intracellular pH indicator.** *Biophys J* 1998, **74:**1591-1599.
30. Mello C, Fire A: **DNA transformation.** In *Caenorhabditis elegans: Modern Biological Analysis of an Organism Volume 48*. 1st edition. Edited by: Epstein HF and Shakes DC. San Diego, Academic Press; 1995:451-481.
31. Stinchcomb DT, Shaw JE, Carr SH, Hirsh D: **Extrachromosomal DNA transformation of Caenorhabditis elegans.** *Mol Cell Biol* 1985, **5:**3484-3496.
32. Clokey GV, Jacobson LA: **The autofluorescent "lipofuscin granules" in the intestinal cells of Caenorhabditis elegans are secondary lysosomes.** *Mech Ageing Dev* 1986, **35:**79-94.
33. Poole B, Ohkuma S: **Effect of weak bases on the intralysosomal pH in mouse peritoneal macrophages.** *J Cell Biol* 1981, **90:**665-669.
34. Demaurex N: **pH Homeostasis of cellular organelles.** *News Physiol Sci* 2002, **17:**1-5.
35. Brenner S: **The genetics of Caenorhabditis elegans.** *Genetics* 1974, **77:**71-94.
36. Mello CC, Kramer JM, Stinchcomb D, Ambros V: **Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences.** *Embo J* 1991, **10:**3959-3970.
37. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.** *Nature* 1998, **391:**806-811.
38. **WormBase Release WS100** [http://ws100.wormbase.org]
39. **WormBase Release WS110** [http://ws110.wormbase.org]
40. **Wormbase BLAST or BLAT Search** [http://www.wormbase.org/db/searches/blat]
41. **GenBank** [http://www.ncbi.nlm.nih.gov/Genbank/index.html]
42. Chomczynski P, Sacchi N: **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.** *Anal Biochem* 1987, **162:**156-159.
43. **Protein families database of alignments and HMMs** [http://www.sanger.ac.uk/Software/Pfam]
44. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
45. **Swiss-Prot / TrEMBL database** [http://www.expasy.org/sprot/]
46. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1988, **5:**164-166.
47. Felsenstein J: **Confidence-Limits on Phylogenies - an Approach Using the Bootstrap.** *Evolution* 1985, **39:**783-791.
48. Needleman SB, Wunsch CD: **A general method applicable to the search for similarities in the amino acid sequence of two proteins.** *J Mol Biol* 1970, **48:**443-453.
49. Rice P, Longden I, Bleasby A: **EMBOSS: The European molecular biology open software suite.** *Trends in Genetics* 2000, **16:**276-277.
50. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340:**783-795.
51. **PDB - Protein Data Bank** [http://www.rcsb.org/pdb/]
52. Guex N, Peitsch MC: **SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling.** *Electrophoresis* 1997, **18:**2714-2723.
53. **PovRay software** [http://www.povray.org/]
54. Nonet ML, Grundahl K, Meyer BJ, Rand JB: **Synaptic function is impaired but not eliminated in C. elegans mutants lacking synaptotagmin.** *Cell* 1993, **73:**1291-1305.
55. Hartree EF: **Determination of protein: a modification of the Lowry method that gives a linear photometric response.** *Anal Biochem* 1972, **48:**422-427.
56. Gallagher S, Winston SE, Fuller SA, Hurrell JGR: **Analysis of Proteins.** In *Current Protocols in Molecular Biology Volume 2*. 1st edition. Edited by: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K. Hoboken, John Wiley & Sons, Inc.; 1997:10.8.
57. Wenger DA, Williams C: **Screening for Lysosomal Disorders.** In *Techniques in Diagnostic Human Biochemical Genetics: A Laboratory Manual* 1st edition. Edited by: Hommes FA. New York, Wiley-Liss, Inc.; 1991:587-617.

ARTICLE

Mutations in *TMEM76** Cause Mucopolysaccharidosis IIIC (Sanfilippo C Syndrome)

Martin Hřebíček, Lenka Mrázová, Volkan Seyrantepe, Stéphanie Durand, Nicole M. Roslin, Lenka Nosková, Hana Hartmannová, Robert Ivánek, Alena Čížková, Helena Poupětová, Jakub Sikora, Jana Uřinovská, Viktor Stránecký, Jiří Zeman, Pierre Lepage, David Roquis, Andrei Verner, Jérôme Ausseil, Clare E. Beesley, Irène Maire, Ben J. H. M. Poorthuis, Jiddeke van de Kamp, Otto P. van Diggelen, Ron A. Wevers, Thomas J. Hudson, T. Mary Fujiwara, Jacek Majewski, Kenneth Morgan, Stanislav Kmoch,† and Alexey V. Pshezhetsky

Mucopolysaccharidosis IIIC (MPS IIIC, or Sanfilippo C syndrome) is a lysosomal storage disorder caused by the inherited deficiency of the lysosomal membrane enzyme acetyl–coenzyme A:α-glucosaminide *N*-acetyltransferase (*N*-acetyltransferase), which leads to impaired degradation of heparan sulfate. We report the narrowing of the candidate region to a 2.6-cM interval between *D8S1051* and *D8S1831* and the identification of the transmembrane protein 76 gene (*TMEM76*), which encodes a 73-kDa protein with predicted multiple transmembrane domains and glycosylation sites, as the gene that causes MPS IIIC when it is mutated. Four nonsense mutations, 3 frameshift mutations due to deletions or a duplication, 6 splice-site mutations, and 14 missense mutations were identified among 30 probands with MPS IIIC. Functional expression of human *TMEM76* and the mouse ortholog demonstrates that it is the gene that encodes the lysosomal *N*-acetyltransferase and suggests that this enzyme belongs to a new structural class of proteins that transport the activated acetyl residues across the cell membrane.

Heparan sulfate is a polysaccharide found in proteoglycans associated with the cell membrane in nearly all cells. The lysosomal membrane enzyme, acetyl–coenzyme A (CoA):α-glucosaminide *N*-acetyltransferase (*N*-acetyltransferase) is required to *N*-acetylate the terminal glucosamine residues of heparan sulfate before hydrolysis by the α-*N*-acetyl glucosaminidase. Since the acetyl-CoA substrate would be rapidly degraded in the lysosome,[1] *N*-acetyltransferase employs a unique mechanism, acting both as an enzyme and a membrane channel, and catalyzes the transmembrane acetylation of heparan sulfate.[2] The mechanism by which this is achieved has been the topic of considerable investigation, but, for many years, the isolation and cloning of *N*-acetyltransferase has been hampered by its low tissue content, instability, and hydrophobic nature.[3–5]

Genetic deficiency of *N*-acetyltransferase causes mucopolysaccharidosis IIIC (MPS IIIC [MIM 252930], or Sanfilippo syndrome C), a rare autosomal recessive lysosomal disorder of mucopolysaccharide catabolism.[6–8] MPS IIIC is clinically similar to other subtypes of Sanfilippo syndrome.[9] Patients manifest symptoms during childhood with progressive and severe neurological deterioration causing hyperactivity, sleep disorders, and loss of speech accompanied by behavioral abnormalities, neuropsychiatric problems, mental retardation, hearing loss, and relatively minor visceral manifestations, such as mild hepatomegaly, mild dwarfism with joint stiffness and biconvex dorsolumbar vertebral bodies, mild coarse faces, and hypertrichosis.[7] Most patients die before adulthood, but some survive to the 4th decade and show progressive dementia and retinitis pigmentosa. Soon after the first 3 patients with MPS IIIC were described by Kresse et al.,[6] Klein et al.[8,10] reported a similar deficiency in 11 patients who had received the diagnosis of Sanfilippo syndrome, therefore suggesting that the disease is a relatively frequent subtype. The birth prevalence of MPS IIIC in Australia,[11] Portugal,[12] and the Netherlands[13] has been estimated to be 0.07, 0.12, and 0.21 per 100,000, respectively.

The putative chromosomal locus of the MPS IIIC gene was first reported in 1992. By studying two siblings who received the diagnosis of MPS IIIC and had an apparently

From the Institute for Inherited Metabolic Disorders (M.H.; L.M.; L.N.; H.H.; R.I.; A.Č.; H.P.; J.S.; J.U.; V. Stránecký; J.Z.; S.K.) and Center for Applied Genomics (R.I.; A.Č.; V. Stránecký; J.Z.; S.K.), Charles University 1st School of Medicine, and Institute of Molecular Genetics, Academy of Sciences of the Czech Republic (R.I.), Prague; Hôpital Sainte-Justine and Départements de Pédiatrie (V. Seyrantepe; S.D.; J.A.; A.V.P) and Biochimie (A.V.P), Université de Montréal, and Research Institute of the McGill University Health Centre (N.M.R.; T.J.H.; T.M.F.; K.M.), McGill University and Genome Quebec Innovation Centre (P.L.; D.R.; A.V.; T.J.H.; J.M.), and Departments of Human Genetics (T.J.H.; T.M.F.; J.M.; K.M.), Medicine (T.J.H.; T.M.F.; K.M.), and Anatomy and Cell Biology (A.V.P.), McGill University, Montreal; Biochemistry, Endocrinology & Metabolism Unit, UCL Institute of Child Health, London (C.E.B.); Hôpital Debrousse, Lyon, France (I.M.); Department of Medical Biochemistry, Academic Medical Center UVA (B.J.H.M.P.), and Department of Clinical Genetics, VU University Medical Center (J.v.d.K.), Amsterdam; Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands (O.P.v.D.); and Laboratory of Pediatrics and Neurology, University Medical Center, Nijmegen, The Netherlands (R.A.W.)

Received June 8, 2006; accepted for publication August 8, 2006; electronically published September 8, 2006.

Address for correspondence and reprints: Dr. Alexey V. Pshezhetsky, Service de Génétique Médicale, Hôpital Sainte-Justine, 3175 Côte Sainte-Catherine, Montreal, Quebec H3T 1C5, Canada. E-mail: alexei.pchejetski@umontreal.ca

* Footnote added in proof: the gene name has been changed to *HGSNAT*.

† S.K. has led the Prague team.

Am. J. Hum. Genet. 2006;79:807–819. 0002-9297/2006/7905-0004$15.00

balanced Robertsonian translocation, Zaremba et al.[14] suggested that the mutant gene may be located in the pericentric region of either chromosome 14 or chromosome 21, but no further confirmation of this finding was provided. Previously, we performed a genomewide scan on 27 patients with MPS IIIC and 17 unaffected family members, using 392 highly informative microsatellite markers with an average interspacing of 10 cM. For chromosome 8, the scan showed an apparent excess of homozygosity in patients compared with their unaffected relatives.[15] Additional genotyping of 38 patients with MPS IIIC for 22 markers on chromosome 8 identified 15 consecutive markers (from *D8S1051* to *D8S2332*) in an 8.3-cM interval for which the genotypes of affected siblings were identical in state. A maximum multipoint LOD score of 10.6 was found at marker *D8S519,* suggesting that this region includes the locus for MPS IIIC.[15] Recently, localization of the MPS IIIC causative gene on chromosome 8 was confirmed by microcell-mediated chromosome transfer in cultured skin fibroblasts of patients with MPS IIIC.[16]

Here, we report the results of linkage analyses that narrowed the candidate region for MPC IIIC to a 2.6-cM interval between *D8S1051* and *D8S1831* and the identification of the *TMEM76* gene, located within the candidate region, as the gene that codes for the lysosomal *N*-acetyltransferase and, when mutated, is responsible for MPS IIIC.

Material and Methods

Families

In Montreal, 33 affected individuals and 35 unaffected relatives comprising 15 families informative for linkage were genotyped. The families came from Europe, North Africa, and North America. An additional 27 affected individuals and 9 unaffected relatives in uninformative pedigrees, as well as 40 controls, were also genotyped. Eleven of these families and the controls have been reported elsewhere.[15] In addition, 54 individuals from four MPS IIIC–affected families from the Czech Republic were studied in Prague (fig. 1). One family had two affected brothers, whereas the remaining three families each had one affected individual. The families came from various regions of the Czech Republic and were not related within the four most-recent generations. The diagnosis for affected individuals was confirmed by the measurement of *N*-acetyltransferase activity in cultured skin fibroblasts or white blood cells.

Genotyping

The samples in Montreal were genotyped for 22 microsatellite markers in the pericentromeric region of chromosome 8 spanning 8.9 cM on the Rutgers map, version 2.0.[17] The genotyping was performed as described by Mira et al.[18] at the McGill University and Genome Quebec Innovation Centre on an ABI 3730xl DNA Analyzer platform (Applied Biosystems). Alleles were assigned using Genotyper, version 3.6 (Applied Biosystems). The random-error model of SimWalk2, version 2.91,[19,20] was used to detect potential genotyping errors, with an overall error rate of 0.025. Nine genotypes for which the posterior probability of being incorrect was >0.5 were removed before subsequent analyses. In addition, nine genotypes for one marker in one family were removed because of a suspected microsatellite mutation. The samples from the Czech Republic were genotyped in Prague for 18 microsatellite markers in an 18.7-cM region that includes the 8.9-cM region mentioned above. The genotyping was performed on an LI-COR IR2 sequencer by use of Saga genotyping software (Li-Cor) as described elsewhere.[21] Genotypes were screened for errors by use of the PedCheck program.[22]

Linkage Analysis

For the families genotyped in Montreal, multipoint linkage analysis was performed using the Markov chain–Monte Carlo (MCMC) method implemented in SimWalk2, version 2.91,[19] since one pedigree was too large to be analyzed by exact computation. A fully penetrant autosomal recessive parametric model was used with a disease-allele frequency of 0.0045. Marker-allele frequencies were estimated by counting alleles in the available parents of patients with MPS IIIC and in control individuals. To check the consistency of the results, the MCMC analysis was repeated four times.

N-acetyltransferase activity was measured in all participants of the four families from the Czech Republic.[23] Individuals were classified as affected, carriers, or unaffected on the basis of the results of this assay. Mean affected and carrier activities were determined from the five affected individuals and their seven obligate heterozygote parents, respectively, whereas the mean control activity was determined from a sample of 89 unrelated individuals. Four individuals were unable to be classified because their values were within 2 SDs of the means of both the control and carrier groups. Multipoint linkage analysis was performed using a codominant model with a penetrance of 0.99 and a phenocopy rate of 0.01, to account for the possibility of misclassification or genotyping errors. The same disease-allele frequency of 0.0045 was used. Marker-allele frequencies were estimated by counting all genotyped individuals. Exact multipoint linkage analysis was run on 18 microsatellite markers by use of Allegro 1.2c,[24] which was also used to infer haplotypes.

Gene-Expression Analysis

For each of 32 genes located in the candidate interval, a single 5′-amino–modified 40-mer oligonucleotide probe (Illumina) was spotted in quadruplicate on aminosilane-modified microscopic slides and was immobilized using a combination of baking and UV cross-linking. Total RNA (250–1,000 ng) from white blood cells of two patients with MPS IIIC (patients AIV.8 and BIII.5) and four healthy individuals were amplified using the SenseAmp plus RNA Amplification Kit (Genisphere) and were reverse transcribed using 300 ng of poly(A)-tailed mRNA. Reverse transcription and microarray detection were done using the Array 900 Expression Detection Kit (Genisphere) according to the manufacturer's protocol. The two patient samples and four control samples were analyzed in dye-swap mode, in two replicates of each mode. The hybridized slides were scanned with a GenePix 4200A scanner (Molecular Devices), with photomultiplier gains adjusted to obtain the highest-intensity unsaturated images. Data analysis was performed in the R statistical environment (The R Project for Statistical Computing, version 2.2.1) by use of the Linear Models for Microarray Data package (Limma, version 2.2.0).[25] Raw data were processed using loess normalization and a moving minimum background correction on individual arrays and quantile

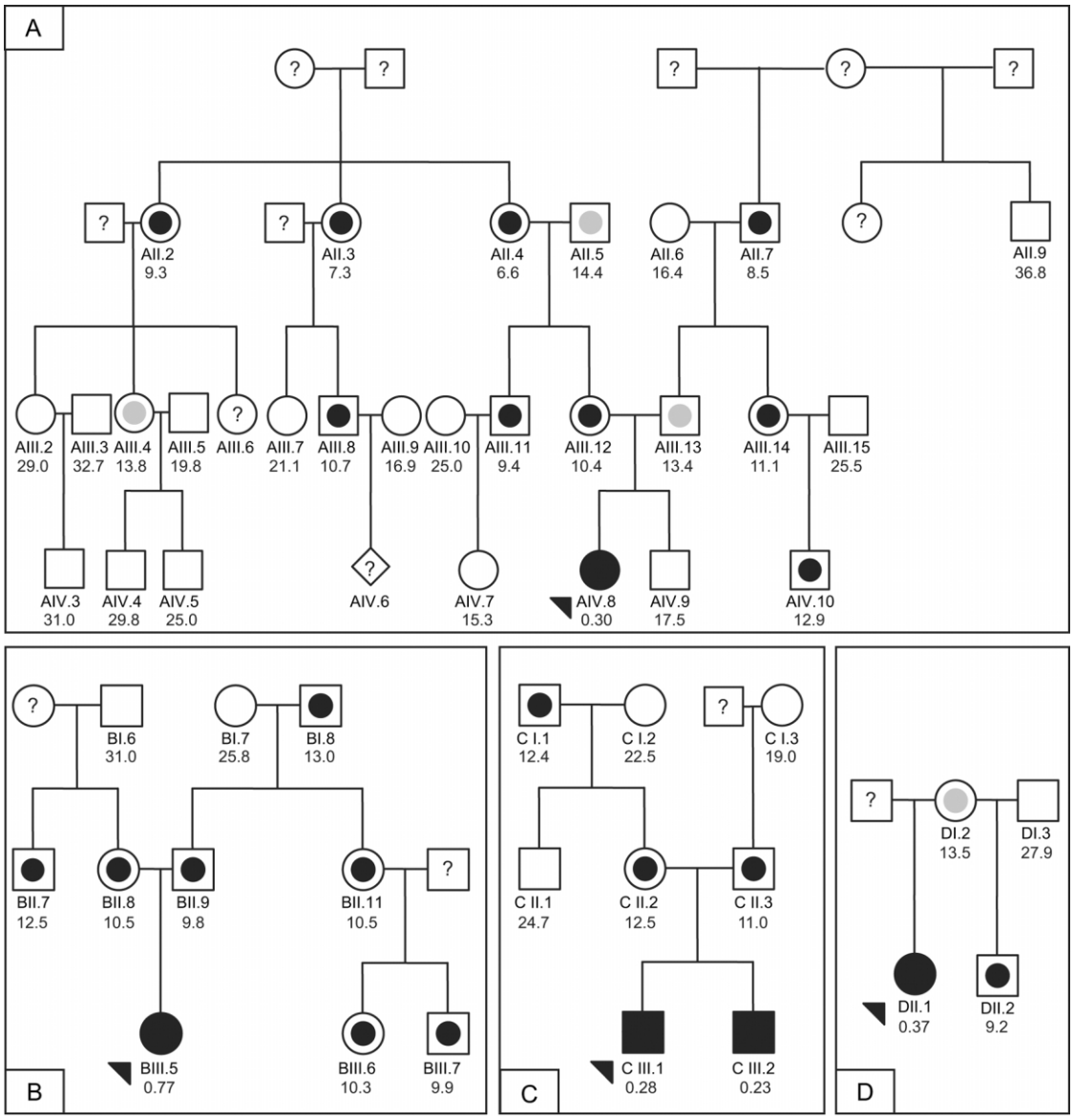


Figure 1. Four families from the Czech Republic used in the linkage and mutation analyses. Fully blackened symbols indicate individuals with MPS IIIC; arrowheads indicate probands. Measurements in seven obligate heterozygotes from these pedigrees (mean ± SD 11.6 ± 1.5 nmol/h/mg) and 89 controls not known to be related to members of the pedigree (mean ± SD 24.4 ± 5.7 nmol/h/mg) were used to establish *N*-acetyltransferase activity ranges for heterozygotes (symbols with blackened inner circle) and normal homozygotes (open symbols). An individual was assigned to a class if his or her enzyme activity was within 2 SDs of the class, unless the value was within the overlap of the upper end of the obligate heterozygotes and the lower end of the controls. Individuals with values within the open interval 13.0–14.6 nmol/h/mg were classified as unknown (symbols with gray inner circle). A symbol with a question mark (?) indicates that no material was available for the enzyme assay. DNA was available for individuals with ID numbers, and *N*-acetyltransferase activity measurements in white blood cells are shown below the ID numbers.

normalization between arrays. The correlation between four duplicate spots per gene on each array was used to increase the robustness. A linear model was fitted for each gene given a series of arrays by use of the lmFit function. The empirical Bayes method[26] was used to rank the differential expression of genes by use of the eBayes function. Correction for multiple testing was performed using the Benjamini and Hochberg false-discovery–rate method.[27] We considered genes to be differentially expressed if the adjusted *P* value was <.01.

DNA and RNA Isolation and Sequencing

Cultured skin fibroblasts from patients with MPS IIIC and normal controls were obtained from cell depositories (Hôpital Debrousse,

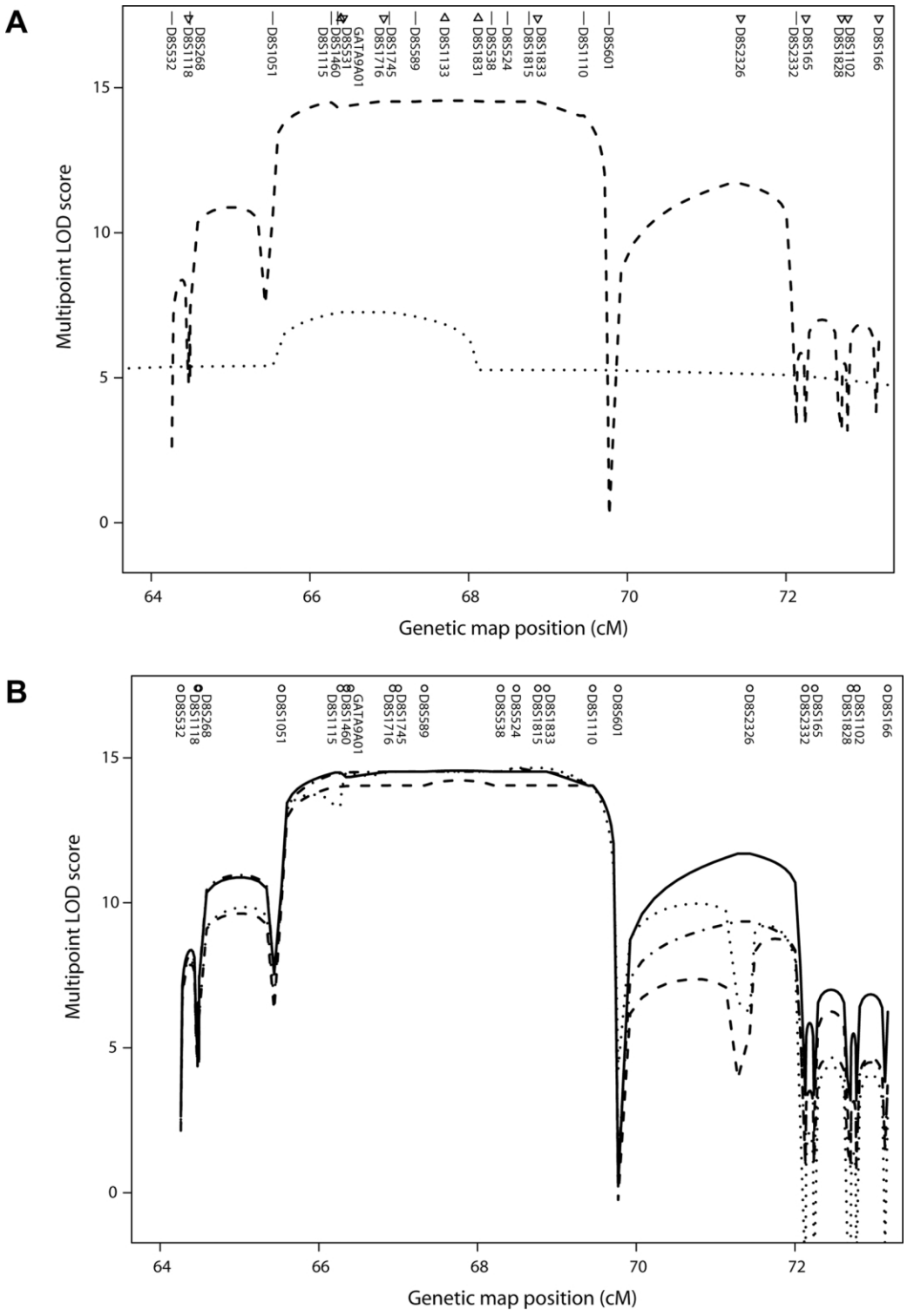


Figure 2. Multipoint linkage analysis of MPS IIIC on chromosome 8. *A,* Multipoint LOD scores in an 8.9-cM interval from two sets of families. Symbols above the marker names indicate the map position. Marker names are listed in the correct order but may be displaced from the symbols for visibility. The dashed line is based on families genotyped in Montreal, and the dotted line on families genotyped in Prague. Straight lines next to marker names indicate that the markers were typed in both data sets. Triangles pointing down indicate markers typed only in the Montreal data set, and triangles pointing up indicate markers typed only in the Prague data set. For the Montreal data, the SimWalk2 run with the highest likelihood is shown. *TMEM76* lies between *D8S1115* and *D8S1460,* and, according to the March 2006 freeze of the human genome sequence from the University of California–Santa Cruz Genome Browser,[30] the order is *D8S1115*–(500 kb)–*TMEM76*–(800 kb)–centromere–(200 kb)–*D8S1460. B,* Multipoint LOD scores from the Montreal data from four runs of SimWalk2, version 2.91,[19] showing the variation between runs.

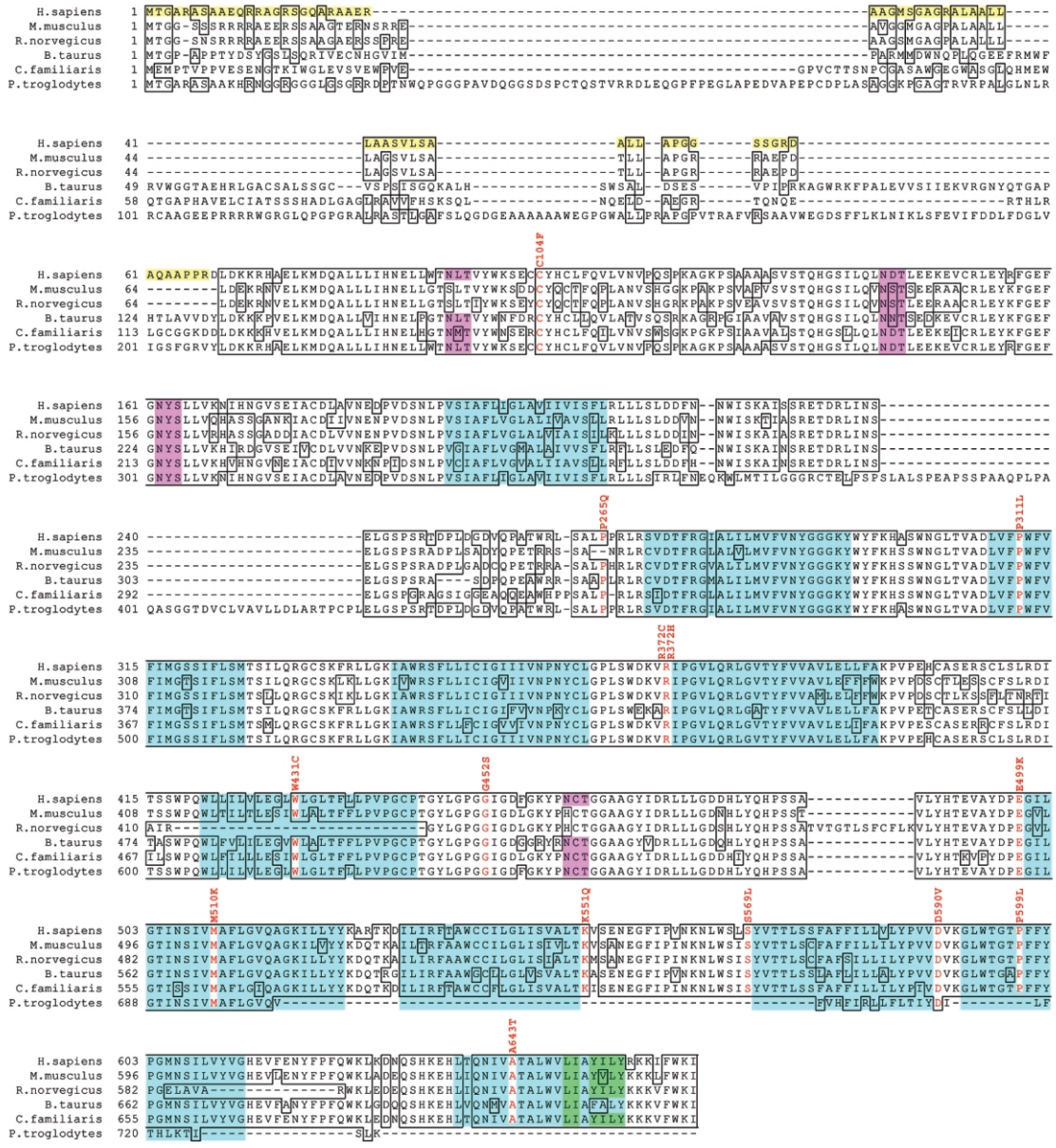


Figure 3. Predicted amino acid sequence of the TMEM76 protein. Amino acid sequence alignment of *Homo sapiens* TMEM76 with orthologs from *Mus musculus* (cloned sequence), *Canis familiaris* (GenBank accession number XP_539948.2), *Bos taurus* (XP_588978.2), *Rattus norvegicus* (XP_341451.2), and *Pan troglodytes* (XP_519741.1) by use of BLAST. All cDNA sequences are predicted except the sequence for *M. musculus*. The identical residues are boxed, the residues with missense mutations in patients with MPS IIIC are shown in red, and the amino acid changes are indicated above the sequence. The first 67 aa of the human sequence shown as black on yellow comprise the predicted signal peptide. The predicted transmembrane domains in the human sequence are shown as black on turquoise. The topology model[5–7] strongly predicts that the N-terminus is inside the lysosome and the C-terminus is outside. Four predicted *N*-glycosylation sites are shown as black on pink, and the predicted motifs for the lysosomal targeting, as black on green.

France; NIGMS Human Genetic Mutant Cell Repository; Montreal Children's Hospital, Canada; and Department of Clinical Genetics, Erasmus Medical Center, The Netherlands). Blood samples from patients with MPS IIIC, their relatives, and controls were collected with ethics approval from the appropriate institutional review boards. DNA from blood or cultured skin fibroblasts was extracted using the PureGene kit (Gentra Systems). Total RNA from cultured skin fibroblasts and pooled tissues (spleen, liver, kidney, heart, lung, and brain) of a C57BL/6J mouse was isolated using Trizol (Invitrogen), and first-strand cDNA synthesis was prepared with SuperScript II (Invitrogen). DNA fragments containing *TMEM76* exons and adjacent regions (~40 bp from each side; primer sequences are shown in appendix A) were amplified by PCR from genomic DNA and were purified with Montage PCR96

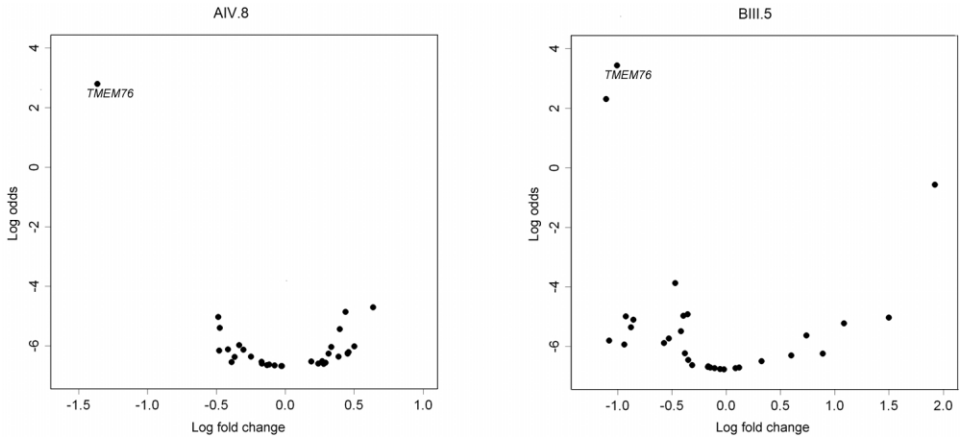


Figure 4. Volcano plot of genes located within the MPS IIIC candidate region, showing significantly reduced expression of the *TMEM76* gene in white blood cells of two patients with MPS IIIC: AIV.8 and BIII.5. The natural logarithm of the probability that the gene is differentially expressed (Log odds) is plotted as a function of the logarithm of the gene-expression $\log_2$ fold change (Log fold change) between the patient and control samples.

filter plates (Millipore). Each sequencing reaction contained 2 μl of purified PCR product, 5.25 μl of H_2O, 1.75 μl of 5 × sequencing buffer, 0.5 μl of 20 μM primer, and 0.5 μl of Big Dye Terminator v3.1 (all from Applied Biosystems). In Montreal, PCR products were analyzed using an ABI 3730xl DNA Analyzer (Applied Biosystems). In Prague, PCR products were analyzed on an ALFexpress DNA sequencer (Pharmacia), as described elsewhere.[28] Included in the sequencing analysis were 30 probands with MPS IIIC who were considered unrelated and 105 controls. The controls were unrelated CEPH individuals, and amplified DNAs were combined in pools of two before sequencing.

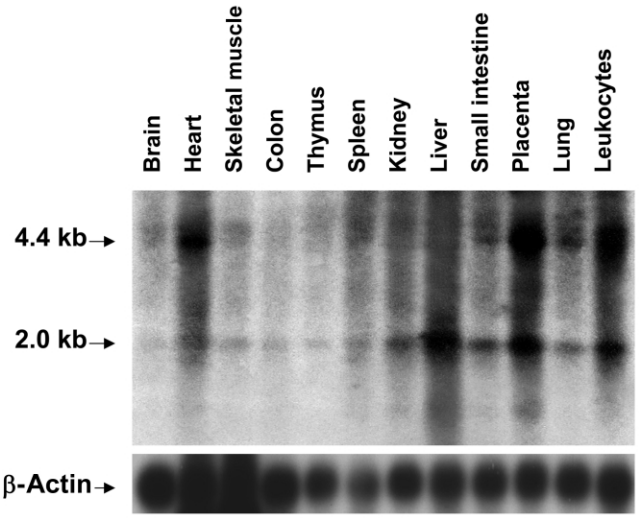


Figure 5. Northern-blot analysis of *TMEM76* mRNA in human tissues. A 12-lane blot containing 1 μg of poly A+ RNA per lane from various adult human tissues was hybridized with a [^{32}P]-labeled 220-bp cDNA fragment corresponding to exons 8–10 of the *TMEM76* gene or β-actin, as described in the Material and Methods section.

Northern Blotting

A 12-lane multiple-tissue northern blot containing 1 μg of poly A+ RNA per lane from various human tissues (BD Biosciences Clontech) was hybridized with the 220-bp cDNA fragment corresponding to exons 8–10 of the human *TMEM76* gene or the entire cDNA of human β-actin labeled with [^{32}P]-dCTP by random priming with the MegaPrime labeling kit (Amersham). Prehybridization of the blot was performed at 68°C for 30 min in ExpressHyb (Clontech). The denatured probes were added directly to the prehybridization solution and were incubated at 68°C for 1 h. The blots were washed twice for 30 min at room temperature with 2 × sodium chloride–sodium citrate (SSC) solution and 0.05% SDS and once for 40 min at 50°C with 0.1 × SSC and 0.1% SDS and were exposed to a BioMax film for 48 h.

Mouse and Human TMEM76 *cDNA Cloning*

Mouse coding sequence was amplified by PCR (forward primer 5′-GAATTCATGACGGGCGGGTCGAGC-3′; reverse primer 5′-ATATGTCGACGATTTTCCAAAACAGCTTC-3′) and was cloned into pCMV-Script, pCMV-Tag4A (Stratagene), and pEGFP-N3 (BD Biosciences Clontech) vectors by use of the *Eco*RI and *Sal*I restriction sites of the primers. The cloned sequence was identical to GenBank accession number AK152926.1, except that an "AT" was needed to complete an alternate ATG initiation codon. GenBank accession number AK149883.1 provides what we consider to be the complete clone and encodes a 656-aa protein. The GenBank sequences differ by 1 aa and three silent substitutions.

A 1,907-bp fragment of the human *TMEM76* cDNA (nt +75 to +1992) was amplified using Platinum High Fidelity *Taq* DNA polymerase (Invitrogen), a sense primer with an *Hin*dIII site (5′-AAGCTTGGCGGCGGGCATGAG-3′), and an antisense primer with an *Sal*I site (5′-GTCGACCTCAGTGGGAGCCATCAGATTTT-3′) and was cloned into pCMV-Script expression vector (Stratagene). Since high GC content (85%) of the 5′ region of human *TMEM76* cDNA prevented its amplification by PCR, a synthetic 186-bp codon-optimized double-stranded oligonucleotide fragment (5′-AAGCTTATGACCGGAGCGAGGGCAAGCGCCGCCG-

AACAAAGAAGAGCCGGACGGTCCGGCCAGGCTAGGGCCGC-AGAGCGAGCTGCTGGCATGTCAGGTGCAGGGCGCGCACTTG-CCGCCTTGCTGCTCGCCGCGAGTGTGCTGAGCGCTGCCCTC-CTGGCTCCCGGAGGCTCTTCCGGGCGGGAC-3′) corresponding to nt +1 to +186 of human *TMEM76* cDNA was purchased from BioS&T. A 177-bp 5′ fragment was combined with rest of the cDNA by use of *Hin*dIII and *Sap*I sites. The cloned sequence is identical to GenBank accession number XM_372038.4 from nt 131 to nt 1946, except for the presence of SNP *rs1126058*.

Cell Culture and Transfection

Skin fibroblasts and COS-7 cells were cultured in Eagle's minimal essential medium (Invitrogen) supplemented with 10% (v/v) fetal bovine serum (Invitrogen) and were transfected with the full-size mouse *Tmem76* (*Hgsnat*) coding sequence subcloned into pCMV-Script, pCMV-Tag4A, and pEGFP-N3 vectors or with the full-size human *TMEM76* coding sequence subcloned into pCMV-Script vector by use of Lipofectamine Plus (Invitrogen) according to the manufacturer's protocol. The cells were harvested 48 h after transfection, and *N*-acetyltransferase activity was measured in the homogenates of *TMEM76*-transfected and mock-transfected cells (i.e., transfected with only the cloning vector).

Enzyme Assay

N-acetyltransferase enzymatic activity was measured using the fluorogenic substrate 4-methylumbelliferyl β-D-glucosaminide (Moscerdam) as described elsewhere.[23] Protein concentration was measured according to the method of Bradford.[29] This assay was used for the activity measurements in cultured skin fibroblasts or white blood cells from patients and all participating members of the Czech families and for the functional expression experiments.

Confocal Microscopy

To establish colocalization of the tagged protein with the lysosomal compartment, the skin fibroblasts expressing mouse TMEM76-EGFP were treated with 50 nM LysoTracker Red DND-99 dye (Molecular Probes), were washed twice with ice-cold PBS, and were fixed with 4% paraformaldehyde in PBS for 30 min. Slides were studied on an LMS 510 Meta inverted confocal microscope (Zeiss).

Results

Linkage Analysis

Previously, we performed a genomewide linkage study that indicated that the locus for MPS IIIC is mapped to an 8.3-cM interval in the pericentromeric region of chromosome 8.[15] To reduce this interval, we genotyped the families from that study as well as newly obtained MPS IIIC–affected families for 22 microsatellite markers (Montreal data). Linkage analysis under an autosomal recessive model resulted in LOD scores >14 in the 4.2-cM region spanning *D8S1051* to *D8S601,* which included the centromere (fig. 2). The results of multiple MCMC runs showed consistent trends. Linkage was also performed in four families from the Czech Republic by use of an autosomal codominant model (Prague data). For these data, linkage analysis produced a maximum LOD score of 7.8 at 66.4 cM at *D8S531* and reduced the linked region for the Montreal data to a 2.6-cM interval between *D8S1051* and *D8S1831*. This region was defined by inferred recombinants at *D8S1051* in one family in each of the Montreal and Prague data sets, and a recombinant at *D8S1831* in an additional family in the Prague data set. This interval contains 32 known or predicted genes and ORFs.

Identification of a Candidate Gene

On the basis of our previous studies that defined the molecular properties of the lysosomal *N*-acetyltransferase,[31] we searched the candidate region for a gene encoding a protein with multiple transmembrane domains and a molecular weight of ~100 kDa, which allowed us to exclude the majority of the genes in the region. In contrast, the predicted protein product of the *TMEM76* gene has multiple putative transmembrane domains. The predicted coding region in GenBank accession number XM_372038.4 was extended by 28 residues at the 5′ end on the basis of the transcript in GenBank accession number DR000652.1 (which includes 14 of the 28 residues), examination of the genomic sequence in NT_007995.14, and comparison with mouse sequence AK149883.1. We predict that the modified *TMEM76* contains 18 exons, corresponding to an ORF of 1,992 bp, and codes for a 73-kDa protein. A comparison of human TMEM76 with five vertebrate orthologs is shown in figure 3. Furthermore, of all the genes present in the candidate interval, only *TMEM76* showed a statistically significant reduction of the transcript level in the cells of two patients with MPS IIIC (AIV.8 and BIII.5; adjusted P values $< .001$) in the custom oligonucleotide-based microarray assay (fig. 4). Further, we showed that both patients carried nonsense mutations presumably causing mRNA decay (R534X and L349X; see table 2).

Analysis of the TMEM76 *Transcript by Northern Blot and RT-PCR*

Northern-blot analysis identified two major *TMEM76* transcripts of 4.5 and 2.1 kb ubiquitously expressed in various human tissues (fig. 5). The highest expression was detected in leukocytes, heart, lung, placenta, and liver, whereas the gene was expressed at a much lower level in the thymus, colon, and brain, which is consistent with the expression patterns of lysosomal proteins. Consistent with the northern-blot results, a full-length 4.5-kb cDNA containing 1,992 bp of coding sequence and two polyadenylation signals as well as two shorter transcripts were amplified by RT-PCR from the total RNA of normal human skin fibroblasts, white blood cells, and skeletal muscle. In one transcript, exons 9 and 10 were spliced out, leading to an in-frame deletion of 64 aa, which contains the predicted transmembrane domains III and IV. Most likely, this transcript does not encode an active enzyme, since it was also detected in the RNA of two patients with MPS IIIC (patients CIII.1 and CIII.2) who had almost complete loss of

N-acetyltransferase activity. Another transcript lacked exons 3, 9, and 10.

The deduced amino acid sequence predicts 11 transmembrane domains and four potential *N*-glycosylation sites (fig. 3), consistent with the molecular properties of lysosomal *N*-acetyltransferase.[31] The first 67 aa may comprise the signal peptide, with length and composition resembling those of lysosomal proteins. According to the predictions made by empirical computer algorithms,[32–34] the C-terminus of the TMEM76 protein is exposed to the cytoplasm and contains conserved Tyr-X-X-Θ and Leu-Leu sequence motifs involved in the interaction with the adaptor proteins responsible for the lysosomal targeting of membrane proteins.[35]

Mutation Analysis

We identified 27 *TMEM76* mutations in the DNA of 30 MPS IIIC–affected families (table 1) that were not found in DNA from 105 controls. Among the identified mutations, there were 4 nonsense mutations, 14 missense mutations, 3 predicted frameshift mutations due to deletions or duplications, and 6 splice-site mutations. All the missense mutations occur at residues conserved among five species with the most homologous TMEM76 sequences (fig. 3), except for P265Q, which is not conserved in the mouse, and W431C, which is not conserved in the rat. There were three instances of two mutations on the same allele that were found in patients who were homozygous, and these are designated as complex mutations in table 1. cDNA sequencing of one of the patients homozygous for the splice-site mutation in intron 2 and a missense mutation (P265Q) demonstrated that the splice-site mutation disrupts the consensus splice-site sequence between exon 2 and intron 2 and causes exon 2 skipping and a frameshift (not shown).

Consanguinity was reported in 4 of the 13 families in which the patients were homozygous for *TMEM76* mutations: the two Moroccan families, the French family with two missense mutations (W431C and A643T), and the Turkish family with the splice-site mutation in intron

Table 1. Mutations in *TMEM76* Identified in Patients from 30 Families with MPS IIIC

Mutation Group and Mutation[a]	Predicted Effect on Protein	No. of Alleles	Location in *TMEM76*
Nonsense mutations:			
c.1031G→A	p.W344X	2	Exon 10
c.1046T→G	p.L349X	1	Exon 10
c.1234C→T	p.R412X	8	Exon 12
c.1600C→T	p.R534X	1	Exon 15
Missense mutations:			
c.311G→T	p.C104F	1	Exon 2
c.932C→T	p.P311L	3	Exon 9
c.1114C→T	p.R372C	3	Exon 11
c.1115G→A	p.R372H	1	Exon 11
c.1354G→A	p.G452S	2	Exon 13
c.1495G→A	p.E499K	3	Exon 14
c.1529T→A	p.M510K	1	Exon 14
c.1706C→T	p.S569L	4	Exon 17
c.1769A→T	p.D590V	1	Exon 17
c.1796C→T	p.P599L	1	Exon 17
Frameshift mutations:			
c.1118_1133del	p.I373SfsX3	1	Exon 11
c.1420_1456dup	p.V488GfsX22	1	Exon 13
c.1834delG	p.V612SfsX16	1	Exon 18
Splice-site mutations:			
c.202+1G→A	p.L69EfsX32[b]	1	Intron 1
c.577+1G→A	p.P193HfsX20[b]	1	Intron 4
c.935+5G→A	p.F313X	1	Intron 9
c.1334+1G→A	p.G446X[b]	1	Intron 12
c.1810+1G→A	p.S567NfsX14	2	Intron 17
Complex mutations:			
c.[318+1G→A; 794C→A]	p.[D68VfsX19; P265Q]	6	Intron 2; exon 7
c.[577+1G→A; 1650A→C]	p.[P193HfsX20; K551Q]	2	Intron 4; exon 16
c.[1293G→T; 1927G→A]	p.[W431C; A643T]	2	Exon 12; exon 18

[a] Mutation names were assigned according to the guidelines of the Human Genome Variation Society and on the basis of the cDNA sequence from GenBank accession number NT_007995.14, except that the first exon includes 84 nt 5′ of the stated ATG initiation codon. Thus, +1 corresponds to the A of the ATG at nt 13315945 (instead of nt 13316029).

[b] The mutations were named under the assumption that no exon skipping takes place; cDNA sequencing was not done.

17. The two Moroccan families were not known to be related to each other or to the Spanish patient homozygous for the same mutations (table 2). The parents of the French patient are second cousins in two ways (see family F1 in the work of Ausseil et al.[15]).

The splice-site mutation in the above-mentioned Turkish family disrupts the consensus splice-site sequence between exon 17 and intron 17 and causes exon 17 skipping and a frameshift in all transcripts, as detected by sequencing of multiple RT-PCR clones (not shown). The two affected siblings in this family (family F8 in the work of Ausseil et al.[15]) had a severe form of MPS IIIC and showed almost complete loss of *N*-acetyltransferase activity in cultured skin fibroblasts. Among other severely affected patients with MPS IIIC, a patient of French origin was homozygous for a nonsense mutation (W344X) in exon 10, which may result in the synthesis of a truncated protein or RNA decay. A patient of Polish origin was a compound heterozygote for a 37-bp duplication in exon 13 and a missense mutation (S569L) in exon 17 (table 2). The duplication results in a frameshift, whereas the substitution of a strictly conserved small polar Ser for a bulky hydrophobic Leu may have a significant structural impact (fig. 3).

The five patients from four Czech families are all compound heterozygotes for eight different mutations (table 2). Five of the eight mutations are predicted to result in truncated products (three nonsense mutations, one 16-bp deletion, and one splice-site mutation leading to the inclusion of 89 bases from the 5′ end of intron 9 and the splicing out of exon 10 in the transcript, and the remaining three are missense mutations affecting residues conserved among multiple species and located either in the predicted transmembrane regions (fig. 3) or in their close vicinity, suggesting that they may have a serious structural impact. In the Czech families, the mutations completely segregated with reduced enzyme activity. That is, all individuals assigned to be heterozygotes on the basis of the enzyme assay as well as the four individuals who were within 2 SD of the lower end of the controls (symbols with gray inner circle in fig. 1) were found to carry *TMEM76* mutations.

Functional Expression Studies

The fibroblast cell line from a patient homozygous for a splice-site mutation in intron 17 with negligible *N*-acetyltransferase activity was transfected with plasmids containing human *TMEM76* cDNA or cDNA of the mouse ortholog of TMEM76 carrying a FLAG tag on the C-terminus or of a fusion protein of mouse TMEM76 with enhanced green fluorescent protein (EGFP). All constructs increased the *N*-acetyltransferase activity in the mutant fibroblast cells to approximately normal level (fig. 6*A*). Significant increase in activity was also observed in transfected COS-7 cells, confirming that the TMEM76 protein

Table 2. TMEM76 Predicted Mutations in Probands from 30 Families with MPS IIIC

Patient Group and Mutation 1	Mutation 2	No. of Patients	Geographic Origin of Patient(s)
Patients from Czech families:			
p.I373SfsX3	p.R534X	1	Czech Republic
p.L349X	p.M510K	1	Czech Republic
p.F313X	p.R412X	1	Czech Republic
p.R372H	p.P599L	1	Czech Republic
Patients homozygous for *TMEM76* mutations:			
p.[D68VfsX19; P265Q]	p.[D68VfsX19; P265Q]	3	Morocco, Morocco, and Spain
p.[P193HfsX20; K551Q]	p.[P193HfsX20; K551Q]	1	France
p.P311L	p.P311L	1	United Kingdom
p.W344X	p.W344X	1	France
p.R372C	p.R372C	1	United Kingdom
p.R412X	p.R412X	2	Turkey and Poland
p.[W431C; A643T]	p.[W431C; A643T]	1	France
p.G452S	p.G452S	1	Canada
p.E499K	p.E499K	1	Canada
p.S567NfsX14	p.S567NfsX14	1	Turkey
Patients compound heterozygous for *TMEM76* mutations:			
p.C104F	…	1	Belarus
p.E499K	p.D590V	1	France
p.P193HfsX20	p.R412X	1	Canada
p.P311L	p.R372C	1	France
p.R412X	…	1	Poland
p.R412X	p.G446X	1	Poland
p.S569L	…	2	France and Portugal
p.S569L	p.L69EfsX32	1	United States
p.V488GfsX22	p.S569L	1	Poland
p.V612SfsX16	…	1	Finland
Families with no mutations identified to date	…	2	North Africa and Portugal

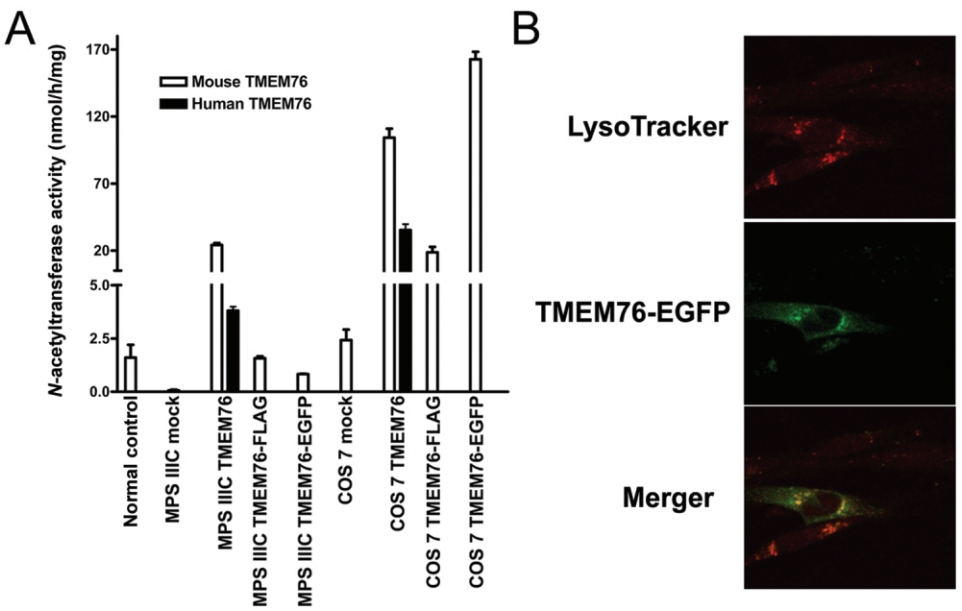


Figure 6. Functional expression of human and mouse TMEM76 protein. *A,* The full-size human and mouse *TMEM76* coding sequences subcloned into pCMV-Script, pCMV-Tag4A, and pEGFP-N3 vectors were expressed in COS-7 cells and in cultured skin fibroblasts from a patient with MPS IIIC. The cells were harvested 48 h after transfection, and *N*-acetyltransferase activity was measured in the homogenates of TMEM76-transfected and mock-transfected fibroblast or COS-7 cells by use of the artificial fluorometric substrate 4-methylumbelliferyl-β-D-glucosaminide.[23] Values represent means $\pm$ SD of four independent experiments. *B,* The intracellular localization of TMEM76 was studied by expressing the fusion protein of the mouse TMEM76 with EGFP. Before fixation, the cells were treated for 45 min with 50 nM lysosomal marker, LysoTracker Red DND-99 dye. Slides were analyzed on an LMS 510 Meta confocal microscope (Zeiss). Magnification $\times$1000. The image was randomly selected from 30 studied panels, all of which showed a similar localization of TMEM76-EGFP. The fluorescence of EGFP was not quenched as it would have been if the fluorophore had been exposed to the acidic lysosomal microenvironment, confirming that the C-terminus of TMEM76 faces the cytoplasmic side of the lysosomal membrane.

by itself has *N*-acetyltransferase activity. Confocal fluorescent microscopy shows that TMEM76-EGFP (fig. 6*B*) or TMEM-FLAG (not shown) peptides are targeted in human fibroblasts to cytoplasmic organelles, colocalizing with the lysosomal-endosomal marker LysoTracker Red.

Discussion

Degradation of heparan sulfate occurs within the lysosomes by the concerted action of a group of at least eight enzymes: four sulfatases, three exo-glycosydases, and one *N*-acetyltransferase, which work sequentially at the terminus of heparan sulfate chains, producing free sulfate and monosaccharides. The inherited deficiencies of four enzymes involved in the degradation of heparan sulfate cause four subtypes of MPS III: MPS IIIA (heparan *N*-sulfatase deficiency [MIM 252900]), MPS IIIB (α-*N*-acetylglucosaminidase deficiency [MIM 252920]), MPS IIIC (acetyl-CoA:α-glucosaminide acetyltransferase deficiency), and MPS IIID (*N*-acetylglucosamine 6-sulfatase deficiency [MIM 252940]). Since the clinical phenotypes of all these disorders are similar, precise diagnosis relies on the determination of enzymatic activities in patients' cultured skin fibroblasts or leukocytes. The biochemical defect in MPS IIIC was identified 30 years ago as a deficiency of an enzyme that transfers an acetyl group from cytoplasmically derived acetyl-CoA to terminal α-glucosamine residues of heparan sulfate within the lysosomes, resulting in the accumulation of heparan sulfate. Therefore, for identification of the molecular basis of this disorder, we used two complementary approaches. First, we performed a partial purification of human and mouse lysosomal *N*-acetyltransferase, which suggested that the enzyme has properties of an oligomeric transmembrane glycoprotein, with an ~100-kDa polypeptide containing the enzyme active site.[31] Second, by linkage analysis, we narrowed the locus for MPC IIIC to a 2.6 cM-interval (*D8S1051–D8S1831*) and, third, compared the level of transcripts of the genes present in the candidate region between normal control cells and those from patients with MPS IIIC. Thus, an integrated bioinformatic search and gene-expression analysis both pinpointed a single gene, *TMEM76,* which encodes a 73-kDa protein with predicted multiple transmembrane domains and glycosylation sites. DNA mutation analysis showed that patients with MPS IIIC harbor *TMEM76* mutations incompatible with the normal function of the predicted protein, whereas expression of human *TMEM76* and the mouse ortholog proved that the protein has *N*-acetyltransferase activity and lysosomal lo-

calization, providing evidence that *TMEM76* is the gene that codes for the lysosomal *N*-acetyltransferase.

The TMEM76 protein does not show a structural similarity to any known prokaryotic or eukaryotic *N*-acetyltransferases or to other lysosomal proteins, on the basis of sequence homology searches. Thus, we think that it belongs to a new structural class of proteins capable of transporting the activated acetyl residues across the cell membrane. Moreover, TMEM76 shares homology with a conserved family of bacterial proteins COG4299 (uncharacterized protein conserved in bacteria) (Entrez Gene GeneID 138050). All 146 members of this family are predicted proteins from diverse bacterial species, including Proteobacteria, Cyanobacteria, and Deinococci. Since many of these bacteria are capable of synthesizing heparan sulfate and other structurally related glycosaminoglycans and perform reactions of transmembrane acetylation, it is tempting to speculate that this activity may also be performed by the proteins of the COG4299 family. Previous studies suggested two contradictory mechanisms of transmembrane acetylation. Bame and Rome[2,36,37] proposed that it is performed via a ping-pong mechanism. First, the acetyl group of acetyl-CoA is transferred to an His residue in the active site of the enzyme. This induces a conformational change that results in the translocation of the protein domain containing the acetylated residue to the lysosome, where the acetyl residue is transferred to the glucosamine residue of heparan sulfate. In contrast, Meikle et al.[38] were unable to demonstrate any specific acetylation of the lysosomal membranes and proposed an alternative mechanism that involved the formation of a tertiary complex of the enzyme, acetyl-CoA, and heparan sufate. Identification of *N*-acetyltransferase as a 73-kDa protein with multiple transmembrane domains, together with our previous data that showed that *N*-acetyltransferase is acetylated by [^{14}C]acetyl-CoA in the absence of glucosamine,[31] strongly supports the ping-pong mechanism of transmembrane acetylation.

For 23 of the 30 probands included in this study for mutation analysis, *TMEM76* mutations were identified in both alleles. Five probands were heterozygous for a missense mutation, with a second mutation yet to be identified. In two probands from North Africa and Portugal, we did not identify any mutations in the coding regions or immediate flanking regions. These patients are homozygous for the microsatellite markers throughout the entire MPS IIIC locus and may be homozygous for a yet-to-be-identified *TMEM76* mutation; however, we cannot formally exclude defects in other genes. Additional studies have been initiated to search for mutations in the introns and promoter regions. The patients with MPS IIIC with the identified frameshift and nonsense mutations all have a clinically severe early-onset form. The almost complete deficiency of *N*-acetyltransferase activity in cultured skin fibroblasts from these patients is consistent with the predicted protein truncations and/or nonsense-mediated mRNA decay. Further expression studies are necessary to confirm the impact of the identified substitutions of the conserved amino acids on enzyme activity. Nevertheless, the identification of the lysosomal *N*-acetyltransferase gene which, when mutated, accounts for the molecular defect in patients with MPS IIIC sets the stage for DNA-based diagnosis and genotype-phenotype correlation studies and marks the end of the gene-discovery phase for lysosomal genetic enzymopathies.

Acknowledgments

We thank the patients, their families, and the Czech Society for Mucopolysaccharidosis, for participating in our study, and members of the sequencing and genotyping facilities at the McGill University and Genome Quebec Innovation Centre, for their technical support. We also acknowledge Nina Gusina, Joe Clarke, and Tony Rupar, for providing cell lines from patients with MPS IIIC; Mila Ashmarina, Milan Elleder, J. Loredo-Osti, and Johanna Rommens, for helpful discussions; Karine Landry, for technical support; and Maryssa Canuel, for help with confocal microscopy. The Montreal study was supported by operating grants from the Sanfilippo Children's Research Foundation (to A.V.P.) and by the Canadian Networks of Centres of Excellence Program—the Mathematics of Information Technology and Complex System network (to K.M.). The Prague study was supported by grants NR8069-1 and 1A/8239-3 from the Grant Agency of the Ministry of Health of the Czech Republic. Institutional support was provided by Ministry of Education of the Czech Republic grant MSM0021620806. A.V.P. is a National Investigator of the Fonds de la Recherche en Santé du Québec.

Appendix A

Table A1. Exon-Flanking Primers Used for PCR Amplification and Sequencing of the Exons in the Human *TMEM76* Gene

Primer	Sequence (5′→3′)
TMEM76_Exon1_F	CTCCCGAAGACAAACACTCC
TMEM76_Exon1_R	GCGAAGTCGCAGCAACAGC
TMEM76_Exon2_F	AAGCTTTTGAGAAGCACTACTGG
TMEM76_Exon2_R	GAAGGGCTTTAGACATGAGAGC
TMEM76_Exon3_F	GGAAAAGTCATGTCAGGATCTCC
TMEM76_Exon3_R	GAATAATACATGTTCCTGGGTACG
TMEM76_Exon4_F	TTATTCTGCCTCCATGATATTAGC
TMEM76_Exon4_R	CTACAGAAAGCGTCATGGACTGC
TMEM76_Exon5_F	GGAAATTCAGCATGAGAATATAGG
TMEM76_Exon5_R	GCCACTTGAGGGTGACAGC
TMEM76_Exon6_F	GAATATGAGCTTTAATTTTATTTCC
TMEM76_Exon6_R	TTAGGAATACGGGAGCTACAACC
TMEM76_Exon7_F	CAAAATGAAATTTACCCCTTAGC
TMEM76_Exon7_R	ACATCCAAGAAATCCTTCCTAGC
TMEM76_Exon8_F	CCTTCCTTTTCACATAGCAAACC
TMEM76_Exon8_R	GCTCTGTGAAGGACGTATATAAGC
TMEM76_Exon9_F	CCCCTGGGTTTACTTTCTATACC
TMEM76_Exon9_R	CCAGCATCATCTGAAAAACAGG
TMEM76_Exon10_F	GGGGCTATATTCTGAACTCTTCC
TMEM76_Exon10_R	ACCTGAGATGGAGGAATTGC
TMEM76_Exon11_F	CTGGGATGAGAGGAGAAGTCC
TMEM76_Exon11_R	ACTTGAAGCCAGGAGTGAGG
TMEM76_Exon12_F	CCTTCTATTTGCATTTAGTTCACC
TMEM76_Exon12_R	GAGAATTCCTCTGACTCGAGACC
TMEM76_Exon13_F	TTTTATTCTTGTCCCTCTGTTCG
TMEM76_Exon13_R	CACTTCTGAAAGCCTGAGTTCC
TMEM76_Exon14_F	TTGGTCTAGGAGCTGTTTGTACG
TMEM76_Exon14_R	CCATAGCACAAGAGAGAATATGC
TMEM76_Exon15_F	TCTTTGTCAGGTAGTTAAGACAGTGG
TMEM76_Exon15_R	GTGAAGGAAAGGAATTTTAGC
TMEM76_Exon16_F	ACAAGTTTCAGCCCTCTCTACG
TMEM76_Exon16_R	GTGGAGGAGACGTTTCAGTGC
TMEM76_Exon17_F	ATGCTGAAATTGGATTTGTTCC
TMEM76_Exon17_R	ACCAAGGATGCTCCAGAGG
TMEM76_Exon18_F	AGTAGCCAACAATGGAAGTGC
TMEM76_Exon18_R	GAGCCGTGTCACAGTTAACC

NOTE.—For bidirectional sequencing on the ALFexpress DNA sequencer, all primers have the universal overhang synthesized on the 5′ end (AATACGACTCACTATAG for forward [F] primers and CAGGAAACAGCTATGAC for reverse [R] primers).

Web Resources

Accession numbers and URLs for data presented herein are as follows:

BLAST, http://www.ncbi.nlm.nih.gov/blast/ (used to identify ortholog protein sequences)

Entrez Gene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene (for GeneID 138050)

GenBank, http://www.ncbi.nih.gov/Genbank/ (for accession numbers AK152926.1, AK149883.1, DR000652.1, XM_372038.4, NT_007995.14, XP_539948.2, XP_588978.2, XP_341451.2, and XP_519741.1)

Human Genome Variation Society, http://www.hgvs.org/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for MPS IIIA, IIIB, IIIC, and IID)

The R Project for Statistical Computing, http://www.r-project.org/

References

1. Rome LH, Hill DF, Bame KJ, Crain LR (1983) Utilization of exogenously added acetyl coenzyme A by intact isolated lysosomes. J Biol Chem 258:3006–3011
2. Bame KJ, Rome LH (1985) Acetyl-coenzyme A:α-glucosaminide *N*-acetyltransferase: evidence for a transmembrane acetylation mechanism. J Biol Chem 260:11293–11299
3. Pohlmann R, Klein U, Fromme HG, von Figura K (1981) Localisation of acetyl-CoA: α-glucosaminide *N*-acetyltransferase in microsomes and lysosomes of rat liver. Hoppe Seylers Z Physiol Chem 362:1199–1207
4. Hopwood JJ, Freeman C, Clements PR, Stein R, Miller AL (1983) Cellular location of N-acetyltransferase activities toward glucosamine and glucosamine-6-phosphate in cultured human skin fibroblasts. Biochem Int 6:823–830
5. Meikle PJ, Whittle AM, Hopwood JJ (1995) Human acetyl-coenzyme A:α-glucosaminide *N*-acetyltransferase: kinetic characterization and mechanistic interpretation. Biochem J 308:327–333
6. Kresse H, von Figura K, Bartsocas C (1976) Clinical and biochemical findings in a family with Sanfilippo disease, type C. Clin Genet 10:364
7. Bartsocas C, Grobe H, van de Kamp JJ, von Figura K, Kresse H, Klein U, Giesberts MA (1979) Sanfilippo type C disease: clinical findings in four patients with a new variant of mucopolysaccharidosis III. Eur J Pediatr 130:251–258
8. Klein U, Kresse H, von Figura K (1978) Sanfilippo syndrome type C: deficiency of acetyl-CoA:α-glucosaminide *N*-acetyltransferase in skin fibroblasts. Proc Natl Acad Sci USA 75: 5185–5189
9. Sanfilippo SJ, Podosin R, Langer LO Jr, Good RA (1963) Mental retardation associated with acid mucopolysacchariduria (heparitin sulfate type). J Pediatr 63:837–838
10. Klein U, van de Kamp JJP, von Figura K, Pohlmann R (1981) Sanfilippo syndrome type C: assay for acetyl-CoA:α-glucosaminide *N*-acetyltransferase in leukocytes for detection of homozygous and heterozygous individuals. Clin Genet 20:55–59
11. Meikle PJ, Hopwood JJ, Clague AE, Carey WF (1999) Prevalence of lysosomal storage disorders. JAMA 281:249–254
12. Pinto R, Caseiro C, Lemos M, Lopes L, Fontes A, Ribeiro H, Pinto E, Silva E, Rocha S, Marcao A, Ribeiro I, Lacerda L, Ribeiro G, Amaral O, Sa Miranda MC (2004) Prevalence of lysosomal storage diseases in Portugal. Eur J Hum Genet 12: 87–92
13. Poorthuis BJ, Wevers RA, Kleijer WJ, Groener JE, de Jong JG, van Weely S, Niezen-Koning KE, van Diggelen OP (1999) The frequency of lysosomal storage diseases in The Netherlands. Hum Genet 105:151–156
14. Zaremba J, Kleijer WJ, Juijmans JG, Poorthuis B, Fidzianska E, Glogowska I (1992) Chromosomes 14 and 21 as possible candidates for mapping the gene for Sanfilippo disease type IIIC. J Med Genet 29:514
15. Ausseil J, Loredo-Osti JC, Verner A, Darmond-Zwaig C, Maire I, Poorthuis B, van Diggelen OP, Hudson TJ, Fujiwara TM, Morgan K, Pshezhetsky AV (2004) Localization of a gene for mucopolysaccharidosis IIIC to chromosome region 8p11-8q11. J Med Genet 41:941–945

16. Seyrantepe V, Tihy F, Pshezhetsky AV (2006) The microcell-mediated transfer of human chromosome 8 restores the deficient *N*-acetylytransferase activity in skin fibroblasts of mucopolysaccharidosis type IIIC patients. Hum Genet 120:293–296
17. Kong X, Murphy K, Raj T, He C, White PS, Matise TC (2004) A combined linkage-physical map of the human genome. Am J Hum Genet 75:1143–1148
18. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, Mai CP, Nguyen TH, Nguyen NB, Pham XK, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, Van De Vosse E, Hudson TJ, Abel L, Schurr E (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427:636–640
19. Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. Am J Hum Genet 58:1323–1337
20. Sobel E, Papp JC, Lange K (2002) Detection and integration of genotyping errors in statistical genetics. Am J Hum Genet 70:496–508
21. Hodanova K, Majewski J, Kublova M, Vyletal P, Kalbacova M, Stiburkova B, Hulkova H, Chagnon YC, Lanouette CM, Marinaki A, Fryns JP, Venkat-Raman G, Kmoch S (2005) Mapping of a new candidate locus for uromodulin-associated kidney disease (UAKD) to chromosome 1q41. Kidney Int 68:1472–1482
22. O'Connell JR, Weeks DE (1998) PedCheck: a program for identifying genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266
23. Voznyi YV, Karpova EA, Dudukina TV, Tsvetkova IV, Boer AM, Janse HC, van Diggelen OP (1993). A fluorimetric enzyme assay for the diagnosis of Sanfilippo disease C (MPS III C). J Inher Metab Dis 16:465–472
24. Gudbjartsson DF, Jonasson K, Frigge M, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13
25. Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds) Bioinformatics and computational biology solutions using R and Bioconductor. Springer, New York, pp 397–420
26. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:article 3
27. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300
28. Kmoch S, Hartmannova H, Stiburkova B, Krijt J, Zikanova M, Sebesta I (2000) Human adenylosuccinate lyase (ADSL), cloning and characterization of full-length cDNA and its isoform, gene structure and molecular basis for ADSL deficiency in six patients. Hum Mol Genet 9:1501–1513
29. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254
30. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, et al (2006) The UCSC Genome Browser Database: update 2006. Nucleic Acids Res 34:D590–D598
31. Ausseil J, Landry K, Seyrantepe V, Trudel S, Mazur A, Lapointe F, Pshezhetsky AV (2006) An acetylated 120-kDa lysosomal transmembrane protein is absent from mucopolysaccharidosis IIIC fibroblasts: a candidate molecule for MPS IIIC. Mol Genet Metab 87:22–31
32. Kahsay RY, Gao G, Liao L (2005) An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. Bioinformatics 21:1853–1858
33. Jensen LJ, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Nielsen H, Staerfeldt HH, Rapacki K, Workman C, Andersen CA, Knudsen S, Krogh A, Valencia A, Brunak S (2002) Prediction of human protein function from post-translational modifications and localization features. J Mol Biol 319:1257–1265
34. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4:1633–1649
35. Bonifacino JS, Traub LM (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem 72:395–447
36. Bame KJ, Rome LH (1986a) Acetyl-coenzyme A:α-glucosaminide *N*-acetyltransferase: evidence for an active site histidine residue. J Biol Chem 261:10127–10132
37. Bame KJ, Rome LH (1986b) Genetic evidence for transmembrane acetylation by lysosomes. Science 233:1087–1089
38. Meikle PJ, Whittle AM, Hopwood JJ (1995) Human acetyl-coenzyme A:α-glucosaminide *N*-acetyltransferase: kinetic characterization and mechanistic interpretation. Biochem J 308:327–333

Appended submitted manuscript

Members of glycoside hydrolases family 31 in *Caenorhabditis elegans*: Search for the nematode's acid alpha–glucosidase

Jana Uřinovská*[1], Jakub Sikora*[§1], Helena Poupětová[1], Filip Majer[1], Jitka Hlavatá[1], Karel Jelínek[2], Marta Kostrouchová[3], Jana Ledvinová[1], Martin Hřebíček[§1]

* These authors contributed equally to this work
§ Corresponding authors

Institutional addresses:

[1] Charles University in Prague, 1st Faculty of Medicine, Institute of Inherited Metabolic Disorders, Ke Karlovu 2, Prague 2, Czech Republic

[2] Charles University in Prague, Faculty of Science, Department of Physical and Macromolecular Chemistry, Albertov 6, Prague 2, Czech Republic.
current address: Mathematical and Statistical Projects, Podskalska 14, Prague 2, Czech Republic

[3] Charles University in Prague, 1st Faculty of Medicine, Institute of Inherited Metabolic Disorders, Laboratory of Molecular Biology and Genetics

e-mail addresses:
JU: jhujo@lf1.cuni.cz
JS: jakub.sikora@lf1.cuni.cz
HP: helena.poupetova@lf1.cuni.cz
JH: jhlava@lf1.cuni.cz
FM: fmaje@lf1.cuni.cz
KJ: kjelinek@ikj.cz
MK: marta.kostrouchova@lf1.cuni.cz
MH: mhreb@lf1.cuni.cz

Abstract

Background

Human acid alpha-glucosidase (GAA, EC 3.2.1.20) which catalyses lysosomal glycogen degradation is a member of glycoside hydrolase family 31 (GH31). Hereditary deficiency of GAA results in lysosomal glycogen storage disease type II (GSDII). The aim of this study was to assess GH31 proteins in *Caenorhabditis elegans* (*C. elegans*), to identify the GAA ortholog with acidic pH optimum and to evaluate *C. elegans* as a model organism for GSDII utilizing RNA mediated interference (RNAi).

Results

Searches for GAA orthologs in the *C. elegans* genome revealed four acid alpha-glucosidase related (*aagr-1-4*) genes. These predicted ORFs were amplified from mixed stage N2 cultures cDNA and sequenced including 5´ and 3´ untranslated regions.

Multiple protein sequence alignment of GH31 members and AAGR-1-4 demonstrated their evolutionary conservation which was further used for homology modeling based on available GH31 protein structures. Phylogenetic analyses suggested clustering of AAGR-1 and 2 with acid-active, and AAGR-3 and 4 with neutral pH optimum enzymes.

Transcriptional GFP tagging was used to evaluate spatial and temporal expression patterns of AAGR-1 and AAGR-2. The expression of both genes was limited to intestinal cells and six coelomocytes.

RNAi of each of the *aagrs* did not reveal any morphological phenotype. The RNAi effect was assessed by GAA activity measurements at two pH values (4.0 and 6.5) with and without addition of inhibitor acarbose (pH 4.0). We observed predominant neutral or acidic glucosidase activities corresponding to individual AAGRs as well as variable influence of acarbose on their enzymatic activity. AAGR-1 was found to possess acidic glucosidase

activity with relatively pronounced acarbose resistance, a result further replicated in *aagr-1* deletion mutant.

Conclusions

We analysed *C. elegans* (AAGR-1-4) orthologs of human GAA by combination of bioinformatic, cellular and biochemical approaches and determined that even though AAGR-2 has prevalent acid and AAGR-3 predominant neutral glucosidase activity. Nevertheless, AAGR-1 is the least acarbose sensitive acidic AAGR and thus represents the most probable ortholog of human GAA. Even though, separate RNAi assays did not provide sufficient enzymatic expression knock-down for efficient modeling of lysosomal glycogen storage, the study rendered additional insights into function of GH31 proteins.

Background

Human acid alpha-glucosidase (GAA) is a well-characterized lysosomal enzyme belonging to glycoside hydrolases family 31 (GH31, clan D). This family of proteins includes a number of confirmed or predicted hydrolases from different prokaryotic and eukaryotic organisms [1-3] and has remote but significant homologies with glycoside hydrolases from other clans [4]. GH31 enzymes represent a wide set of proteins with diverse substrate specificities and functions. The common property of enzymes belonging to GH31 family is their ability to cleave terminal carbohydrate moieties from di-, oligo- or polysaccharides of variable size. The hydrolysis of glycosidic bond by GH31 enzymes is exerted via acid/base-catalyzed retention mechanism with glycoside-enzyme intermediate. Two aspartic residues, in the otherwise conserved GH31 active site [5], have been previously identified as the active nucleophile and acid/base catalyst [6-8]. The amino acid sequence (Wi<u>D</u>MnE) surrounding the active site Asp nucleophile (underlined) constitutes an evolutionary signature of GH31

family. Even though conserved, variances in this particular sequence stretch rendered bioinformatic and phylogenetic basis for further stratification of GH31 hydrolases into four additional subgroups [9]. Proteins orthologous to GAA have been identified in many animal and plant species and also in bacteria [1, 10]. There are at least four other human GH31 proteins besides GAA [1, 11]: dual catalytic sucrase-isomaltase (SUIS) and maltase-glucoamylase (MGA) that evolved by duplication [11, 12]; and alpha-glucosidase AB (GANAB) and alpha-glucosidase C (GANC). These four enzymes differ in their pH activity optima – acid (GAA, SUIS, MGA) *vs.* neutral (GANAB, GANC) [10] .

Protein crystallographic structural data are currently available for three GH31 proteins – bacterial YicI [9] from *Escherichia coli*, archeal MalA [13] from *Sulfolobus solfataricus* and N-terminal subunit of human maltase-glucoamylase (NtMGA) [14]. Besides providing valuable insights into molecular structure, catalytic function and structural evolution of GH31 proteins, these data (YicI and MalA) have already been used as structural templates for homology modeling of other GH31 proteins including catalytic domain of human GAA [9, 15, 16]. It is of importance to note that a subset of NtMGA crystallographic data was obtained in the presence of a potent artificial MGA inhibitor (acarbose) which is also routinely used in the diagnostic assays for GAA deficiency (see below) [17, 18].

Glycogen storage disease type II (GSDII, Pompe disease), is a human autosomal recessive lysosomal storage disorder resulting from the enzymatic deficiency of GAA. Its deficiency causes lysosomal accumulation of glycogen with normal structure in a wide set of tissues. Phenotypic presentation, related mainly to the varying degree of skeletal and cardiac myopathy and hepatopathy, spans a range of clinical phenotypes from severe early-onset to slowly progressive adult-onset forms. Animal models currently available for the study of GSDII are either naturally occurring variants or were established by targeted disruption of

GAA gene [10]. *Caenorhabditis elegans* (*C. elegans*) has not, so far, been considered as a model organism for GSDII.

Evolutionary conservation of the nematode's genomic sequence [19], relative ease of genetic manipulation and cultivation allowed its use as a relevant model for different human diseases including those resulting from malfunction of the late endosomal/lysosomal membrane proteins (for recent review see de Voer *et al.* (2008) [20]). In addition, there are numerous reports demonstrating the importance of the nematode's glycogen intermediary carbohydrate metabolism [21, 22] - a fact that supports its suitability for GAA studies including GSDII modeling.

Based on these prerequisites, our work focused on the predicted GAA orthologs (acid alpha-glucosidase related genes – *aagr*) in *C. elegans* genome and evaluated them by bioinformatic and phylogenetic means including homology modeling of their tertiary structures. DNA sequences of *aagr* genes were verified by sequencing on the cDNA level and were further studied for their expression patterns and biological roles during the nematode's life cycle by means of green fluorescent protein (GFP) tagging and gene specific RNA mediated interference (RNAi). In addition, we managed to acquire deletion mutants of two *aagr* genes via *C. elegans* Gene Knockout Consortium [23] and compared their biochemical phenotype with findings observed after RNAi.

Results and discussion

BLAST searches, cDNA amplification and sequencing

Repetitive BLAST [24] searches for orthologs of human GAA in the Wormpep database revealed five predicted coding sequences or sequences annotated as confirmed by partially overlapping EST clones - D2096.3 (*aagr-1*), R05F9.12 (*aagr-2*), F40F9.6 (*aagr-3*), F52D1.1 (*aagr-4*) and F53F4.8 (*aagr-5*). E-values against human GAA for each of the found sequences were as follows: 499 e –141 (AAGR-1), 493 e-139 (AAGR-2), 288 e-77 (AAGR-3)

and 334 e-91 (AAGR-4). E-values and overall similarity of the amino acid sequences with the human GAA sequence supported their annotated status [25] as GH31 proteins. No predicted or confirmed double active site proteins comparable to human SUIS or MGA were found in the *C. elegans* genome, identical number and annotation of proteins present in *C.elegans* GH31 is available via CAZy database inteface [1].

In order to evaluate the available *in silico* predictions, mRNAs (cDNAs) of *aagr* 1-4 genes were amplified in overlapping PCR fragments including 5´*trans* spliced ends and 3´untranslated regions including polyadenylation sites. Number of EST clones was noted in the critical genomic regions and was taken into account for PCR primer design. Gene organization (exon/intron boundaries) and the lengths of all four transcripts and deriving proteins were found identical to the available database predictions (see Figure 1).

The sequence analyses surprisingly revealed SL1 *trans* splicing of all four *aagr* genes although *aagr-1* lies in a close downstream vicinity of the transcriptional operon CEOP4284 [26] (intergenic region ~ 650bp, see Figure 5). We therefore explored cDNA sequence and *trans* splicing of the gene D2096.12 located immediately upstream of *aagr-1* and found that it is SL2 *trans* spliced and that its sequence completely agreed with the available prediction [25]. We searched the available protein databases for potential orthologs of D2096.12. The highest scoring ortholog was *Caenorhabditis briggsae* gene of unknown function CBG18464 (E-value 6.79e-102), other significantly less scoring BLAST hits were also mostly gene products of unknown or only inferred function. The first gene in the CEOP4284 operon is UDP- glucuronic acid decarboxylase (*sqv-1*) [27] which is SL1 *trans* spliced. CEOP4284 operon and *aagr-1* thus represent local clustering of a gene involved in glycosaminoglycan neogenesis (*sqv-1*) and a GH31 member protein (*aagr-1*) involved in carbohydrate catabolism.

We were able to detect only a single splicing variant of *aagr-3* corresponding to F40F9.6b [25], despite the additional annotated alternative splicing variant - F40F9.6a.

It was further not possible to amplify any PCR product for the predicted gene F53F4.8 (*aagr-5*) despite extensive amplification efforts and relatively high E-value (426 e-119). We therefore propose that F53F4.8 represents a pseudogene (in compliance with recent change of this Wormbase entry [25]) and therefore we excluded it from our additional studies.

We performed signal sequence predictions by SignalP [28, 29] server which suggested N-terminal signal sequences for all AAGRs with high probability (AAGR-1 – 0.996, AAGR-2 – 1.0, AAGR-3 – 0.911, AAGR-4 – 0.999). Signal sequence cleavage sites and thus the overall lengths of signal sequences (see Figure 3) were predicted between residues 22-23 (AAGR-1), 18-19 (AAGR-2), 19-20 or 22-23 (AAGR-3) and 16-17 (AAGR-4). Additional predictions of intracellular protein targeting by TargetP server [29, 30] proposed all four AAGRs as non-mitochondrial / non-secretory proteins.

Multiple protein alignment and phylogenetic analyses

Multiple alignment of amino acid sequences of selected eukaryotic proteins from GH31 family and AAGRs (see Methods) revealed significant conservation of primary protein sequence in all aligned entries within the extent of GH31 module [11]. The alignment delineated two groups of proteins: those with predominating acidic pH activity optimum (GAAs, MGAs and SUISs) and those with predominating neutral pH activity optimum (GANABs and GANCs) [10]. AAGR-1 and 2 clearly segregated with the first and AAGR-3 and 4 with the second group of proteins, respectively. As expected, the most conserved sequence stretch was the immediate vicinity (Wi<u>D</u>MnE) of the active site Asp (underlined) nucleophile [5]. Even though variable in the WiDMnE sequence all AAGRs can be sub categorized into the previously introduced subgroup 1 of GH31 proteins [13] (see Additional

Figure 1) which comprises GH31 proteins with multitude of activities (xylosidases, alpha-glucosidases, glucoamylases, and sucrase-isomaltases).

This bioinformatic analysis further showed presence of unique insertions in the sequences of some of the AAGRs with either very low or no similarity to other aligned GH31 proteins. These insertions in AAGR-1 and 2 were located in the regions corresponding to structural inserts 1 and 2 of the catalytic catalytic$(\beta/\alpha)_8$ barrel [9, 13, 14] (see below). The proximal (N-terminally located) insertions in AAGR-1 and 2 lacked any similarity to sequences previously suggested to contribute to substrate specificity between NtMGA and CtMGA [14]. In addition, AAGR-3 and 4 did not contain, similarly to all GANC and GANAB proteins, sequence of the structural insert 2 located C-terminally from the WiDMnE sequence (see Additional Figure 1, Figure 3 and Homology modeling for further results and discussion).

Unrooted phylogenetic trees based on the three initial multiple alignments (see Methods) employing three different computation algorithms (neighbor joining, maximum parsimony and maximum likelihood) with prior bootstrapping and final tree evaluation were constructed to assess the evolutionary relationships among the sequences [31, 32]. Topology of the individual trees employing full length sequences did not substantially differ with the mathematical algorithm used. Tree validation based on bootstrapping showed satisfactory values for all the constructed cladograms. Tree organization was always in general agreement with accepted evolutionary concepts and confirmed and further extended prior assumptions that related AAGR-1 and AAGR-2 to acid pH functioning proteins and AAGR-3 and AAGR-4 to their neutral pH functioning counterparts (Figure 2 and Additional Figure 2).

Cladograms employing separated GH31 modules (as in *Naumov (2007)* [11]) and sequences corresponding to catalytic domains of GH31 proteins demonstrated closer evolutionary relationship of AAGR-1 and 2 to CtMGA and CtSUIS in comparison to

NtMGA, NtSUIS and GAA proteins (data not shown). These findings agree with previously reported close evolutionary relations of g1065946 (currently *aagr-1*) and g6425187 (currently pseudogene *aagr-5*) [12] - the only GH31 genes predicted in *C. elegans* genome at that time - to C-terminal subunits of dual catalytic GH31 proteins. AAGR-3 and 4 preserved their clustering to GANAB and GANC proteins in these alternatively prepared phylogenetic trees.

Homology modeling

At present, the structural data of three members of GH31 family are available - *E. coli* YicI [9] with predominating xylosidase activity, archeal *S. solfataricus* MalA [13] and human NtMGA in complex with acarbose inhibitor [14]. These proteins represent suitable template structures for AAGRs as they exhibit considerable primary structure identity and contain distinct structural GH31 motifs [9] (see Figure 3 and Additional Table 1).

All four models of AAGR proteins were individually built by Modeller 9 using multiple sequence alignment with all three templates (Figure 3). Resulting models were compared with templates using DaliLite server [33]. NtMGA was the best scoring template for all AAGRs. The final models shown in Figure 4 represent architectural folds sharing template (NtMGA) - target structural identity above 35% and also include loop insertions without sequence similarity (see below and Additional Table 3). The most conserved part of all AAGRs was the catalytic domain. When compared, the sequence identity of the catalytic domain between the NtMGA template and AAGR-1 and 2 was higher in comparison to AAGR-3 and 4 (Figure 3, Additional Table 1 and 3). AAGR-1 and 2 possess both previously described structural inserts (proximal-1/distal-2) in the catalytic $(\beta/\alpha)_8$ barrel domain present in NtMGA and MalA proteins. Moreover, each of these inserts in AAGR-1 and 2 contains unique additional insertions distant both from -1 and +1 sugar-binding subsites of the active site. On the contrary, AAGR-3 and 4 do not possess distal structural insert 2 (also absent in YicI) as well as the unique insertions present in AAGR-1 and 2. Furthermore, in comparison

to AAGR-1 and 2, the entire N-terminal domain of AAGR-3 and 4 shares less identity with all templates (Figures 3 and 4, Additional Figure 1, Additional Table 1) due to their clustering with GANAB and GANC proteins (see above). General fold of the proximal C-terminal domain is preserved in all AAGRs despite lower identity values in comparison to catalytic domain.

Superimposition of the models and templates displayed high conservation of the amino acids forming previously proposed GH31 active site demonstrating only few substitutions in AAGR-1-4. Nevertheless, it should be pointed out that due to the low identity of the N-terminal domain and the absence of distal structural insert in AAGR-3 and 4, the participation of residues from these protein regions in the formation of the active site is highly uncertain using the available templates (see Figure 4 and Additional Table 2). Therefore we focused on the GH31 active site residues present in all template and model molecules. The most apparent and also difference in residues closest to -1 and +1 substrate sugar units between templates and models was tyrosine (Y299) of NtMGA aligning to Y184 in MalA and Phe 277 in YicI, a residue previously suggested to cause low alpha-glucosidase activity of YicI. These residues structurally correspond to Y304 in AAGR-2 but are substituted by bulkier tryptophan (W316, W361 and W353) in AAGR-1, 3 and 4. Furthermore, only these active site residues represent nonrotamer conformation when acarbose is inserted into the active site.

To obtain a better insight on the impact of this substitution on ligand binding, docking of acarviosine (terminal pseudo-disaccharide moiety of the pseudo-tetrasaccharide acarbose) into both AAGR-1 and 2 was performed using Autodock 4 [34]. Structural properties such as pocket shape of the active site (Figure 4C) and rigidity of the inhibitor (acarbose) represent good prerequisites for this type of analysis. Acarviosine was selected because it represents the interacting part of acarbose with the active site as previously evidenced in acarbose NtMGA

crystallographic complex. All rotatable bonds in acarviosine and side chains of residues W316 (AAGR-1) or Y304 (AAGR-2) were made flexible. Repeated runs of Autodock gave consistent results. Increasing the number of energy evaluations (see Methods) showed superior RMSD cluster convergence for AAGR-2 over AAGR-1 (data not shown). However, the lowest estimated binding energy for acarviosine complexes were close for both models, with values -9.57 kcal/mol corresponding to inhibition constant (Ki) 96 nM and -10.5 kcal/mol (Ki = 26 nM) for AAGR-1 and 2, respectively. These values characterizing the docked states suggest predominant contribution of conserved intermolecular interactions and impact of additional atoms surrounding the active site on the resulting affinity potential of the models. Furthermore, necessity of torsion changes of side chain of the half buried bulky tryptophan during the binding of the inhibitor may affect enzyme characteristics and may have been reflected by inferior clustering convergence.

Predicted best docked positions of acarviosine in the active site pockets of either AAGR-1 or AAGR-2 were similar (RMSD = 0.97 and 1.03 Å) to the original crystallographic position of acarviosine in NtMGA inhibitor complex (Figure 4). Moreover, the same intermolecular interactions with the exception of bonding through solvating water molecules (not included in the models) were observed. These docking experiments provide additional support for RNAi experiments (see below) which evaluate acarbose sensitivity of AAGR-1 in comparison to AAGR-2. It is worth noting that the presence of aromatic residues at this position represents a conserved feature in GH31 enzymes, moreover proteins with previously reported acarbose sensitivity contain tyrosine residue, whereas those with less pronounced acarbose sensitivity (GAA and neutral glucosidases – not shown in the presented protein alignment Figures) possess tryptophan residue at this position in the active site [17].

Additional subtle differences (see Figure 4C) in the amino acid composition of the -1 and +1 sugar binding subsites were observed in all four AAGRs. Relatively distant

substitutions (~6 Å) in the binding cavity of AAGR-3 and 4 are tyrosines (Y398 and Y390) whereas NtMGA and AAGR-1 and 2 contain aspartates at this position. At the edge of the binding pocket is a glycine (G602) in NtMGA template, serine (S652) in AAGR-1 and threonine (T647) in AAGR-2, and AAGR-3 and 4 contain histidines (H638 and H630). Contribution of these more distant residues to direct enzyme substrate interactions can be expected to be much less in all modeled AAGRs.

In this perspective, the use of rationally designed [35] or other described GH31 substrates/inhibitors (e.g. analogs of salacinol and kotalanol [36]) and methods like site directed mutagenesis would further elucidate the enzyme specificity and physiological functions of the nematode's AAGRs. We demonstrated advantageous use of lately released protein structure of NtMGA over MalA or YicI structures for metazoan GH31 protein homology modeling. We therefore propose preferential use of NtMGA structural data while modeling human GAA (data not shown) in situations like prediction of structural impact of GSDII pathogenic variations or docking of substances suitable for enzyme enhancement therapy (performed previously using MalA and YicI templates [15, 16]. Even though not reported yet, it is plausible to speculate about potential misinterpretations (pseudo deficiency) of GAA activity assay with acarbose inhibition in GSDII diagnostics caused by pathogenic substitutions in the presumed GAA active site residues (for their list [37] including tryptophan (W376) see Additional Table 2).

Biochemical studies

Standard bacteria/nematode separation protocol based on sucrose flotation was omitted [38] due to the known effects of sucrose on the lysosomal physiology [39]. We therefore resorted to repetitive washes of the nematode cultures as described in Methods.

We found detectable GAA activities in normal *C. elegans* N2 strain homogenates when measured at two different pH values (4.0 and 6.5) and two different incubation

temperatures (25°C and 37°C) with methylumbelliferyl (MU) substrates. The values measured at acidic pH by this protocol reflect cumulative enzymatic activities of several enzymes that can be expected in *C. elegans*. Based on human studies, it is especially MGA that interferes with direct GAA measurements.

The overall values of alpha-glucosidase activity measured at 25°C were 53 and 67.5 nmol $mg^{-1}h^{-1}$ for pH values 4.0 and 6.5, respectively. For activity values measured at 37°C see Tables 1-3.

Addition of acarbose in the 3-9 μM range to the reaction mixture at acidic pH values completely inhibits MGA activity without affecting GAA activity. The effect of acarbose on the neutral glucosidases has also been previously shown as insignificant [17, 18]. Values measured with acarbose at acidic pH should thus directly reflect GAA activity. The robustness of this simple assay modification has already resulted in its introduction into the clinical diagnostics of GSDII including GAA activity measurements in dried blood spots [40]. Impact of addition of acarbose into glucosidase assays is demonstrated in Tables 1-3. Approximately 68-72% of alpha-glucosidase activity (pH 4.0) can be inhibited by addition of acarbose in the control samples. Values of glucosidase activity measured at pH 6.5 were virtually unaffected by addition of acarbose into enzymatic assay.

We assessed alpha-glucosidase activities after selective RNAi experiments and in deletion mutant strains in order to further discriminate and stratify all four AAGR proteins.

RNA mediated interference and the effects of acarbose on glucosidase activities

We performed separate selective RNA interference experiments (see Methods) of all four *aagr* genes to evaluate the impact of RNAi on the nematodes´ phenotype and alpha-glucosidase activity profiles both in the presence and absence of acarbose inhibitor. The selection of the sequence regions for cloning into the double RNA polymerase bearing vector

was performed on the basis of cDNA multiple sequence alignment of all four *aagrs*. The overall identity in the alignment was 14.1% but the similarity was over 90% (due to nucleotide sequence alignment).

We tried to implement low similarity/identity regions for RNAi assays to avoid or at least minimize cross-interference and thus possible off-target RNAi effects between otherwise homologous sequences [41]. The selected cDNA sequence stretches (for details see Methods) had following identity values in comparison to the rest of the multiple sequence alignment: *aagr-1*- 25.2%; *aagr-2*- 10.7%; *aagr-3*- 12.3%; *aagr-4*- 11.6%. It is worth noting, that the differences in enzymatic activity values measured in individual experiments are not surprising due to the previously reported variability in the RNAi efficiency [42].

We were not able to observe any morphological or other observable phenotype after RNAi of any of the *aagr* genes despite extensive and very detailed examination of the animals. We even assayed D2096.12, upstream gene from *aagr-1* formerly assigned to same operon (see above), but also with negative results. This fact was not surprising with respect to our previous findings [38] of minimal RNAi phenotype of other lysosomal hydrolase ortholog in *C. elegans* - *gana-1*. These findings most probably reflect high residual enzymatic activity (see later) after RNAi.

We therefore resorted to biochemical evaluation of the impact of RNAi treatment due to the absence of morphological changes. Measurement of GAA activities at neutral and acidic pH (6.5 and 4.0) and parallel acarbose inhibition showed separation of the neutral and acid alpha-glucosidase activities to different AAGRs (see Tables 1 and 2). In general, we were able to demonstrate a general decrease of alpha-glucosidase activity after RNAi of *aagr-1-3.* The decrease of alpha-glucosidase activities in *aagr-4* RNAi experiments was not that prominent and with respect to analyses performed in the later described mutant strain RB1307 harboring *aagr-4* (*ok1423*) deletion, it can be hypothesized that AAGR-4 enzymatic potential

is limited. The most pronounced selective changes in activity values were found for *aagr-2* and *aagr-3* RNAi experiments, with the drop of acidic and neutral values to the levels of 20% of controls, respectively. RNAi of *aagr-1* decreased GAA activity values more dramatically in the acidic ranges - average value 68% of controls; compared to neutral values - 86% of control values (both average values calculated from Table 2). These changes were less pronounced when compared to *aagr-2* and *aagr-3*, despite the inclusion of active nucleophile surroundings and the highest identity values (25%) of the sequence used for *aagr-1* RNAi.

Influence of acarbose on alpha-glucosidase activity in the individual RNAi experiments was variable and suggested that AAGR-1 is the least acarbose sensitive acidic AAGR (for summary see Table 2). This result was futher supported by the analysis of mutant strain RB1790 carrying *aagr-1* deletion (*ok2317* - see below). On the basis of these observations we assume that AAGR-1 is the most probable ortholog of acid alpha-glucosidase in *C.elegans*.

Isolation and characterization of *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) deletion mutants

Both deletion mutants, *aagr-1 – ok2317* (RB 1790 strain) and *aagr-4 – ok1423* (RB 1307 strain), were generated by the *C. elegans* Gene Knock-out Consortium [23]. We obtained only these two mutants out of the four *aagr* genes at the time of the manuscript's submission, generation of *aagr-2* and *aagr- 3* deletion mutants is a pending request. The extent of the deletions were analyzed according to the PCR primer data provided by the Consortium.

The sequence analysis of gDNA of *aagr-1* (*ok2317*) deletion mutant revealed 1162 bp long deletion resulting in the loss of 224 C-terminal base pairs of exon 9, introns/exons 9, 10, 11 and 65 base pairs from intron 12 (Figure 1).

Aagr-1 deletion corresponds to the loss of 218 amino acids of the AAGR-1 amino acid sequence. The deletion covers more than half of the predicted catalytic domain (205 missing amino acids), including α4-helix and final 4 βα repetitions in the $(\beta\alpha)_8$ - core barrel according to the multiple amino acid sequence alignment (Figure 3) and structural model of AAGR-1 (Figure 4).

We did not observe any abnormal morphological phenotype despite thorough microscopic analysis of *ok2317* worms. To evaluate the impact of the deletion on the enzymatic activity of this predicted acid-active glucosidase, we performed GAA activity measurements at pH values 4.0 and 6.5 and parallel acarbose inhibition in the homogenates of mutated and standard N2 Bristol strain. These measurements demonstrated considerable activity decrease at acidic pH in comparison to standard N2 Bristol strain and sensitivity of the residual acidic activity to the acarbose (Table 3). The results are comparable with the values acquired by RNAi assays of *aagr-1* (see Tables 1 and 2). As noted above, these experiments further support AAGR-1 as the least acarbose sensitive acidic AAGR and thus as the most probable ortholog of human acid α-glucosidase. Unfortunately, *aagr-1* deletion mutant does not clearly replicate GSD type II cellular phenotype in terms of lysosomal storage and autophagic buildup [43]. We explain this finding by the existence of the second enzyme with acidic α-glucosidase activity in *C. elegans* (AAGR-2), which may compensate the deficient acidic glucosidase activity of *AAGR-1*. Our results from RNAi of *aagr-1* and *aagr-2* and biochemical analysis of the *aagr-1* deletion mutant strain demonstrate that contribution of *aagr*-2 to the overall acidic glucosidase activity is proportionally greater in comparison to *aagr-1*. We may speculate that the extent of the overall substrate turnover in *C. elegans* and duration of nematode's life cycle is short to allow buildup of full blown sub-cellular lysosomal storage phenotype. We also hypothesize that the double knockout of both

C. elegans enzymes with acidic GAA activity (AAGR-1 and 2) can presumably provide a suitable model of GSD type II in *C.elegans*.

The sequence analysis of the *aagr-4* deletion mutant (*ok1423*) revealed 1832 bp long deletion resulting in the loss of 421 C-terminal base pairs of exon 4, complete intron 4 and exon 5 and 45 base pairs from intron 5 on the level of the gDNA (Figure 1). This deletion comprises 281 amino acids from the conceptual translation of *aagr-4* gene, including the first 6 βα repetitions in the $\beta_8\alpha_8$ - core barrel of the catalytic domain according to the multiple amino acid sequence alignment (Figure 3) and structural model of AAGR-4 (Figure 4). Similarly to *aagr-1* deletion mutant, detailed microscopic analysis of the nematodes carrying *ok1423* deletion did not reveal any abnormal morphological or other observable phenotype.

In order to evaluate the impact of the deletion on the enzymatic activity of this predicted neutral activity glycoside hydrolase we performed GAA activity measurements at pH values 4.0 and 6.5. The measured GAA activity in the homogenates of back-crossed nematodes was compared to the results from the standard N2 Bristol strain. Surprisingly, we were not able to detect any significant activity decrease in *aagr-4* deletion mutant when measured at both pH levels (Table 3).

The complementing contribution of the second predicted neutral-active enzyme AAGR-3 to the global neutral GAA activity or minor enzymatic activity of AAGR-4 is highly probable, based on these results. RNAi assays against *aagr-3* in *aagr-4* deletion mutant were performed to evaluate this hypothesis.

The activity values measured at neutral pH reached the values acquired by the above described RNAi experiments for AAGR-3 alone. Activity values measured at acidic pH values did not differ from the controls (Table 3).

This described combined approach allows us to conclude that the dominant contribution to the neutral GAA activity originates from AAGR-3 and that AAGR-4 provides

only a minor portion of the total neutral GAA activity in *C.elegans*. Our assumptions are based on the expected profound impact of both deletions (*ok2317* and *ok1423*) both on mRNA processing or protein function.

Transcriptional GFP fusions of *aagr-1* and *aagr-2* genes

We prepared several transcriptional GFP fusion constructs to evaluate the expression patterns of the two most probable orthologs (*aagr-1* and *aagr-2*) of human GAA. In addition, we further evaluated differential expression patterns of the upstream genes of *aagr-1* (D2096.12 and *sqv-1*) organized into CEOP4284 operon. GFP tagging was used for its well-characterized properties and relative ease of evaluation (see Methods) [44].

Aagr-1

The GFP expression pattern of the pJS8 construct (1340 bp upstream of *aagr-1* ATG) was observed in membrane endowed vacuoles of 6 coelomocytes and also in the intestinal cells. There was no other GFP positive tissue found in a number of transgenic worms. The expression was observed only in L4 and adult nematode stages (Figure 5C). We were not able to detect GFP expression in any other developmental stage of the nematodes. It is worth noting that the differentiation of the specific GFP signal from the extensive intestinal autofluorescence in the elderly worms was based on the spectral fluorescence foot printing (for details see Methods). The use of spectrally resolving confocal microscope and subsequent linear unmixing of reference spectra provided unambiguous resolution of GFP emission and contaminating autofluorescence signal and thus effectively overcame methodological limitations of band-pass fluorescence emission signal filtering [45]. Based on these findings we can conclude that *aagr-1* possesses its own promoter region despite relatively short intergenic sequence (see above and Figure 5).

The regulatory sequence of CEOP4284 operon was divided into three fragments for the purposes of evaluation of the expression patterns on the length of the DNA regulatory sequence (for details about construct structure see Methods). Construct pJS1 and pJS4 covered *sqv-1*/D2096.12 intergenic region and 3kb of regulatory sequence upstream of D2096.12 ATG, respectively. These two constructs showed two different expression patterns: pJS1 – anal sphincter; and pJS4 – intestinal cells and pharyngeal muscles (data not shown). The temporal expression patterns of pJS1 and 4 were identical to pJS6 (see below).

Construct pJS6 including 5 kb of D2096 genomic region upstream of D2096.12 ATG revealed homogeneous GFP expression in intestinal cells, coelomocytes, pharyngeal and body wall muscle cells, pharyngeal and rectal glands, epidermal cells and rectal sphincter. We also noted expression in a population of head neurons. The expression was detectable during all developmental larval stages of the nematodes starting at the 50 cell embryonic stage. See Figure 6A/B for schematic representation of the cloning strategy and GFP expression patterns.

Aagr-2

The expression pattern of the pJS9 construct (3431 bp upstream of *aagr-2* ATG initiation codon) was similar to that of *aagr-1* (construct pJS8 – see above). GFP signal was observed in membrane endowed vacuoles of 6 coelomocytes and also in the pseudocoelomic cavity. The pseudocoelomic signal most probably resulted from the coelomocyte exocytosis of GFP. There was no other GFP positive tissue found in number of transgenic worms. Temporally the expression was observed in L4 and adult nematode stages (Figure 6D), earlier temporal expression was not observed, even after repetitive and extensive evaluation.

Similarly to *Aagr-1*, *Aagr-2* expression was not observed in body wall muscle cells.

Aagr-3

Expression pattern of *aagr-3* was evaluated as part of a high throughput project analyzing gene expression patterns in *C.elegans* [46]. The expression is broad and includes a number of tissues such as intestine, neurons, pharynx and others (for details see Wormbase [25])

Conclusions

There are at least four genes in *C. elegans* genome belonging to GH31 family that are homologous to human GAA. None of the four genes showed dual catalytic properties as observed in higher organisms, suggesting both low level of substrate specificity and the later evolutionary occurrence of the duplication/ fusion event. Simple bioinformatic tools (e.g BLASTP search) cannot discriminate between the substrate specificity and/or pH optima of these four gene products. We were able to delineate two genes coding proteins with preferential acid activity (*aagr-1* and *2*) and additional two with preferential neutral activity (*aagr-3* and *4*) based on the phylogenetic and homology modeling data. However, the analysis of the *aagr-4* deletion mutant disclosed this gene product as a minor contributor to the overall glucosidase activity.

RNA interference data showed selective drops of pH dependant glucosidase activity pointing to AAGR-2 and AAGR-3 as the dominant contributors of acidic and neutral glucosidase activity in *C. elegans*, respectively. Docking of acarbose into modeled structures and measurements of GAA activities with parallel acarbose inhibition suggest AAGR-1 as the least acarbose sensitive acidic AAGR and so as the most probable ortholog of human acid α-glucosidase. This finding was furher supported by the analysis of *aagr-1* deletion mutant. RNA interference approach did not efficiently generate structural storage changes in *C.*

elegans. We base this finding on the high residual enzymatic activities after RNA interference treatment. Nevertheless, our study demonstrates several interesting aspects of glycoside hydrolase family 31 protein function that may be reflected in GSDII diagnostics and possibly in design of substances suitable for enzyme enhancement therapy of this disorder.

We hypothesize that double deletion mutant of both AAGR-1 and 2 could be a promising model for GSD type II and may provide important information about advisability of *C. elegans* as a model organism for this human lysosomal storage disorder.

Methods

C. elegans methods and strains

The wild type Bristol N2 strain was used for all experiments and was handled under standard laboratory conditions [47]. Standard methods were used for DNA and dsRNA synthesis [48]. The nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center [49], which is funded by the NIH National Center for Research Resources (NCRR). The names of the genes (AAGR) were consulted and approved by Dr. Jonathan Hodgkin (CGC Genetic Map and Nomenclature Curator) prior to the submission of the manuscript.

BLAST search

Wormbase (2002-2008 consecutive versions, and freeze versions WS100 – WS190) databases [25] were repeatedly searched for orthologs of human GAA using BLASTP program set at default values [24]. E-values given in the results correspond to WS188 and Wormpep databases. Amino acid sequence of human GAA (acc.no. NP_000143) [50] was used as a query sequence in all searches.

cDNA amplification, sequencing and signal sequence predictions

Total RNA was isolated from mixed stage N2 cultures [51] and reverse transcribed with oligodT-T7 (5´- AATACGACTCACTATAG(A19)) primer and Superscript reverse transcriptase (Invitrogen). Intragenic D2096.12, D2096.3 (*aagr-1*), R05F9.12 (*aagr-2*), F40F9.6 (*aagr-3*) and F52D1.1 (*aagr-4*) coding sequences were PCR amplified in 1.2-1.5kb partially overlapping fragments with primers designed according to available Wormbase [25] and Worfdb [52] databases (individual primer sequences can be provided by the authors on request). SL1 and SL2 primers (5´-GGTTTAATTACCCAAGTTTGAG, 5´-GGTTTTAACCCAGTTACTCAAG) with gene specific primers were used to amplify 5´ UTR of the cDNAs and to evaluate the *trans* splicing mode of the particular mRNA; 3´UTRs were amplified with T7 and gene specific primers to characterize the polyadenylation sites.

The individual PCR products were cloned using TOPO TA cloning kit (Invitrogen) into the pCR 2.1 vector. Positive clones were sequenced using Li-Cor automated fluorescent sequencer; resultant sequences were aligned with reference cosmid sequences in the AlignIR software (Li-Cor) to evaluate splicing boundaries and overall gene organization.

Signal sequences of AAGR-1-4 were predicted using SignalP server [28, 29] and intracellular protein targeting of all four AAGRs was predicted by TargetP [29, 30] server. Predictions (SignalP and TargetP) were done under predefined default values.

Multiple alignment and phylogenetic analyses

The following amino acid sequences of selected eukaryotic GH31 proteins listed in the CAZy database [1] were downloaded from SwissProt/TrEMBL database [53] and were used for multiple alignment and consecutive phylogenetic analyses - *Homo sapiens* GAA (P10253), SUIS (P14410), MGA (O43451), GANAB (Q14697), GANC (Q8TET4); *Bos taurus* GAA (Q9MYM4); *Mus musculus* GAA (P70699); *Coturnix japonica* GAA (O73626); *Oryctolagus cuniculus* SUIS (P07768); *Suncus murinus* SUIS (O62653); *Rattus norvegicus*

SUIS (P23739); *Sus scrofa* GANAB (P79403); *Drosophila melanogaster* GAA-like (Q7KMM4); and confirmed conceptual translations of *Caenorhabditis elegans* AAGR-1-4. The sequences were selected on the basis of the status of their annotation, number of predicted GH31 protein sequences [11] was omitted from these analyses.

The bootstrap values used for tree validation were chosen as 1000 for distance matrix and maximum parsimony methods and 100 for maximum likelihood methods.

Initial multiple alignment for phylogenetic tree construction was performed using ClustalW algorithm [54] and Blosum62 matrix (gap penalty -10, gap extension penalty 0.1) and was visually checked and edited. Three different multiple alignments were evaluated for comparative reasons: one including the entire protein sequences, one with separated GH31 modules [11] of dual activity proteins (SUIS, MGA) and one employing sequences of the GH31 catalytic domain based on YicI, MalA and NtMGA structural protein data [9, 13, 14].

PHYLIP version 3.57c software package was used for distance matrix and maximum parsimony algorithm based methods and version 3.6 for maximum likelihood algorithm methods [32]. Alignments were bootstrapped using SEQBOOT program prior to all analyses. Final consensus trees were constructed by CONSENSE program.

Bootstrapped data were used for PROTDIST using Dayhoff PAM 001 matrix in case of distance matrix methods. Tree construction was performed in NEIGHBOR program separately employing Neighbor-Joining and UPGMA algorithms.

PROTPARS and ProML were used for maximum parsimony and maximum likelihood tree construction with the second algorithm employing Jones, Taylor and Thorton probability model [31].

Homology modeling and molecular docking

Multiple alignment (see Figure 3) including *E. coli* YicI [9], *S. solfataricus* MalA [13] and *H. sapiens* NtMGA [14] and all four *C.elegans* AAGRs and human GAA (data shown) was prepared using ClustalW algorithm [54] and Blosum62 matrix.

Structural models of AAGR-1-4 were generated using Modeller 9 [55] program which implements comparative protein structure modeling by satisfaction of spatial restraints. All these three available crystallographic structures of GH31 proteins (*H. sapiens* NtMGA PDB entry:2QMJ, *E. coli* YicI 2XSJ and *S. solfataricus* MalA 2G3M) were used as input structural templates. Unstructured loop regions (sequence insertions not present in the template structures) were energy refined and the final loop arrangements were selected both on the basis of Discrete Optimized Protein Energy (DOPE) score and visual check. Quality of the models was assessed by PROCHECK [56] and MolProbity [57] programs, individual structural domains of the template and target structures were pairwise aligned and analysed using DaliLite server [33].

Docking of acarviosine was conducted in Autogrid 4 (generation of atom-specific affinity, electrostatic and desolvation potential maps) and Autodock 4 [34] using Lamarckian genetic algorithm searching method and semiempirical free energy force field. Docking settings were set at default values, only the number of evaluations was increased to 5 000 000. Residues in the binding pocket and intermolecular interactions between the best scoring docked ligands and modeled proteins were compared and ascertained with the template crystallographic structures in Autodock.

RNA mediated intereference

Complete cDNA sequences of *aagr-1-4* were aligned using ClustalW algorithm [54] (match scoring 10, mismatch scoring 0, gap opening penalty 15, gap extension penalty 6.66)

in order to maximize the use of low similarity/identity regions between otherwise related sequences for RNAi protocols.

The following regions of cDNA were amplified by PCR and cloned into the TOPO TA pCR II vector (adenine from ATG initiation codon being nucleotide +1): 1500-2226 bp (*aagr-1*), 1-1161 bp (*aagr-2*), 1925-3054 bp (*aagr-3*) and 55-799 bp (*aagr-4*). Re-sequenced cDNA clones were further transferred into L4440 double promoter vector in the following manner. The *aagr-1* clone was digested from the TOPO TA pCR II vector with *EcoRI* (dual digestion), *aagr-2* and *3* with *AvaI*/*HindIII* and *aagr-4* with *AvaI* (dual digestion). Gel purified fragments were than ligated into L4440 vector linearized by the particular restriction digestions. The constructs for RNAi feeding experiments were further used for transfections into HT115 *E. coli* strain and production of specific dsRNA was induced by IPTG (0,4 mM).

N2 Bristol or back-crossed *aagr-4* deletion mutant (*ok1423*) worms were fed on HT115 *E. coli* carrying L4440 interference construct and worms fed on HT115 *E. coli* carrying L4440 plasmid without insert were used as controls. We performed four separate experiments for each *aagr* each evaluating 1500-2000 worms.

F_1 progeny was screened for morphological phenotypes and biochemical evaluation by means of GAA activity measurements of the interfered worms was performed as described below.

Isolation and characterization of *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) deletion mutants

At our request, two deletion mutants – *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) were kindly generated by the *C. elegans* Gene Knock-out Consortium [23] as the only two from the four *aagr* genes at the time of the manuscript submission. Mutant nematodes were back-crossed according to established protocol with the N2 worms [58]. Worms were PCR screened for the presence of homozygous deletions with the following pairs of primers 5´-CCGGCAAGTAGATTGAGAGC and 5´-CAATATCACGCAGGTTGTGG (*ok1423*) and 5´-

GAT TTT TGC TCG TAG TCC CG and 5′-CGG TGC AAT CAT AAG AGC AG (*ok2317*) during the backcross. PCR products acquired with these primers were directly sequenced on the 3100-*Avant* Genetic Analyzer (Applied Biosystems) in order to evaluate the size and position of the deletions. The backcrossed worms were observed for morphological or otherwise observable phenotypes and the GAA enzyme activities were measured in these nematodes as described below.

RNAi feeding experiment using *aagr-3* construct was performed as described above using back-crossed *aagr-4* (*ok1423*) deletion mutant, GAA enzyme activities were assessed as described below.

Determination of α-glucosidase and β-hexosaminidase activities

Prior to all activity measurements worms were washed from culture plates and repeatedly rinsed/centrifuged in M9 buffer. 4- methylumbelliferyl (MU) -β-D-glucopyranoside (2,5 mM) in the McIlvaine buffer (0.1M citrate/0.2M phosphate buffer at pH 4.7) and 4-MU-α–D-glucopyranoside (1,26 mM) in the same buffer (pH 4.0 or 6.5) were used as enzyme substrates for β-hexosaminidase and GAA activity measurements, respectively. Selective inhibitor - acarbose (Serva) [17] was added to the reaction mixture in selected measurements to final concentration 8 μM (see Results and Discussion). Reaction mixtures (sample and enzyme substrate) were incubated at 37°C and reactions were terminated by addition of 600 μl of 0.2 M glycine/ NaOH buffer (pH 10.6). The activities were also determined at 25°C in normal N2 strain homogenates. Fluorescence signal of the released 4-methylumbelliferone was measured on the luminiscence spectrofotometer LS 50B (Perkin-Elmer) (excitation 365nm and emission 448nm). All measurements were performed in doublets.

Transcriptional GFP fusions of *aagr-1* and *aagr-2* genes

The 6 kb fragment spanning 459-6599 bp of cosmid D2096 [25] was PCR amplified using combination of KlenTaq (Ab Peptides) and DeepVent (New England Biolabs) DNA polymerases from the genomic DNA isolated from the mixed stage N2 nematode culture. Resultant PCR product was used as template for three second round PCR reactions with artificial 5´PstI restriction site introduced by sense primers aligning at 711-730 (used for construct pJS6), 2942-2962 (construct pJS4) and 5446-5463nt (construct pJS1) of the D2096 cosmid sequence and universal antisense primer aligning at 6220-6241 nt of the D2096 cosmid with *BamH*I artificial restriction site. Acquired PCR products 796bp (pJS1), 3300bp (pJS4) and 5531 bp (pJS6) were cloned into pCR-XL-TOPO vector using TOPO XL PCR cloning kit (Invitrogen). Inserted PCR fragments were cleaved by sequential *Pst*I and *BamH*I digestions and ligated into *Pst*I-*BamH*I linearised pPD95.75 vector (supplied by A. Fire, Stanford School of Medicine, Stanford University, United States) in-frame with open reading frame of GFP. pJS1 includes regulatory sequence between the *sqv-1* and D2096.12 genes of CEOP4284 operon. pJS4 and pJS6 include 2969 and 5200 bp upstream of D2096.12 ATG, respectively.

Regulatory sequence of *aagr-1* (D2096.3) in the CEOP4284 operon was cloned into pPD95.75 GFP vector in the following fashion. 3720bp fragment spanning 6242-9961nt of D2096 was PCR amplified with primer containing C-terminal *SmaI* restriction site. This PCR product (3720bp) was cloned into pCR-XL-TOPO vector. The resultant construct was further digested by *SmaI* and *HindIII* to obtain 2106bp long fragment containing approximately 1340bp of *aagr-1* ATG upstream regulatory sequence and complete *aagr-1* intron 1. This fragment was ligated into *SmaI/HindIII* linearized pPD95.75 vector, the resultant construct was named pJS8.

The 3862 bp DNA fragment (*aagr-2*) spanning overlapping R05F9 and ZK546 cosmids was PCR amplified as described above and used as a template for second round PCR

with 40533-40553 nt (R05F9 cosmid) and 11999-12019nt (ZK546 cosmid) primers with artificial 5′ *BamH*I restriction sites. Insertion of the resultant PCR product (3684 bp) into the pCR-XL-TOPO vector was followed by *BamH*I digestion and in-frame re-cloning to *BamH*I linearised pPD95.75 vector containing open reading frame of GFP. This resultant construct – designated pJS9 – included 3431 bp of aagr-2 upstream regulatory sequence.

All extra chromosomal expression arrays were generated by co-injecting plasmid DNA (individual constructs) along with the marker plasmid pRF4 (*su1006*) at the final total concentration of 100ng/μl in to the germinal syncytium of young adult hermaphrodite worms [44]. The transgenic progeny was screened for GFP signal under Olympus SZX12 epifluorescent stereo microscope equipped with appropriate selective filter blocks. Final images were acquired on Nikon TE2000E C1*si* confocal microscope[59]. Excitation of GFP was done by 488 nm laser line. The spectral cross-talk of specific GFP signal and autofluorescence emission of granules in intestinal cells was eliminated by single value decomposition / linear unmixing algorithm embedded in the microscope's firmware (Nikon). Reference spectra for unmixing were acquired from the samples with only a single spectrum emission (GFP or autofluorescence, see Figure 5).

List of abbreviations

AAGR - acid alpha-glucosidase related; BLAST – basic local alignment and search tool; bp – base pair; cDNA – complementary DNA; CGC - *Caenorhabditis* Genetics Center; Ct – C-terminal; DOPE – discrete optimized protein energy; DNJ – 1-deoxynojirinomycin; dsRNA – double stranded RNA; EST – expressed sequence tags; GAA – acid alpha-glucosidase; GANAB – neutral alpha-glucosidase AB; GANC - neutral alpha-glucosidase C; gDNA – genomic DNA; GFP – green fluorescent protein; GH31 - glycoside hydrolase family 31; GSD II – glycogen storage disease type II; Ki – inhibition constant; MGA - maltase-glucoamylase; mRNA – messenger RNA; MU – methylumbelliferyl; Nt – N-terminal; ORF –

open reading frame; PCR – polymerase chain reaction; RNAi – RNA interference; RMSD – root mean square deviation; SL – splice leader RNA; SUIS - sucrase-isomaltase; UPGMA - unweighted pair group method with arithmetic mean

Authors' contributions

JU carried out RNAi, some expression experiments and enzyme activity measurements and participated on the manuscript preparation, JS performed bioinformatics and participated on homology modeling, cDNA sequencing, participated in designing the RNAi and expression experiments and prepared the manuscript including the final version, HP, JH, and JL participated on the biochemical analyses, FM performed homology modeling and docking analysis, KJ participated on homology modeling, MK advised the work on *C.elegans* and provided the laboratory support for *C.elegans* experiments, MH conceived the project, provided fundraising and participated on the preparation of the final version of the manuscript.

Acknowledgements

This work was supported from research project 0021620806 from Ministry of Education, Youth and Sports of the Czech Republic.

Figure legends

Figure 1 – *aagr-1-4* gene structure

Schematic representations of **(A)** *aagr-1*, **(B)** *aagr-2*, **(C)** *aagr-3* and **(D)** *aagr-4* gene structure. The size of each of the genes is given in base pairs on the level of genomic DNA as well as mRNA/cDNA (in the brackets). The values include ATG initiation codon and the lengths of 3´ untranslated regions (values on the right side). The green rectangles below **(A)** *aagr-1* and **(D)** *aagr-4* schematically correspond to the extent of *ok2317* and *ok1423* deletions

in the genomic DNA. For additional details about organization and genomic context of *aagr-1-4* see Wormbase [25]

Figure 2 – Cladogram of AAGRs and selected eukarytic GH31 proteins

Cladogram of AAGR-1-4 and other selected GH31 proteins computed by maximum likelihood algorithm (see Methods). The values at the branch nodes represent bootstrap values (maximum 100). GH31 modules [11] of dual catalytic proteins (SUIS and MGA) were not separated for the purpose of tree construction. The topology of the cladogram demonstrates the separation of the proteins into two groups (predominantly acidic vs. neutral enzymes) and separation of AAGR-1+2 and AAGR-3+4 with these groups.

Note the bootstrap values at the AAGR-1/2 branch (52) and D. melanogaster GAA-like (54) and compare the position of these proteins in cladograms computed by UPGMA and maximum parsimony algorithms (Additional Figure 2).

Figure 3 – Multiple sequence alignment of *C. elegans* AAGRs and *E. coli* YicI, *S. solfataricus* MalA and *H. sapiens* NtMGA

Multiple alignment of *E. coli* YicI [9], *S. solfataricus* MalA [13], *H. sapiens* NtMGA [14] and four *C. elegans* AAGR proteins which was used for homology modeling of their tertiary structures. The extent of structural elements $(\beta\alpha)_8$ present in the catalytic domain of NtMGA (structural inserts included) is shown above the alignment. Side chains of amino acid residues in the boxes correspond to those previously suggested to form -1 and +1 sugar binding sites (YicI, MalA and NtMGA), in AAGR proteins boxed residues correspond to those whose side chains lie with the diameter of ~ 5 Å of the docked acarvosine (for comparative summary see Additional Table 2). Sequences in brown (AAGR-1 and 2) correspond to insertions within the structural inserts (compare with Additional Figure 1 and Figure 4). Note the absence of structural insert 2 in YicI and AAGR3 and 4.

Residues in violet correspond to predicted signal sequences. Underlined sequences in AAGR-1 and 4 mark the extent of deletions in *ok2317* and *ok1423*, respectively (compare with small insets in Figure 4).

Primary amino acid sequences of YicI, MalA and NtMGA were taken from the particular structural .pdb files (see Methods). AAGR sequences are putative translations of the experimental cDNA sequencing data.

Identical residues are in yellow, conserved residues are in red and blocks of similar amino acid residues are in green color.

Figure 4 – AAGR-1-4 protein models

(A) Overview of the best scoring template - NtMGA [14] and structural models of AAGR1-4 in ribbon representation with secondary structure elements. Structural inserts in 1 and 2 are shown in blue and insertions in AAGR-1 and 2 are in purple red. Acarbose (yellow) was inserted by superimposition from NtMGA structure. Note that only the parts most conserved with template molecules are shown for all four AAGRs (see Additional Table 1). The extent of deletions *ok2317* and *ok1423* in AAGR-1 and AAGR-4 models, respectively, are shown in green on the small inset images.

(B) Binding active site pocket of NtMGA, AAGR-1 and AAGR-2 represented as solvent accessible surface. Contributions of both structural inserts are colored in blue, the substitution Tyr/Trp (discussed in the text) is in green and docked acarvosine is in yellow.

(C) Differences in active site of NtMGA and AAGR models. Models were superposed with the NtMGA template. Side chains of active site amino acids conserved in all models and NtMGA are shown in light orange for template only (for comparison see Additional Table 2). Non-equivalent residues are coloured according to the annotation in the picture. Molecular surface corresponds to NtMGA residues. An arrow in the top view image indicates a viewing direction for the side view image.

N-terminal domain of AAGR-3 and 4 is not shown in this Figure and the distal structural insert 2 in the catalytic domain of these two proteins is reduced to a short loop (is not present, see Figure 3 and Additional Tables 1 and 3).

Images were prepared in PyMol [60].

Figure 5 – Transcriptional GFP constructs of *aagr-1* and *aagr-2* with the corresponding GFP expression

(A) Schematic representation of the extent of the regulatory sequences of *aagr-1* and *aagr-2* used for creating transcriptional GFP constructs pJS1 (796 bp), pJS4 (3300 bp), pJS6 (5531 bp), pJS8 (2106 bp) and pJS9 (3684 bp).

(B) Broad GFP expression in the transgenic worm carrying the pJS6 construct in intestinal cells, coelomocytes, pharyngeal muscle cells and head epidermal cells.

(C) GFP expression limited to coelomocytes (arrowheads) in the transgenic worm carrying the pJS8 construct. The expression can also be observed in intestinal cells. Non-specific intestinal autofluorescence and GFP signal were discriminated by linear unmixing.

(D) GFP expression in the transgenic worm carrying the pJS9 construct limited to coelomocytes (arrowheads).

Scale bars in all images represent 40 μm.

Tables legends

Table 1 – Residual alpha-glucosidase activities in selected individual RNAi experiments against *aagrs* - influence of acarbose inhibitor (8 μM)

Table shows variable impact of acarbose inhibition on residual alpha-glucosidase activities measured at two different pH values (4.0 and 6.5) after selected separate *aagr-1-4* RNAi treatment of N2 worms in comparison to the controls (shown as % of controls). Note the differences in residual activity measured both in the presence and absence of acarbose in

RNAi experiments against *aagr-1* and *2*. The high residual activity in RNAi against *aagr-1* in the absence of acarbose (58%) reflects the contribution of *aagr-2*. The effect of acarbose on neutral activities (AAGR-3, 4) was much less pronounced. Compare to Table 2.
All measurements were done at 37°C.

Table 2 – residual alpha-glucosidase activities after separate *aagr-1-4* RNAi treatments – influence of acarbose inhibitor (8 μM)

Table shows residual alpha-glucosidase activities measured at pH 4,0 after eight separate RNAi experiments with or without 8 μM acarbose in the reaction mixture. The last column demonstrates the influence of acarbose – the proportion of residual alpha-glucosidase activity inhibitable by acarbose after RNAi. The most prominent drop of residual alpha-glucosidase activity due to addition of acarbose was observed after RNAi of *aagr-1* due to inhibition of *aagr-2* (values marked by ◄). Compare to predominant contribution of *aagr-2* to the overall acidic activity and also its acarbose sensitivity demonstrated by values in the last column.
All measurements were done at 37°C.

Table 3 – residual alpha-glucosidase activities in *aagr-1* (*ok2317*) and *aagr-4* (*ok1423*) deletion mutants - influence of acarbose inhibitor (8 μM)

Table shows residual alpha-glucosidase activities measured in the deletion mutants of *aagr-1* and *aagr-4* compared to the control N2 strain. The last column demonstrates the influence of acarbose on the acid alpha-glucosidase activity of *aagr-1* mutants. The last two rows of the table show values of residual alpha-glucosidase activity measured after RNAi assay against *aagr-3* in *ok1423* strain (aagr-4 deletion) which are comparable to *aagr-3* RNAi assays in N2 worms. Compare to Tables 1 and 2.
All measurements were done at 37°C.

Additional files legends

Additional Figure 1 - Multiple protein sequence aligment of AAGRs with other glycoside hydrolase family 31 members

Part of the global multiple alignment of selected GH31 proteins with AAGR 1-4 covering the extent of structural insert 1, α3 helix, β4 sheet and structural insert 2 as described in the crystal structure of human NtMGA [14] (above the alignment). N-terminal and C-terminal GH31 modules [11] of MGA and SUIS proteins were included as separate alignment entries. Numbering of amino acid residues corresponds to the reference sequences listed in the Methods.

The alignment demonstrates two segregating groups of sequences corresponding well with previously reported enzymatic pH optima [10]. This pattern is preserved in the GH31 signature sequence (Wi/nDMnEp/v) surrounding the active Asp nucleophile of AAGR-1/ 2 and AAGR-3/4, respectively. The alignment further shows the differences in the presence and primary amino acid sequence of structural inserts 1 and 2 a feature that further differentiates AAGR-1,2 from AAGR-3,4. Sequence stretches previously suggested to contribute to substrate specificities between NtMGA and CtMGA (light blue) [14] are not present in any of the AAGR proteins. Sequence corresponding to structural insert 2 is not present in AAGR-3 and 4 as well as in any of GANC and GANAB proteins.

Compare the positions of the gaps introduced into this alignment and into the alignment shown in Figure 3; a feature that reflects different entry sequences. Color coding of conservation of individual residues corresponds to Figure 3.

Additional Figure 2 - Cladograms of AAGRs and selected eukarytic GH31 proteins (UPGMA and maximum parsimony algorithms)

(A) Unweighted Pair Group Method with Arithmetic Mean (UPGMA), **(B)** Maximum Parsimony. GH31 modules [11] of dual catalytic proteins (SUIS and MGA) were not

separated for the purpose of tree construction. The values at the branch nodes represent bootstrap values (maximum 100).

The topology of the cladograms is virtually identical between these two computational algorithms. The positions of AAGRs are slightly different to the maximum likelihood cladogram (Figure 3), nevertheless the separation into two groups is preserved.

Additional Table 1 – Comparison of the extent and presence of structural elements in GH31 protein structures and AAGR proteins

Table compares the extent and overall length of individual structural elements in the templates [9, 13, 14] and target molecules. Numbering, which corresponds to the multiple protein alignment presented in Figure 3, was additionally checked by overall inspection of structural conservation of the templates and target structures and was further used for pairwise DaliLite structural alignments (Additional Table 3).

Color coding of individual domains and structural features corresponds to Figure 4A. AAGR-1-4 model structures were N- and C-terminally cropped because of low template/target similarity at positions given by numbers in brackets. Italicized numbers provide the comparison of overall lengths of structural inserts in the catalytic domain and catalytic domain itself. Values for the "distal insert (2)" in YicI and AAGR-3 and 4 represent absence of the insert (see elsewhere).

aa – amino acid

Additional Table 2 – Summary of residues predicted to form -1 and +1 sugar binding subsites of AAGR proteins – comparison with template structures (YicI, MalA and NtMGA)

Summary of residues previously reported or expected to participate in the formation of the GH31 active site. Numbering corresponds to multiple protein alignment shown in Figure 3. Residues in blue originate from structural inserts, Tyr/Trp substitution

discussed in detail in the text is in green (coloring corresponds to Figure 4B). Grey residues represent subtle distant substitutions also discussed in the text.

Descriptive legend to individual residues as previously described [9, 13, 14] : [1] active site nucleophile; [2] acid/base catalysts; [3] residues lining the active site pocket; [4] residues explicitly mentioned as forming -1 sugar binding subsite; [5] residues forming +1 subsite; [6] active site residues making contacts through water molecules; [7] F277 suggested to cause low α-glucosidase activity of YicI; [8] predicted active site residues of human GAA (homology modeling – MalA template), identical residues were predicted to interact with 1-deoxynojirinomycin [15, 16].

Human GAA as well as pathogenic GSD II variations reported at these particular positions [37] were included for comparative reasons. Unfortunately, the severity scoring (*sev – severe, int - intermediary, unk - unknown*) of GAA pathogenic variations is based on GAA intracellular processing and not on the activity assessment or sensitivity to acarbose (for comments see text).

Additional Table 3 – Statistics of structural alignment of template and target molecules

Statistics of pairwise comparative structural alignments between individual domains and structural inserts in template and target molecules. Detailed comparison was performed only for NtMGA template. Distal C-terminal domain (NtMGA template) as well as non-catalytic domains for bacterial (YicI, MalA) templates were omitted due to low sequence similarity. The last column compares values of initial structural models of AAGR-1-4 catalytic domain with YicI [9] used as the only structural template prior the release of MalA [13] and NtMGA [14] structures.

The alignments were performed using DaliLite server [33], protein structures of template and model molecules were fragmented prior the analysis according to the values given in Additional Table 1.

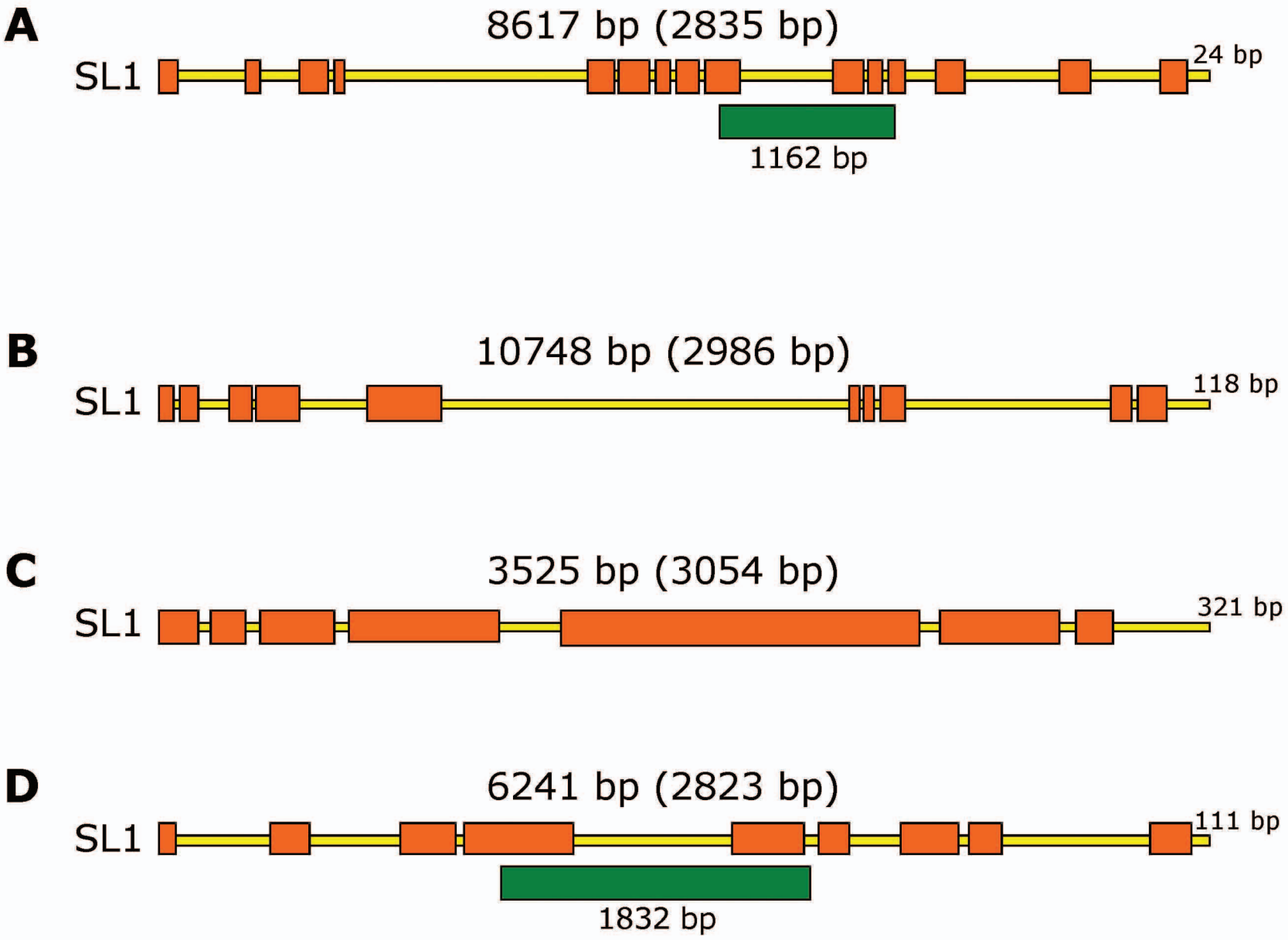
References

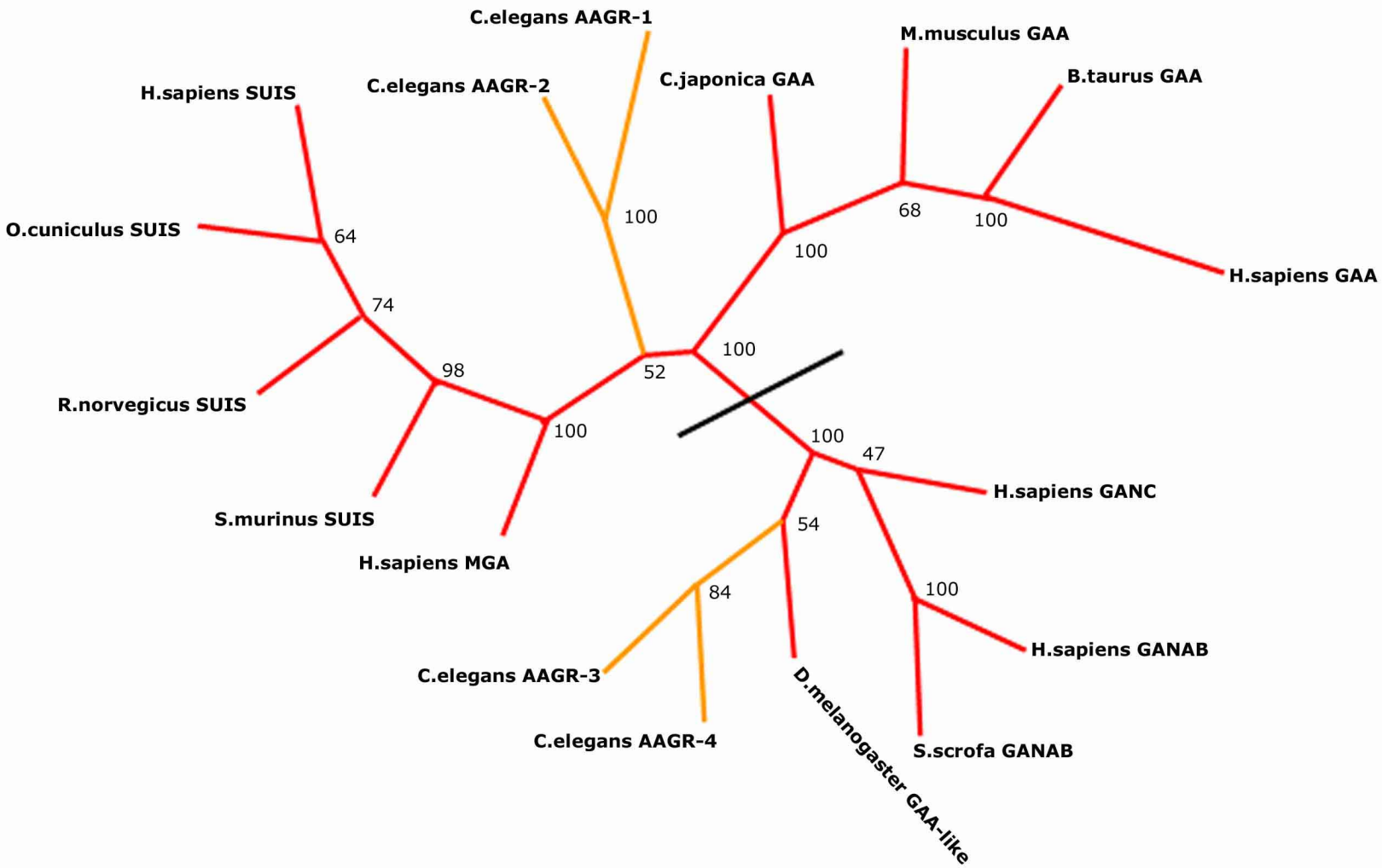
1. **Carbohydrate-Active Enzymes Server** [http://www.cazy.org]
2. **Protein families database of alignments and HMMs** [http://www.sanger.ac.uk/cgi-bin/Pfam/]
3. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities**. *The Biochemical journal* 1991, **280 (Pt 2)**:309-316.
4. Janecek S, Svensson B, MacGregor EA: **A remote but significant sequence homology between glycoside hydrolase clan GH-H and family GH31**. *FEBS Lett* 2007, **581**(7):1261-1268.
5. Hermans MM, Kroos MA, van Beeumen J, Oostra BA, Reuser AJ: **Human lysosomal alpha-glucosidase. Characterization of the catalytic site**. *J Biol Chem* 1991, **266**(21):13507-13512.
6. Kimura A, Somoto A, Mori H, Sakai O, Matsui H, Chiba S: **Identification of essential ionizable groups in active site of Aspergillus niger alpha-glucosidase**. *Biosci Biotechnol Biochem* 1997, **61**(3):475-479.
7. Lee SS, He S, Withers SG: **Identification of the catalytic nucleophile of the Family 31 alpha-glucosidase from Aspergillus niger via trapping of a 5-fluoroglycosyl-enzyme intermediate**. *The Biochemical journal* 2001, **359**(Pt 2):381-386.
8. Okuyama M, Okuno A, Shimizu N, Mori H, Kimura A, Chiba S: **Carboxyl group of residue Asp647 as possible proton donor in catalytic reaction of alpha-glucosidase from Schizosaccharomyces pombe**. *Eur J Biochem* 2001, **268**(8):2270-2280.
9. Lovering AL, Lee SS, Kim YW, Withers SG, Strynadka NC: **Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate**. *J Biol Chem* 2005, **280**(3):2105-2115.
10. Hirschhorn R, Reuser AJJ: **Glycogen Storage Disease Type II: Acid α-Glucosidase (Acid Maltase) Deficiency**. In: *The Metabolic & Molecular Bases of Inherited Disease.* Edited by Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW, Vogelstein B, vol. 3, 8th edn. New York: The McGraw-Hill Companies, Inc.; 2001: 3389-3420.
11. Naumov DG: **[Structure and evolution of mammalian maltase-glucoamylase and sucrase-isomaltase genes]**. *Mol Biol (Mosk)* 2007, **41**(6):1056-1068.
12. Nichols BL, Avery S, Sen P, Swallow DM, Hahn D, Sterchi E: **The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities**. *Proc Natl Acad Sci U S A* 2003, **100**(3):1432-1437.
13. Ernst HA, Lo Leggio L, Willemoes M, Leonard G, Blum P, Larsen S: **Structure of the Sulfolobus solfataricus alpha-glucosidase: implications for domain conservation and substrate recognition in GH31**. *J Mol Biol* 2006, **358**(4):1106-1124.
14. Sim L, Quezada-Calvillo R, Sterchi EE, Nichols BL, Rose DR: **Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity**. *J Mol Biol* 2008, **375**(3):782-792.
15. Tajima Y, Matsuzawa F, Aikawa S, Okumiya T, Yoshimizu M, Tsukimura T, Ikekita M, Tsujino S, Tsuji A, Edmunds T *et al*: **Structural and biochemical studies on Pompe disease and a "pseudodeficiency of acid alpha-glucosidase"**. *Journal of human genetics* 2007, **52**(11):898-906.
16. Yoshimizu M, Tajima Y, Matsuzawa F, Aikawa S, Iwamoto K, Kobayashi T, Edmunds T, Fujishima K, Tsuji D, Itoh K *et al*: **Binding parameters and thermodynamics of the interaction of imino sugars with a recombinant human**

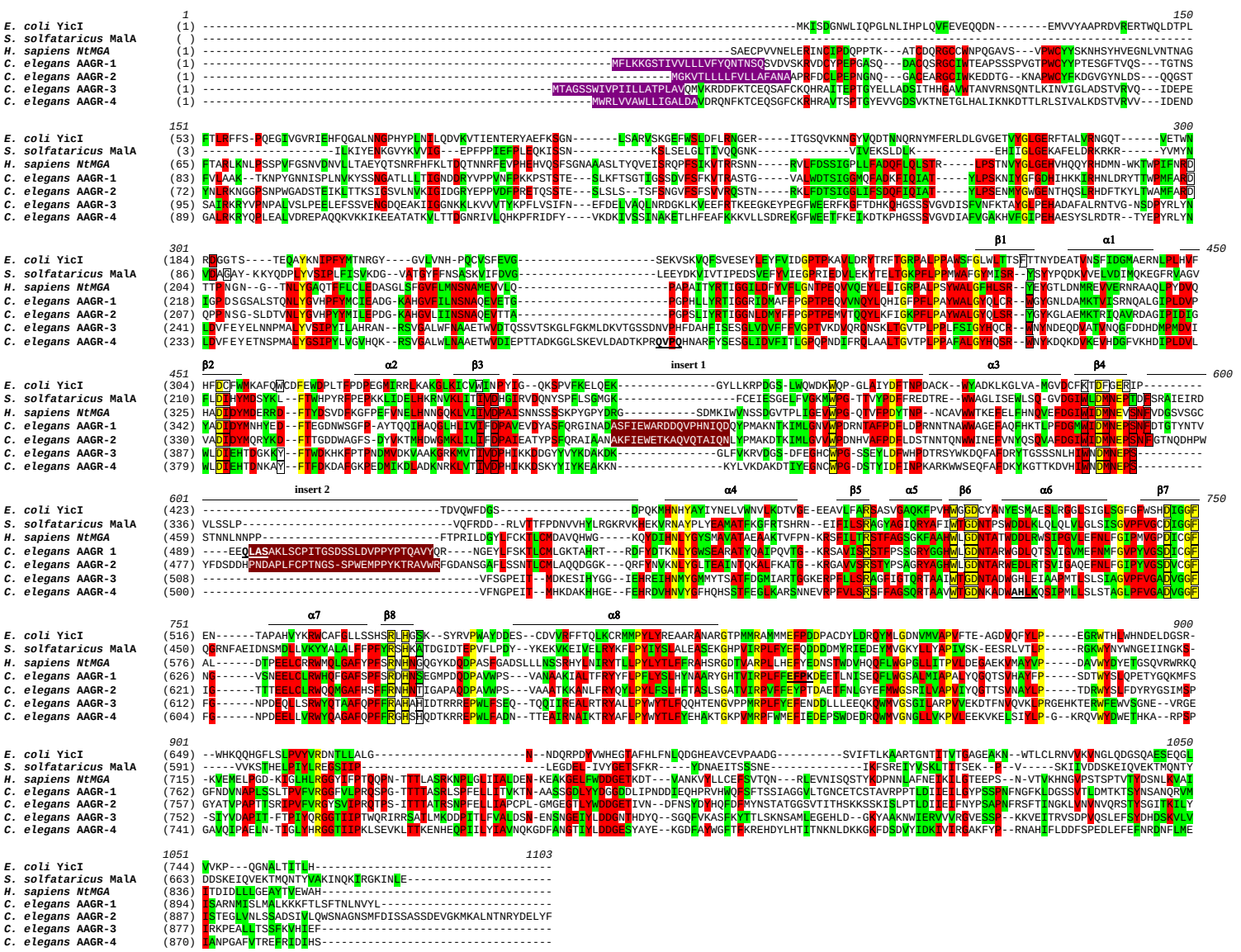
acid alpha-glucosidase (alglucosidase alfa): insight into the complex formation mechanism. *Clinica chimica acta; international journal of clinical chemistry* 2008, **391**(1-2):68-73.
17. Okumiya T, Keulemans JL, Kroos MA, Van der Beek NM, Boer MA, Takeuchi H, Van Diggelen OP, Reuser AJ: **A new diagnostic assay for glycogen storage disease type II in mixed leukocytes**. *Mol Genet Metab* 2006, **88**(1):22-28.
18. Winchester B, Bali D, Bodamer OA, Caillaud C, Christensen E, Cooper A, Cupler E, Deschauer M, Fumic K, Jackson M *et al*: **Methods for a prompt and reliable laboratory diagnosis of Pompe disease: report from an international consensus meeting**. *Mol Genet Metab* 2008, **93**(3):275-281.
19. **Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium**. *Science* 1998, **282**(5396):2012-2018.
20. de Voer G, Peters D, Taschner PE: **Caenorhabditis elegans as a model for lysosomal storage disorders**. *Biochimica et biophysica acta* 2008, **1782**(7-8):433-446.
21. Hanover JA, Forsythe ME, Hennessey PT, Brodigan TM, Love DC, Ashwell G, Krause M: **A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout**. *Proc Natl Acad Sci U S A* 2005, **102**(32):11266-11271.
22. Holt SJ, Riddle DL: **SAGE surveys C. elegans carbohydrate metabolism: evidence for an anaerobic shift in the long-lived dauer larva**. *Mech Ageing Dev* 2003, **124**(7):779-800.
23. **C.elegans Gene Knock-out Consortium** [http://www.celeganskoconsortium.omrf.org/default.aspx]
24. **Wormbase BLAST or BLAT Search** [http://www.wormbase.org/db/searches/blat]
25. **Wormbase** [http://www.wormbase.org/]
26. Blumenthal T, Gleason KS: **Caenorhabditis elegans operons: form and function**. *Nat Rev Genet* 2003, **4**(2):112-120.
27. Hwang HY, Horvitz HR: **The SQV-1 UDP-glucuronic acid decarboxylase and the SQV-7 nucleotide-sugar transporter may act in the Golgi apparatus to affect Caenorhabditis elegans vulval morphogenesis and embryonic development**. *Proc Natl Acad Sci U S A* 2002, **99**(22):14218-14223.
28. **SignalP 3.0 server** [http://www.cbs.dtu.dk/services/SignalP/]
29. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP and related tools**. *Nat Protoc* 2007, **2**(4):953-971.
30. **TargetP 1.1 Server**.
31. Felsenstein J: **Confidence-Limits on Phylogenies - an Approach Using the Bootstrap**. *Evolution* 1985, **39**(4):783-791.
32. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1988, **5**:164-166.
33. Holm L, Park J: **DaliLite workbench for protein structure comparison**. *Bioinformatics* 2000, **16**(6):566-567.
34. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ: **Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function**. *Journal of Computational Chemistry* 1998, **19**(14):1639-1662.
35. Majer F, Pavlickova L, Majer P, Hradilek M, Dolejsi E, Hruskova-Heidingsfeldova O, Pichova I: **Structure-based specificity mapping of secreted aspartic proteases of Candida parapsilosis, Candida albicans, and Candida tropicalis using**

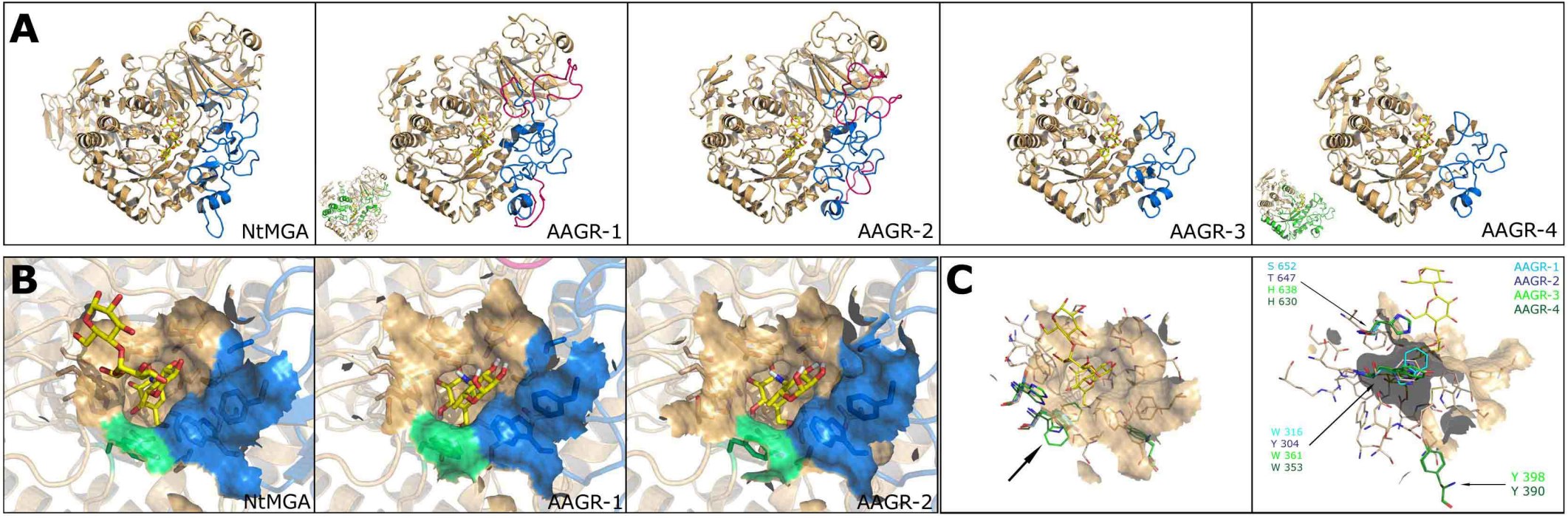
peptidomimetic inhibitors and homology modeling. *Biological chemistry* 2006, **387**(9):1247-1254.
36. Nasi R, Sim L, Rose DR, Pinto BM: **New chain-extended analogues of salacinol and blintol and their glycosidase inhibitory activities. Mapping the active-site requirements of human maltase glucoamylase**. *The Journal of organic chemistry* 2007, **72**(1):180-186.
37. Kroos M, Pomponio RJ, van Vliet L, Palmer RE, Phipps M, Van der Helm R, Halley D, Reuser A: **Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating**. *Hum Mutat* 2008, **29**(6):E13-26.
38. Hujova J, Sikora J, Dobrovolny R, Poupetova H, Ledvinova J, Kostrouchova M, Hrebicek M: **Characterization of gana-1, a Caenorhabditis elegans gene encoding a single ortholog of vertebrate alpha-galactosidase and alpha-N-acetylgalactosaminidase**. *BMC Cell Biol* 2005, **6**(1):5.
39. Meikle PJ, Brooks DA, Ravenscroft EM, Yan M, Williams RE, Jaunzems AE, Chataway TK, Karageorgos LE, Davey RC, Boulter CD *et al*: **Diagnosis of lysosomal storage disorders: evaluation of lysosome- associated membrane protein LAMP-1 as a diagnostic marker**. *Clin Chem* 1997, **43**(8 Pt 1):1325-1335.
40. Kallwass H, Carr C, Gerrein J, Titlow M, Pomponio R, Bali D, Dai J, Kishnani P, Skrinar A, Corzo D *et al*: **Rapid diagnosis of late-onset Pompe disease by fluorometric assay of alpha-glucosidase activities in dried blood spots**. *Mol Genet Metab* 2007, **90**(4):449-452.
41. Jackson AL, Linsley PS: **Noise amidst the silence: off-target effects of siRNAs?** *Trends Genet* 2004, **20**(11):521-524.
42. Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J: **Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans**. *Genome Biol* 2000, **2**(1).
43. Raben N, Takikita S, Pittis MG, Bembi B, Marie SK, Roberts A, Page L, Kishnani PS, Schoser BG, Chien YH *et al*: **Deconstructing Pompe disease by analyzing single muscle fibers: to see a world in a grain of sand**. *Autophagy* 2007, **3**(6):546-552.
44. Mello CC, Kramer JM, Stinchcomb D, Ambros V: **Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences**. *Embo J* 1991, **10**(12):3959-3970.
45. Zimmermann T, Rietdorf J, Pepperkok R: **Spectral imaging and its applications in live cell microscopy**. *FEBS Lett* 2003, **546**(1):87-92.
46. Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A *et al*: **High-throughput in vivo analysis of gene expression in Caenorhabditis elegans**. *PLoS Biol* 2007, **5**(9):e237.
47. Brenner S: **The genetics of Caenorhabditis elegans**. *Genetics* 1974, **77**(1):71-94.
48. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans**. *Nature* 1998, **391**(6669):806-811.
49. **Caenorhabditis Genetics Center** [http://biosci.umn.edu/CGC/CGChomepage.htm]
50. **GenBank** [http://www.ncbi.nlm.nih.gov/Genbank/index.html]
51. Chomczynski P, Sacchi N: **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction**. *Anal Biochem* 1987, **162**(1):156-159.
52. **The C. elegans ORFeome cloning project** [http://worfdb.dfci.harvard.edu/]
53. **Swiss-Prot / TrEMBL database** [http://www.expasy.org/sprot/]
54. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-**

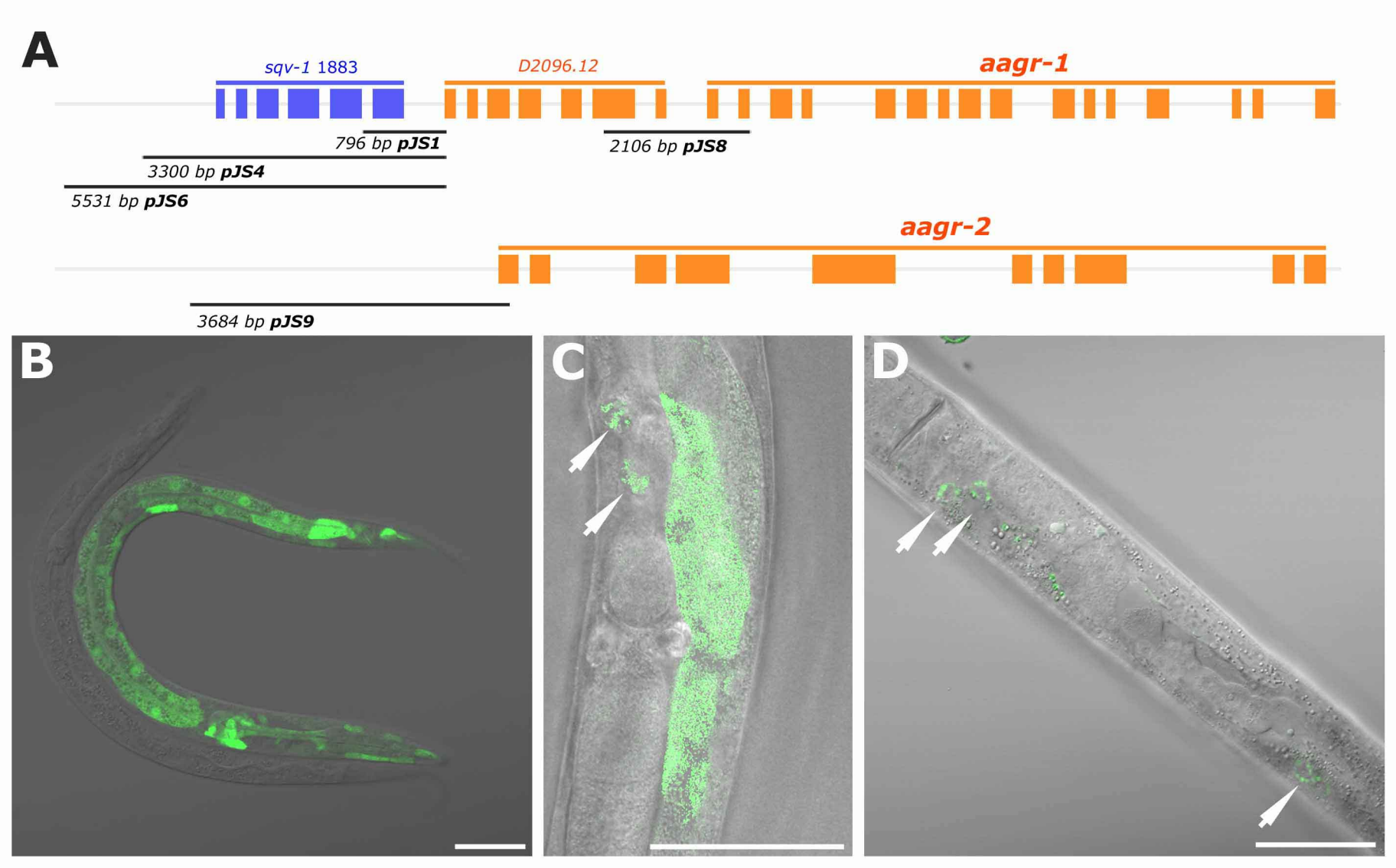
specific gap penalties and weight matrix choice. *Nucleic acids research* 1994, **22**(22):4673-4680.
55. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints**. *J Mol Biol* 1993, **234**(3):779-815.
56. Laskowski RA, Macarthur MW, Moss DS, Thornton JM: **Procheck - a Program to Check the Stereochemical Quality of Protein Structures**. *Journal of Applied Crystallography* 1993, **26**:283-291.
57. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS *et al*: **MolProbity: all-atom contacts and structure validation for proteins and nucleic acids**. *Nucleic acids research* 2007, **35**(Web Server issue):W375-383.
58. Hodgkin J: **Conventional genetics**. In: *Celegans: A Practical Approach.* Edited by Hope IA. Oxford: Oxford University Press; 1999: 245-269.
59. Lerner JM, Zucker RM: **Calibration and validation of confocal spectral imaging systems**. *Cytometry A* 2004, **62**(1):8-34.
60. DeLano WL: **Pymol - DeLano Scientific LLC**. In.; 2006.


A
8617 bp (2835 bp)
SL1
24 bp
1162 bp
B
10748 bp (2986 bp)
SL1
118 bp
C
3525 bp (3054 bp)
SL1
321 bp
D
6241 bp (2823 bp)
SL1
111 bp
1832 bp


C.elegans AAGR-1
M.musculus GAA
C.elegans AAGR-2
C.japonica GAA
B.taurus GAA
H.sapiens SUIS
O.cuniculus SUIS
64
100
68
100
100
H.sapiens GAA
74
100
98
52
R.norvegicus SUIS
100
100
47
H.sapiens GANC
S.murinus SUIS
54
H.sapiens MGA
84
100
H.sapiens GANAB
C.elegans AAGR-3
D.melanogaster GAA-like
C.elegans AAGR-4
S.scrofa GANAB


E. coli YicI
S. solfataricus MalA
H. sapiens NtMGA
C. elegans AAGR-1
C. elegans AAGR-2
C. elegans AAGR-3
C. elegans AAGR-4
β1
α1
β2
α2
β3
insert 1
α3
β4
insert 2
α4
β5
α5
β6
α6
β7
α7
β8
α8


A
NtMGA
AAGR-1
AAGR-2
AAGR-3
AAGR-4
B
NtMGA
AAGR-1
AAGR-2
C
S 652
T 647
H 638
H 630
W 316
Y 304
W 361
W 353
AAGR-1
AAGR-2
AAGR-3
AAGR-4
Y 398
Y 390


A
sqv-1 1883
D2096.12
aagr-1
796 bp pJS1
2106 bp pJS8
3300 bp pJS4
5531 bp pJS6
aagr-2
3684 bp pJS9
B
C
D

residual alpha-glucosidase activity [nmol/$mg_{protein}$*h] after RNAi

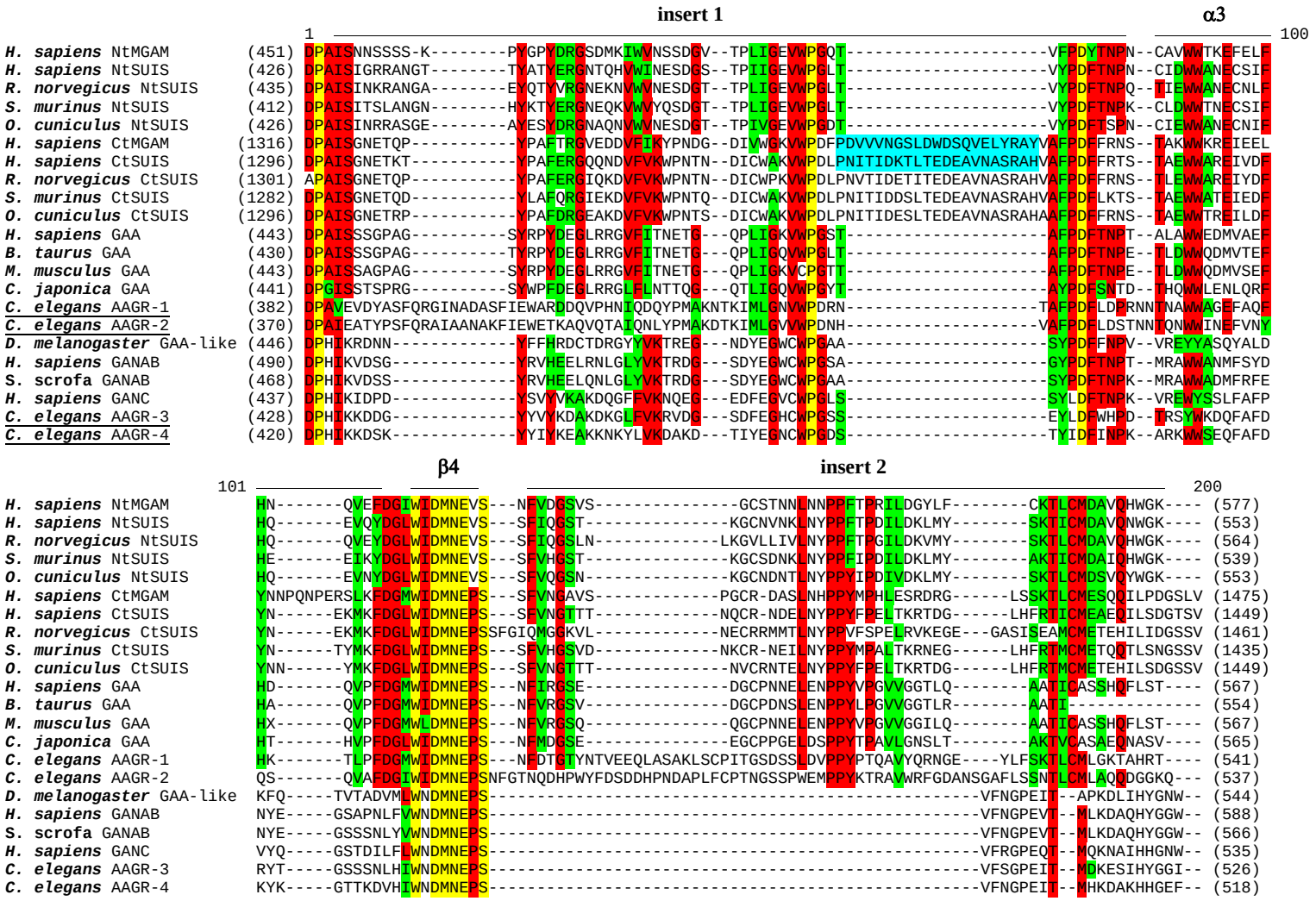
		pH 4,0				**pH 6,5**				
		acarbose -		acarbose +		acarbose -		acarbose +		
Experiment No.	RNAi target	residual activity	% of control	residual activity	% of control	residual activity	% of control	residual activity	% of control	
I.	***aagr-1***	141	**54**	10	**15**	252	**94**	215	**103**	acidic
	aagr-2	26	**10**	20	**31**	188	**70**	183	**88**	
	control	261	**100**	64	**100**	268	**100**	208	**100**	
V.	***aagr-3***	178	**106**	40	**103**	62	**25**	19	**9**	neutral
	aagr-4	166	**99**	37	**94**	203	**83**	157	**76**	
	control	167	**100**	39	**100**	246	**100**	206	**100**	

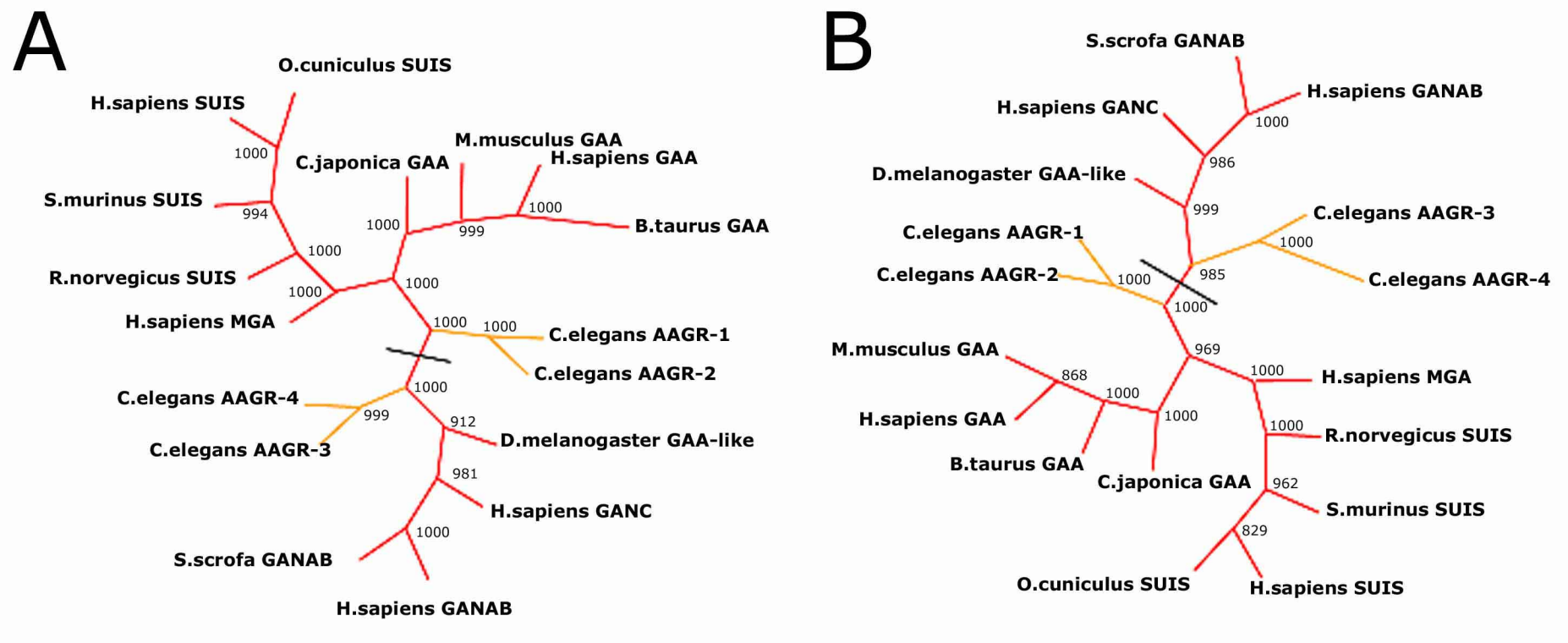
alpha-glucosidase activity [nmol/$mg_{protein}$*h], pH 4,0 after RNAi

Experiment No.	RNAi target	acarbose -		acarbose +		% of residual activity inhibited by acarbose after RNAi	
		residual activity	% of control	residual activity	% of control		
I.	***aagr-1***	119	**55**	18	**26**	**85◄**	acidic
	aagr-2	21	**10**	16	**23**	22	
	control	216	**100**	70	**100**	68	
II.	***aagr-1***	83	**61**	12	**31**	**85◄**	
	aagr-2	19	**14**	13	**32**	32	
	control	137	**100**	40	**100**	71	
III.	***aagr-1***	143	**129**	21	**76**	**86◄**	
	aagr-2	30	**27**	20	**72**	35	
	control	110	**100**	27	**100**	76	
IV.	***aagr-1***	104	**80**	17	**44**	**84◄**	
	aagr-2	34	**26**	22	**57**	34	
	control	129	**100**	38	**100**	70	
V.	***aagr-3***	143	**99**	46	**109**	68	neutral
	aagr-4	136	**95**	41	**99**	70	
	control	144	**100**	42	**100**	71	
VI.	***aagr-3***	165	**98**	53	**114**	68	
	aagr-4	155	**92**	44	**95**	71	
	control	168	**100**	47	**100**	72	
VII.	***aagr-3***	198	**104**	64	**115**	68	
	aagr-4	205	**107**	61	**109**	70	
	control	191	**100**	56	**100**	71	
VIII.	***aagr-3***	160	**99**	58	**114**	64	
	aagr-4	154	**95**	46	**91**	70	
	control	162	**100**	51	**100**	69	

residual alpha-glucosidase activity [nmol/$mg_{protein}$*h] in deletion mutants

	pH 4,0				**pH 6,5**		% of residual alpha-glucosidase activity inhibited by acarbose (pH 4,0)
	acarbose -		acarbose +		acarbose -		
	residual activity	% of control	residual activity	% of control	residual activity	% of control	
aagr-1 - ok2317	147	**75**	12	**33**	329	**83**	**92◄**
control N2	198	**100**	37	**100**	398	**100**	81
aagr-4 - ok1423	134	**82**			264	**95**	
control N2	165	**100**			277	**100**	
RNAi of *aagr-3* in *ok1423* deletion mutant	260	**110**			99	**26**	
control in *ok1423*	238	**100**			375	**100**	


insert 1
α3
β4
insert 2
1
100
101
200
H. sapiens NtMGAM (451)
H. sapiens NtSUIS (426)
R. norvegicus NtSUIS (435)
S. murinus NtSUIS (412)
O. cuniculus NtSUIS (426)
H. sapiens CtMGAM (1316)
H. sapiens CtSUIS (1296)
R. norvegicus CtSUIS (1301)
S. murinus CtSUIS (1282)
O. cuniculus CtSUIS (1296)
H. sapiens GAA (443)
B. taurus GAA (430)
M. musculus GAA (443)
C. japonica GAA (441)
C. elegans AAGR-1 (382)
C. elegans AAGR-2 (370)
D. melanogaster GAA-like (446)
H. sapiens GANAB (490)
S. scrofa GANAB (468)
H. sapiens GANC (437)
C. elegans AAGR-3 (428)
C. elegans AAGR-4 (420)
(577)
(553)
(564)
(539)
(553)
(1475)
(1449)
(1461)
(1435)
(1449)
(567)
(554)
(567)
(565)
(541)
(537)
(544)
(588)
(566)
(535)
(526)
(518)


A
O.cuniculus SUIS
H.sapiens SUIS
1000
M.musculus GAA
C.japonica GAA
H.sapiens GAA
S.murinus SUIS
994
1000
1000
999
B.taurus GAA
1000
R.norvegicus SUIS
1000
1000
H.sapiens MGA
1000
1000
C.elegans AAGR-1
C.elegans AAGR-2
1000
C.elegans AAGR-4
999
912
D.melanogaster GAA-like
C.elegans AAGR-3
981
H.sapiens GANC
1000
S.scrofa GANAB
H.sapiens GANAB
B
S.scrofa GANAB
H.sapiens GANAB
1000
H.sapiens GANC
986
D.melanogaster GAA-like
999
C.elegans AAGR-3
1000
C.elegans AAGR-1
C.elegans AAGR-2
1000
985
C.elegans AAGR-4
1000
969
M.musculus GAA
868
1000
H.sapiens MGA
1000
H.sapiens GAA
1000
1000
R.norvegicus SUIS
B.taurus GAA
C.japonica GAA
962
S.murinus SUIS
829
O.cuniculus SUIS
H.sapiens SUIS

			pdb	N-terminal domain			catalytic domain										C-terminal domain			
				trefoil				proximal insert (1)					distal insert (2)				proximal C-term		distal C-term	
								C.e. insertion					C.e. insertion							
templates	*E.coli*	YicI	1XSJ	1	-	247	248	348			391 *(43 aa)*		420			437 *(17 aa)*	588 *(340 aa)*	589	665	666	773
	S.sulfataricus	MalA	2GM3	1 (3)	-	155	156	252			295 *(43 aa)*		324			372 *(48 aa)*	530 *(374 aa)*	531	604	605	693
	H. sapiens	**NtMGA**	**2QMJ**	**1 (7)**	**52**	**270**	**271**	**367**			**417 *(50 aa)***		**447**			**497 *(50 aa)***	**652 *(381 aa)***	**653**	**731**	**732**	**869**
targets	*H. sapiens*	GAA		1	135	353	354	444			492 *(48 aa)*		528			572 *(44 aa)*	724 *(370 aa)*	724	820	821	936
	C.elegans	AAGR-1		1 (25)	73	293	294	383	400	417	445 *(62 aa)*		477	492	519	548 *(71 aa)*	699 *(405 aa)*	700	780	781	(824) 955
	C.elegans	AAGR-2		1 (16)	62	281	282	371	388	405	433 *(62 aa)*		465	484	511	543 *(78 aa)*	694 *(412 aa)*	695	775	776	(821) 910
	C.elegans	AAGR-3		1 (37)	-	(330) 338	339	429			474 *(45 aa)*		508			525 *(17 aa)*	686 *(347 aa)*	687	768	769	(814) 903
	C.elegans	AAGR-4		1 (24)	-	(322) 330	331	421			466 *(45 aa)*		500			517 *(17 aa)*	678 *(347 aa)*	679	758	759	(805) 952

organism	protein	*.pdb*																							
E. coli	YicI	1XSJ	D 185 [6]		F 277 [3, 7]	D 306 [3]	C 307 [3]	W 315 [3]	W 345 [3]	- -	W 380 [3]	K 414 [3]	D 416 [1]	F 417 [3]	R 420	- -	R 466	W 479 [3,6]	G 481	D 482 [2]	D 511 [6]	F 515 [3]	R 538	H 540	S 542
S. sulfataricus	MalA	2GM3	D 87 [5]	G **89**	Y 184 [3]	D 212 [4]	I 213 [3]		I 249 [3]	D 251 [6]	W 284 [5]	W 318 [3]	D 320 [1]	M 321 [5]	T 325	F 327 [5]	R 400 [4, 5]	W 413 [3]	G 415	D 416 [2]	D 445 [6]	F 449 [3]	R 478	H 480 [4]	A 482
H. sapiens	**NtMGA**	**2QMJ**	**D 203** [5]		**Y 299** [3]	**D 327** [4]	**I 328** [3]		**I 364** [3]	**D 366**	**W 406** [5]	**W 441** [3]	**D 443** [1]	**M 444** [3]	**S 448**	**F 450** [5]	**R 526** [4, 5]	**W 539**	**G 541**	**D 542** [2]	**D 571**	**F 575**	**R 598**	**H 600** [4]	**G 602**
C. elegans	AAGR-1		D 217		W 316	D 344	I 345		I 380	D 382	W 433	W 471	D 473	M 474	S 478	F 480	R 576	W 589	G 591	D 592	D 621	F 625	R 648	H 650	S 652
C. elegans	AAGR-2		D 206		Y 304	D 332	I 333		I 368	D 370	W 421	W 459	D 461	M 462	S 466	F 468	R 571	W 584	G 586	D 587	D 616	F 620	R 643	H 645	T 647
C. elegans	AAGR-3				W 361	D 389	I 390	Y 398	I 426	D 428	W 463	W 500	D 502	M 503	S 507		R 562	W 575	G 577	D 578	D 607	F 611	R 634	H 636	H 638
C. elegans	AAGR-4				W 353	D 381	I 322	Y 390	I 418	D 420	W 455	W 492	D 494	M 495	S 499		R 554	W 567	G 569	D 570	D 599	F 603	R 626	H 628	H 630
H. sapiens	GAA		D 282		W 376	D 404 [8]	L 405		I 441 [8]	D 443	W 481 [8]	W 516 [8]	D 518 [8]	M 519 [8]	S 523	F 525	R 600 [8]	W 613	G 615	D 616 [8]	D 645	F 649 [8]	R 672	H 674 [8]	S 676
GSDII GAA pathogenic variations					del/ins (*sev*)	D/N (*sev*)	L/P (*sev*)				W/R (sev)	STOP (*sev*)		M/V (*int*) M/T (*sev*)			R/C (*sev*) R/H (*sev*) R/L (*unk*)		G/R (*sev*)		D/N (*sev*) D/Y (*sev*) D/E (*sev*)		R/W (*int*) R/E (*int*)		

part of the model			**NtMGA** (query)					**YicI** (query)	**MalA** (query)	**YicI** (query)
			N-terminal domain	***catalytic domain***			***proximal C-terminal domain***	**catalytic domain**	**catalytic domain**	**catalytic domain**
					proximal insert	***distal insert***				YicI template
	AAGR-1	*Z-score*	36,9	59,6	5,1		17,4	31,9	46,9	40,4
		aligned residues	243	363	40	*no structural similarities*	78	296	349	292
		RMSD (Å)	1,1	1,0	1,1		0,4	2,1	2,3	0,7
		seq. Identity (%)	36	45	18		32	19	30	18
	AAGR-2	*Z-score*	39,2	60,2	2,9		17,1	31,4	46,7	40,5
		aligned residues	254	367	37	*no structural similarities*	79	296	351	292
		RMSD (Å)	1,2	0,9	1,9		0,6	2,2	2,9	0,7
		seq. Identity (%)	37	42	16		28	18	29	18
	AAGR-3	*Z-score*	14,4	50			15,1	355	52	44,3
		aligned residues	199	342	*no structural similarities*	*not present*	77	299	340	295
		RMSD (Å)	3,1	1,2			0,9	1,9	1,2	0,5
		seq. Identity (%)	13	31			26	23	34	23
	AAGR-4	*Z-score*	13,2	50,7			15,5	34,4	52,1	44,2
		aligned residues	186	341	*no structural similarities*	*not present*	78	298	340	295
		RMSD (Å)	3,0	1,2			1,1	2	1,1	0,6
		seq. Identity (%)	13	35			40	23	35	23