Institute of Physical Biology, University of South Bohemia

Institute of Nanobiology and Structural Biology, Academy of Sciences of the Czech Republic

PH. D. THESIS

Modeling Substrate-Enzyme Interactions in Fungal Hydrolases

Author: Natallia Kulik

Supervisor: Doc. RNDr. Rüdiger Ettrich, PhD.

Institute of Nanobiology and Structural Biology of GCRC, Academy of Sciences of the Czech Republic

Institute of Physical Biology of University of South Bohemia

Nove Hrady, Czech Republic, 2011

ACKNOWLEDGEMENTS

It is an honor for me to thank those who made this thesis possible, who helped and supported me.

First of all I would like to thank my parents who taught me the most important things people need in life and always encouraged and supported me.

My supervisor Rudiger Ettrich, who was patient with me, and always inspired us to be creative in our work.

From all the people involved in biocatalysis and biotransformation I would especially like to thank our colleagues from Prague, who did all the experimental work. My personal thanks goes to Prof. Vladimir Kren, Prof. Karel Bezouska and Kristyna Slamova.

Thanks to my teachers at the University of South Bohemia, who advised and supported me during my work – Michal Kuty, Milan Predota and Rudiger Ettrich.

Peter Palencar for help in my first steps in MD with Yasara.

Thanks to my colleagues for healthy criticism and the pleasant working atmosphere in our lab: Zofie Sovova, Milan Melichercik, Vasilina Zayats, Morteza Khabiri.

Petr Hamberger and Martin Safranek - for technical support.

FUNDINGS

This PhD thesis was elaborated within the National Center of Biocatalysis and Biotransformation (MSMT LC06010). Additionally, my PhD studies were supported from the University of South Bohemia by grant GAJU 170/2010/P. A lot of computational work has been done using METACentrum computing facilities provided under the research concept MSM6383917201.

DECLARATION

Prohlašuji, že svoji disertační práce jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí (článku v Appendix) archivovaných Ústavem Fyzikální biologie elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

So you see! There's no end to the things you might know,
Depending how far beyond Zebra you go.

Dr Seuss

ABSTRACT

Modern science utilizes a huge number of methods from different fields for complex investigations of biological systems. The great jump in computational power in recent years allowed modeling and simulation of larger, biologically more relevant, systems of biomolecules.

Advantages of computational methods used to study protein and protein-substrate complexes are their relative low cost, the comparable large number of software that could be used and the variety of properties which could be studied. These methods have in common that they investigate systems on atomic scale and propagate them in time. Molecular dynamics simulations hereby have some limitations like the high computational demand resulting in long calculation times or modeling of transition state or reaction pathway. Homology modeling helps to overcome the gap between the number of sequenced proteins and solved structures. Screening of the NCBI database (http://www.ncbi.nlm.nih.gov/sites/gquery) demonstrates that among the 34621 protein sequences classified as glycosidases just 639 of them have an experimentally solved 3-D structure.

In my PhD work I focused on modeling substrate-enzyme interactions in fungal hydrolases capable of transglycosylation and nitrile hydrolysis with considerable potential in practical applications of biotransformation. The concrete hydrolase systems used for computational studies have been alpha-galactosaminidase from *Aspergillus niger;* beta-N-acetylhexosaminidases from *Aspergillus oryzae* and *Penicillium oxalicum*; and nitrilase from *Aspergillus niger* K10. Hereby, a great part was dedicated to modeling and analysis of substrate-enzyme interaction including substrate and inhibitor docking.

My work was in close collaboration with the experimentalists providing screening studies of enzyme substrate interactions and the selection of studied complexes, substrates or inhibitors as well as my predictions and results were under constant feedback and evaluation from the experimental side.

The investigation of substrate specificity and affinity of beta-N-acetylhexosaminidase to structurally modified substrates is relevant for the design of new effective drugs against glycosidase-induced infections (Carmona *et al.* 2006; Fialová *et al.* 2005b; Varki *et al.* 1999), enzymatic synthesis of important compounds. Modified carbohydrates are regarded

as important therapeutic agents, but it is often tricky to control stereochemistry of synthesized carbohydrates (Elhalabi and Rice 1999).

Enzymatic utilization of nitriles in the environment has a broad ecological significance (Brady 2009; Polaina and MacCabe 2007; Singh *et al.* 2006).

Differences in substrate affinity of beta-N-acetylhexosaminidases from *A. oryzae* and *P. oxalicum* were analyzed based on structural features of the two enzymes.

Carbohydrates are important as independent structures, which could act as substrates or inhibitors, but also as covalently linked to glycoproteins or as part of glycosphingolipids. Glycosylation plays diverse roles, which can be divided in two major functional groups: structural and modulatory; as a marker for recognition by other molecules (Varki *et al.* 1999).

Investigation of carbohydrates is a challenging topic because of the high flexibility of carbohydrate structures. It is rather difficult to model as flexible system by computational methods (Frank and Schloissnig 2010; Lütteke 2009).

In my thesis I first give a literature review of glycosyl hydrolases and nitrilases with the main attention to the studied enzymes alpha-galactosidases and alpha-N-acetylgalactosaminidases from *A. niger*, beta-N-acetylhexosaminidases from *A. oryzae* and *P. oxalicum*, nitrilase from *A. niger* and a general introduction into the field (**Part I**). Then I continue with description of the methods used (**Part II)**, to finally sum up my gained results and connect the papers the thesis is based on (**Part III**). Papers are attached to the work as **Appendix**.

ABBREVIATIONS

Ala – alanine

Arg - arginine

Asp - aspartic acid

Cys – cysteine

EC number - enzyme commission number

GALase(s) - alpha-galactosidase(s)

GalNAc - N-acetylgalactosamine

GH - glycosyl hydrolase

GlcNAc - N-acetylglucosamine

Glu - glutamic acid

HEX(s) - beta-N-acetylhexosaminidase(s)

Lys - lysine

MAF - macrophage activating factor

ManNAc - N-acetylmanosamine

MD – molecular dynamics simulation

Met - methionine

NAGALase(s) - alpha-N-acetyl galactosaminidase(s)

NK - natural killer

pNP - p-nitrophenyl

RMSD - root mean square deviation

Ser – serine

Trp – triptophan

CONTENT

PART I. INTRODUCTION (LITERATURE REVIEW)

1. GLYCOSYL HYDROLASES

Glycosyl hydrolases (also called glycoside hydrolases or glycosidases) are a large group of enzymes, which catalyze the hydrolysis of the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (http://en.wikipedia.org/wiki/Glycoside_hydrolase).

The general reaction glycosidases catalyze is shown in Figure 1.1.

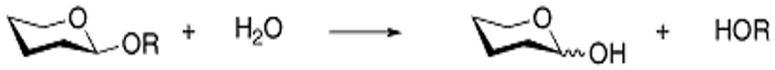

Figure 1.1. Scheme of hydrolysis by glycosidases (picture is from http://en.wikipedia.org/wiki/Glycoside_hydrolase).

1.1 OCCURRENCE AND ROLE IN THE NATURE

Glycosidases are found in all organisms from prokaryotes to eukaryotes. They play an important and diverse roles in nature - participate in “degradation of biomass such as cellulose and hemicellulose, in anti-bacterial defense strategies (eg lysozyme), in pathogenesis mechanisms (eg viral neuraminidases) and in normal cellular function (eg trimming mannosidases involved in N-linked glycoprotein biosynthesis)” (from http://en.wikipedia.org/wiki/Glycoside_hydrolase).

Alpha-galactosidases (GALases) occur widely in microorganisms, plants, and animals, and catalyze the hydrolysis of alpha-1,6-linked galactose residues from oligomeric (meloboise, raffinose) and polymeric (galactomannan) compounds. (Margolles-Clark *et al.* 1996; Manzanares *et al.* 1998).

Human GALase predominantly hydrolyzes ceramide trihexoside (glycosfingolipid, containing of the sugars galactose-galactose-glucose bonded to a ceramide). Fabry's disease (also known as Fabry's disease, Anderson-Fabry disease, angiokeratoma corporis diffusum and GALase A deficiency) (Calhoun *et al.* 1985) is connected with genetic deficiency or decreased efficiency of the GALase A (http://www.livestrong.com/article/244724-about-the-enzyme-alpha-galactosidase/#).
Without this enzyme, ceramide trihexosides cannot be metabolized and accumulate in a variety of tissues (for instance blood vessels, kidneys, heart, and nervous tissue). Enzyme replacement therapy, used for normalization of metabolism, prevents disease progression as well as potentially reverses symptoms; however it is not a cure.

Physiological substrates of alpha-N-acetylgalactosaminidases (NAGALases) are O-glycosidic core structures in glycopeptides and glycoproteins and blood group A determinants in oligosaccharides, glycoproteins and glycosphingolipids. Deficiency of this enzyme in human leads to Schindler disease (one of forms of NAGALase deficiency is named Kanzaki disease) (Kanzaki *et al.* 1989; Schindler *et al.* 1989), when accumulation of glycosphingolipids throughout the body occurres. The only strategy in the treatment today is in the suppression of symptoms (http://www.wrongdiagnosis.com/k/kanzaki_disease/treatments).

The other role of NAGALases in human organism is connected with the immune system. Increase in NAGALase activity in serum of patients with cancer and HIV leads to immune suppression (Yamamoto *et al.* 1995 and 1997; Yamamoto 2006).

The excess of NAGALase is secreted by HIV (Human Immunodeficiency Virus) and cancer infected cells (Yamamoto *et al.* 1995 and 1997; Yamamoto 2006). Serum Gc protein (known as vitamin D3-binding protein) is the precursor for the principal macrophage activating factor (MAF), and has a N-acetylgalactosamine sugar moiety. The MAF precursor activity of Gc protein is lost or reduced in case of disease because the Gc protein is deglycosylated by NAGALase secreted by infected cells.

Biochemical characteristics of NAGALase in tumor cells supported the idea that it is an exo-enzyme with biochemical properties different from normal enzyme (Mohamad *et al.* 2002). Activity of NAGALase can be used to monitor cure rate during GcMAF therapy of HIV-infected patients (Yamamoto *et al.* 2009).

Beta-N-acetylhexosaminidases, also known as N-acetyl-beta-hexosaminidases or hexosaminidases, HEXs, are found in bacteria’s, fungus, plants and animals. HEXs catalyze the hydrolysis of chitobiose into monosaccharides, however the exact role of these enzymes in different organisms differs.

The biological role of HEXs is mostly studied in mammals, especially in humans (found in lysosomes, plasma membrane and cytosol). Lysosomal HEXs are described better than other, as the deficiency of these enzymes causes several genetic diseases: Sandhoff disease (GM2-gangliosidosis, variant 0), Tay-Sachs disease (adult GM2-gangliosidosis) (Sandhoff *et al.* 1968; Tay 1881). Fungal HEX plays an important role in life cycle of the fungus for formation septa, germ tubes and fruit bodies (Gooday *et al.* 1992).

1.2 BIOTECHNOLOGICAL IMPORTANCE OF ENZYMES

1.2.1 NAGALase and GALase

Despite of its high importance in glycoproteins and glycolipids chemistry NAGALases were purified just from some limited sources (Callahan *et al.* 1973; Levy and Aminoff 1980; Sung *et al.* 1980).

Recent high attention to these enzymes is connected with ability of NAGALase and GALase to convert blood type group A and B correspondingly to the most common group O (Lenny *et al.* 1995; Liu *et al.* 2007).

GALases are able to hydrolyze transglycosylation reactions in access of substrate (Goulas *et al.* 2009). Kinetic experiments showed that commercially available GALase from *A. niger* APC-9319 is able to catalyze transglycosylation reaction and produce alpha-galactooligosaccharides in a higher yield than other fungal enzymes (Yamashita *et al.* 2005).

1.2.2 HEXs from *A. oryzae* and *P. oxalicum*

These enzymes found many applications in biotechnology and the chemoenzymatic synthesis of oligosaccharides. HEX from *A. oryzae* is able to take part in reverse hydrolysis (Rajnochova *et al.* 1997) and is useful for transglycosylation (Crout *et al.* 1992).

They ability to accept modified substrates attracts great interest from biochemists. Biosynthesis of complex carbohydrates by HEXs is an advanced method (Carmona *et al.* 2006; Fialová *et al.* 2004, 2005a, 2005b; Ogata *et al.* 2007; Uzawa *et al.* 2003; Zeng *et al.* 2007). The advantage of transglycosylation over direct synthesis is its efficiency and the better economics (Ogata *et al.* 2007). Glycosyltransferases commonly used for synthesis are usually metal ion-dependent, with magnesium or manganese being found in the active site, and are by far more expensive than HEXs. High regioselectivity of HEXs makes them a good producer.

The possibility to use HEXs in preparation of immunoactive structures (which are mostly composed of *N*-acetyl-d-hexosamines – GalNAc, GlcNAc, ManNAc) able to stimulate the immune response of an organism by activation of natural killer (NK) cells through NKR-P1A and other cell surface activation receptors (Hušáková *et al.* 2001, Fialová *et al.*, 2005a, Krist *et al.* 2001) is interesting and challenging.

Another feature of HEXs from fungi that attracts high interest, is their diverse affinity to modified substrates among enzymes from different strains. Fungal enzymes are able to

synthesize nucleotide activated oligosaccharides (particularly from *A. oryzae* and *P. oxalicum*) (Nieder *et al.* 2004).

However, understanding of utilization of HEXs in chemoenzymatic synthesis needs an interdisciplinary approach.

1.3 CLASSIFICATION

Several types of classification exist based on structure, function, catalyzed reaction, and evolutional relationship between enzymes.

1.3.1 Catalyzed reaction

One of the most often used classification is based on the catalyzed substrate - this classification is reflected in the name of enzyme alpha-galactosidase, beta-galactosidase, mannosidase and so on. Enzymes are grouped under certain enzyme commission number (EC number) (Webb 1992). HEXs and GALases discriminate between alpha and beta enantiomers.

Mechanistic classifications imply the mechanism of reaction (discussed later in chapter 1.4) and distinguish retaining and inverting enzymes (Koshland 1953).

Glycosidases cleave substrates at different positions: Either at a terminal position of the substrate (*exo*-glycosidases) or within the middle of a chain (*endo*-). The position of the cleaved substrate within the chain determines the active site topology (pocket, groove or tunnel) (Davies and Henrissat 1995).

1.3.2 Sequence similarity

The Carbohydrate-Active Enzymes database (CAZY) classify enzymes on the base of amino acid sequence similarities, that imply folding similarities and better reveal evolutional relationship between enzymes (http://www.cazy.org/Glycoside-Hydrolases.html). There are 115 families of glycoside hydrolases (GH) in the CAZY database to date (16.02.2010). Some families of GH are clustered based on folds in ‘clans’ with letter numeration from A to H (for example GH-J).

1.3.3 Classification of studied enzymes

GALase from *A. niger* is an *exo*-enzyme, EC 3.2.1.22. It uses the so-called retaining mechanism of hydrolysis and belongs to the GH family 27 (Fujimoto *et al.* 2003).

The studied NAGALases belong to the EC 3.2.1.49, GH family 27, *exo*-enzyme, and as the GALase they use the retaining mechanism of hydrolysis (Garman *et al.* 2002).

HEXs from *A. oryzae* and *P. oxalicum* belong to the EC 3.2.1.52, -*exo* type of retaining enzymes, GH 20 (Matsuo *et al.* 2003; Plíhal *et al.* 2004).

1.4 CATALYTIC MECHANISM

Enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires two critical residues: a proton donor, and nucleophile/base (Koshland 1953). This role is mainly played by aspartic acid (Asp) and/or glutamic acid (Glu) (Henrissat and Bairoch 1993).

There are two types of glycosidases according to their catalytic mechanism – retaining and inverting. Inverting hydrolysis (Figure 1.2A (Bojarova and Kren 2009)) proceeds with one step and products have opposite to the substrate anomeric configuration. Retaining enzymes in contrast act by two steps. At each step the anomeric configuration is inverted leading to the final preservation of the anomeric configuration (Figure 1.2B). Retaining glycosidases show transferase activity (http://en.wikipedia.org/wiki/Glycoside_hydrolase).

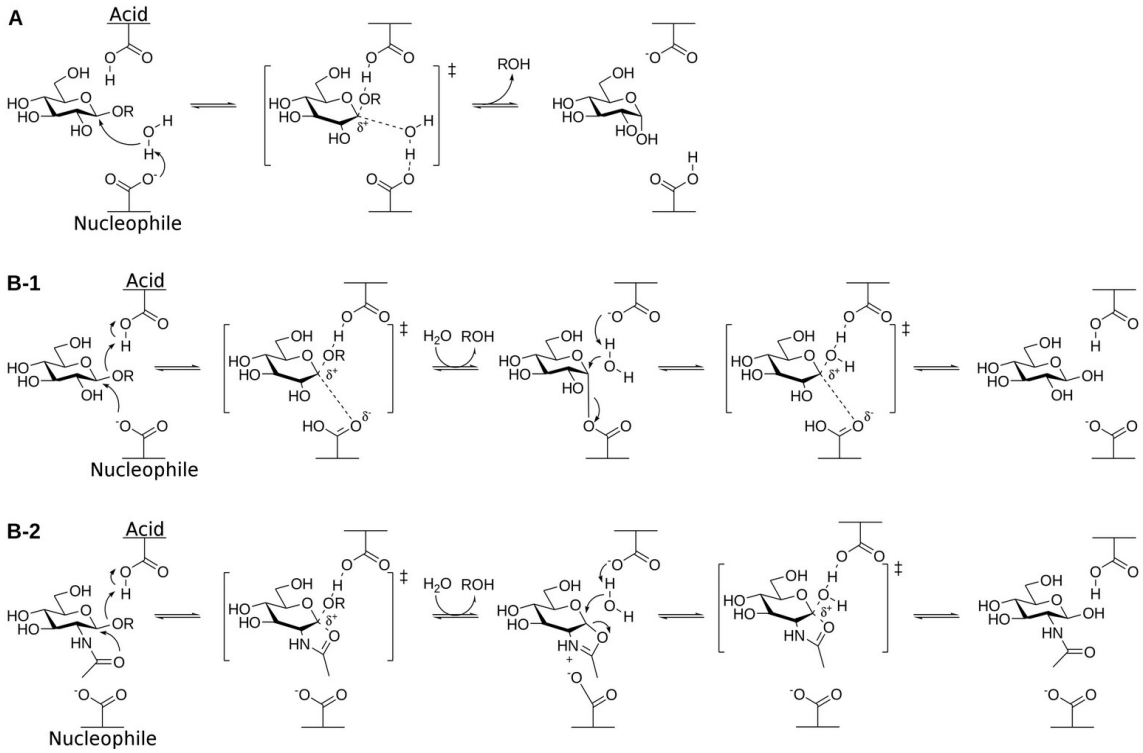

Figure 1.2. Two types of hydrolysis mechanisms (Bojarova and Kren 2009): inverting (A) and retaining (B-1) glycosidase; HEXs using a modified retaining mechanism (B-2).

Certain HEXs use an alternative mechanism for hydrolysis (Figure 1.2B-2) that proceeds through a nucleophilic residue bound to the substrate, rather than being attached

to the enzyme. The role of the nucleophile is played by the acetamido-group, which participates in the formation of an intermediate oxazoline or oxazolinium ion. The two types of retaining hydrolysis differ in the intermediate state.

In both, retaining and inverting enzymes, the position of proton donor (acid) is identical – within hydrogen-bonding distance of the glycosidic oxygen. The distance from nucleophile/catalytic base is different. In retaining enzymes it is in close vicinity of the sugar anomeric carbon, while the distance to the base in inverting enzymes is larger to accommodate a water molecule inbetween. The distance between the catalytic residues in inverting enzymes is approximately 0.95-1 nm, in retaining it is less – 0.55 nm (Ly and Withers 1999; McCarter and Withers 1994).

An important role in hydrolysis is played by carbohydrate ring conformation. Distortion of the hexose geometry into the sofa conformation helps to lower the reaction energy barrier (Davies and Henrissat 1995). The schematical change in position of the C1 atom of the substrate of HEXs is shown in figure 1.3. The initial position (*position 1*) is the conformation before hydrolysis. The conformation corresponding to the first step of hydrolysis (formation of new bond) is *Position 3*. *Position 2* represents the transition state when C-1, C-2, C-5, and O-5 are coplanar (Mark *et al.* 2001; Sinnott 1990).

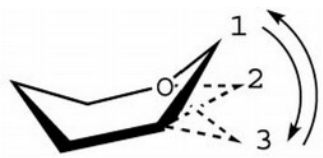

Figure 1.3. Anomeric carbon movement during hydrolysis (from Mark *et al.* 2001).

Accurate values of equilibrium constants for disaccharide hydrolysis is difficult to determine, as there are many isomers in solution.

1.4.1 Catalytic mechanism of NAGALase and GALase

Retaining hydrolysis by NAGALase and GALase usually proceeds not via an intermediate oxazolinium ion, catalytic residues are both Asp (Figure 1.2B-1).

In the first step the oxygen atom from Asp 140 from NAGALase makes a nucleophilic attack on C1 of the terminal alpha-N-acetylgalactosamine of the substrate, cleaving the glycosidic linkage and producing covalent substrate-enzyme intermediate (Garman *et al.* 2002). During the second step, a water molecule, deprotonated by Asp 201, makes a nucleophilic attack on C1 of the substrate, cleaving the covalent enzyme-substrate intermediate complex.

1.4.2 Catalytic mechanism of HEXs

HEX operates through a retaining mechanism (Drouillard *et al.* 1997, Lai and Withers 1994). Kinetic study based of binding with inhibitors (C-2 modified N-acetylcarbohydrates) confirmed that HEXs from the GH20 family use the second retaining mechanism (Figure 1.2B-2) via an intermediate oxazolinium ion (Drouillard *et al.* 1997; Knapp *et al.* 1996; Legler and Bollhagen 1992; Mega *et al.* 1970).

GH 20 class enzymes contain a highly conserved pair of catalytic residues Asp-Glu. Glu acts as a proton donor and substrate 2-acetamido group plays a role of a nucleophile instead of the Asp (Tews *et al.* 1997; Mark *et al.* 2001).

In the first step of the reaction Glu donates its proton to the glycosidic oxygen, which leads to breaking of the O-C1 bond. The non-reducing sugar is bound in a chair conformation (Mark *et al.* 2001). During the first step it is distorted from a clear chair conformation towards a transition sofa-like structure, resulting in the acetamido carboxyl coming closer to the anomeric C1 atom (Tews *et al.* 1996). The first intermediate is stabilized by the oxazolinium ion formation, the reducing carbohydrate leaves the active site and is replaced with a water molecule.

In the second step the hydrolysis of the intermediate proceeds. Water, activated by Glu, attacks the anomeric center C-1 (Mark *et al.* 2001).

1.5 TRANSGLYCOSYLATION REACTION

Transglycosylation – is a mechanism for glycosidic bond formation.

Wide usage of glycosides to catalyse the formation of a new glycosidic bond by a transglycosylation reaction has been reported (Spangenberg 2000; Wang and Huang 2009), including bacterial glycosidases (Tramice *et al.* 2007) and fungal (Sulzenbacher *et al.* 1996; Varki A *et al.*, editors 1999; York and Hawkins 2000). HEX from *A. oryzae* is frequently used (Singh 1995, Varki A *et al.*, editors 1999).

The proposed mechanism for hydrolysis and transglycosylation reactions is depicted in Figure 1.4. The transglycosylation reaction is initiated by simultaneous protonation of the glycosidic bond by a proton donor and a nucleophilic attack on the anomeric carbon of the glucose moiety. The intermediate can react with either water or with another saccharide molecule. The ratio between hydrolysis and transglycosylation is determined by relative concentrations of water and sugar moieties in the active site only (Skov *et al.* 2001).

1.6 DEHYDRATION (CONDENSATION) REACTION

The condensation reaction can be seen as a special case of transglycosylation when the water molecule is released. The condensation reaction catalyzed by HEX from *A. oryzae* is described in (Rajnochova *et al.* 1997). The condensation reaction alone is energetically unfavorable because energy is required to form a bond.

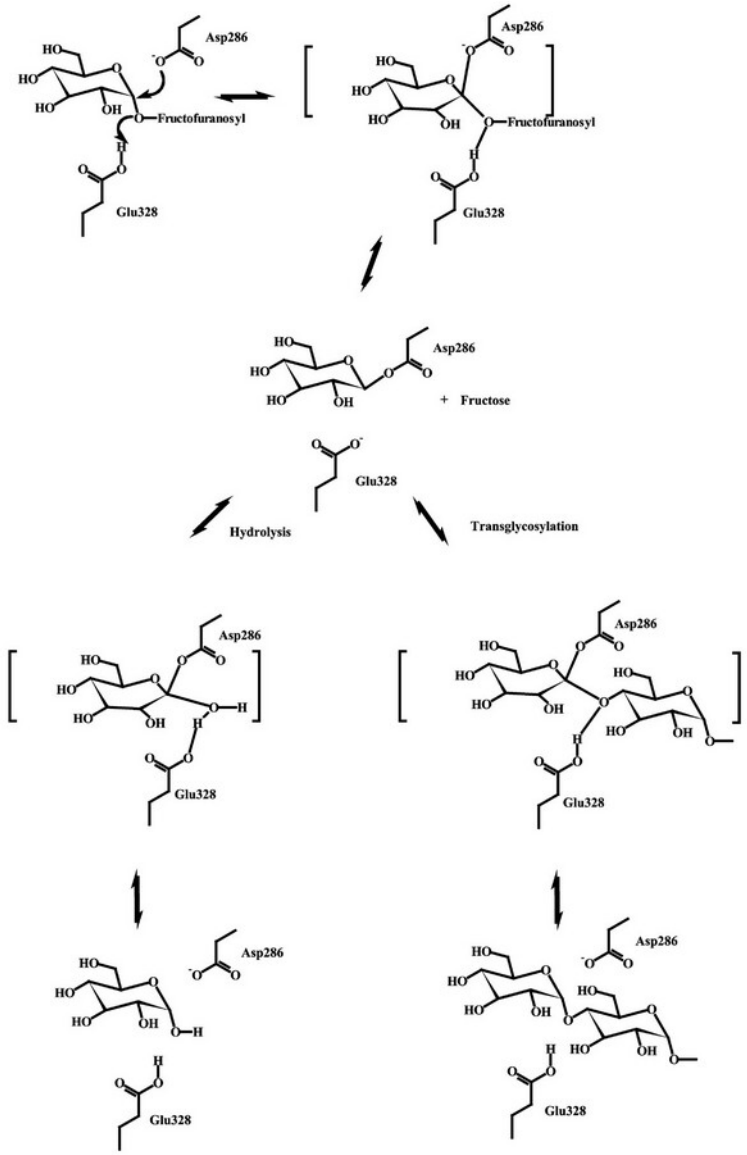

Figure 1.4. Transglycosylation and hydrolysis reaction pathway catalyzed by glycosidases (picture from (Skov *et al.* 2001))

1.7 STRUCTURES OF GLYCOSIDASES

1.7.1 Structure of GALases/NAGALases GH 27

Three-dimensional structures of human (Garman and Garboczi 2004; Lieberman *et al.* 2009), rice (Fujimoto *et al.* 2003), fungal (Fujimoto *et al.* 2009) GALase of GH 27 are available in the PDB database.

In these crystal structures GALase occur as monomers, dimers, trimers and tetramers. The human enzymes were found to be dimeric (Kusiak *et al.* 1978) or even tetrameric (Lieberman *et al.* 2009). Animal enzymes are found to be dimeric, too (Grossmann and Terra 2001).

GALases from plants are mostly studied from seeds and could function as monomers (Fujimoto *et al.* 2003), and tetramers (Del Campillo and Shannon 1982, Dey *et al.* 1983, Dey 1984).

Enzymes from bacteria are highly diverged and are found in all mentioned above forms: monomers (Post and Luebke 2005), dimers (Xiao *et al.* 2000), trimers (Goulas *et al.* 2009), tetramers (Fridjonsson *et al.* 1999). An interesting correlation between the enzyme size of and active pH was mentioned for bifidobacterias. An increase in acidity of the environment was accompanied by decrease in size of the native enzyme (Goulas *et al.* 2009).

Similarly to plants, fungal GALase are either monomers (Zeilinger *et al.* 1993, Miller *et al.* 2001) or tetramers (from *Mortierella vinacea* (Fujimoto *et al.* 2009), *Ganoderma lucidum* (Sripuan *et al.* 2003) *Aspergillus nidulans* (Ríos *et al.* 1993)). GALase isoforms Gal II and IV from *A. niger* (similar/identical to *agl B*) are found to be dimeric, while Gal III is tetrameric (Ademark *et al.* 2001).

Animal NAGALases are either monomers (Garman *et al.* 2002) or diers (Sung and Sweely 1980). Some other animal and bacterial enzymes are classified like multimers (Hoskins *et al.* 1997).

A few NAGALase structures are available in pdb, belonging to the GH109, GH 101 and GH 27. The structure of bacterial NAGALase GH 109 from *Flavobacterium meningosepticum* ATCC 33958 has only one described domain - Rossmann-fold NAD(P)H/NAD(P)(+) binding (NADB) domain. NAD+ is involved in the catalytic mechanism (Liu *et al.* 2007). Just bacterial NAGALases are reported for GH 109 (http://www.cazy.org/GH109_all.html). Enzymes of this family have no homology with other NAGALases.

NAGAlases from GH 101 are *endo*-enzymes with bulk structures. Solved structures of bacterial enzymes from *Streptococcus pneumoniae* (Caines et al. 2008) and *Bifidobacterium longum* (EngBF) (Suzuki et al. 2009) share a common conserved F5/8 type C domain (Baumgartner et al. 1998). Their sequence identity with other NAGALases is lower than 6% (http://www.rcsb.org/pdb/explore/structureCluster.do?structureId=2ZXQ).

For the GH 27 family, human (Clark and Garman 2009), chicken (Garman et al. 2002) and bacterial (*Bacillus halodurans* C-125 (not published)) are availables in PDB. Human and chicken NAGALases have very high sequence (51.9%) and 97% structure identity (http://www.rcsb.org/pdb/workbench/showPrecalcAlignment.do?action=pw_fatcat&pdb1=1KTB&chain1=A&pdb2=3HG3&chain2=A). Additionally they share a melibiase domain, common for GH 27 proteins.

The number of described NAGALases enzymes GH27 is limited: bacterial (Liu *et al.* 2007), human (Clark and Garman 2009), chicken (Garman *et al.* 2002) – only 6 NAGALases compared to more than 40 GALases.

The structures of GALase from rice and NAGALase from chicken GH27 are very similar with 1.49A root mean square deviation of the catalytic domain for C-alpha (Fujimoto *et al.* 2003). However, fungal GALase from *Trichoderma reesei* has a higher sequence similarity with chicken NAGALase than with rice GALase. Architecture of GALase from fungi differs by two longer loops, Trp 136-Ser 205 and Glu367-Trp352, close in space to the active site (Golubev *et al.* 2004).

Structurally GH 27 fold to a $(alpha/beta)_8$-barrel followed by another domain, often a beta-barrel or beta-sandwich, with an undescribed role (Golubev *et al.* 2004).

1.7.2 GALase/NAGALase from *A.niger*

Three genes encoding GALases in *A. niger* are described in the literature – so called *aglA*, *aglB* and *aglC* (Pel *et al.* 2007). Further in the text the gene names (*aglA*, *aglB* and *aglC*) are also used to name the corresponding proteins encoded by these genes. In these cases the word ‘gene’ is omitted. *AglC* has no significant homology with the other two (*aglA* and *aglC* has 32% identity over only 17 % of sequence) and is more related to the bacterial tetrameric enzymes GH36. Differences in the substrate specificities of *aglB* and *aglC* are connected to different physiological roles (Ademark *et al.* 2001; Manzanares *et al.* 1998).

A large screening study of extracellular NAGALase activity of a library of filamentous fungi (42 strains) revealed that the best producer is *A. niger* CCIM K2 (Weignerova *et al.* 2008). This is the first report of NAGALase enzyme activity in *A. niger*. However, in this organism no gene was identified to encode a NAGALase, so the exact sequence-function relationship could not be established and gene identification would be the first step in our work.

1.7.3 Structure of HEXs GH20

There are a couple of published crystal structures of bacterial HEXs - *Actinobacillus actinomycetemcomitans* (Ramasubbu *et al.* 2005), *Paenibacillus* sp. (Sumida *et al.* 2009), *Serratia marcescens* (Prag *et al.* 2000; Tews *et al.* 1996), *Streptococcus gordonii* (Langley *et al.* 2008), *Streptomyces plicatus* (Mark *et al.* 2001; Williams *et al.* 2002).

A first homology model of the alpha-subunit of HEX (HexA) from human has been published already in 1996 (Tews *et al.* 1996), right after the first crystal structure of a representative of family GH20 from bacteria *Serratia marcescens* has been reported. Structural work on human lysosomal HEXs HexA and HexB has been for a long time of a high interest, because mutations at these enzymes cause disorders (Sandhoff *et al.* 1968; Tay 1881). Finally, crystal structure of HexA has been released in 2006 (Lemieux *et al.* 2006) and Hex B in 2003 (Mark *et al.* 2003).

HEX from fungi has not been crystallized yet; however a first homology model of the *A. oryzae* enzymes active site has been suggested already in 2003 by our group (Husakova *et al.* 2003). At this early stage important aspects of the enzyme structure could not be addressed, as glycosylation, dimerization or the large loop forming the lid above the active site.

Animal HEXs exist as dimers. Human HEXs exist as a homodimers (HexB and S) (Maier *et al.* 2003) or heterodimers HexA (Lemieux *et al.* 2006). HEXs from plants exist in several different complex forms. Heterodimeric enzymes were purified from cabbage (Chang *et al.* 1998). HEXs from *Trigonella foenum-graecum* can form trimers, pentamers and hexomers (Bouquelet and Spik 1978). Bacterial HEXs act as monomers and have long loops in vicinity of the active site (Mark *et al.* 2001; Prag *et al.* 2000), while human enzymes don’t have corresponding long loops. The fungal enzymes occur as dimer. HEX from *P. oxalicum* has been also annotated as dimeric (Yamamoto *et al.* 1985).

A common feature of all known structures of HEXs is the $(\beta/\alpha)_8$-barrel (TIM-barrel) architecture of the catalytic domain that houses the active site (Figure 1.5).

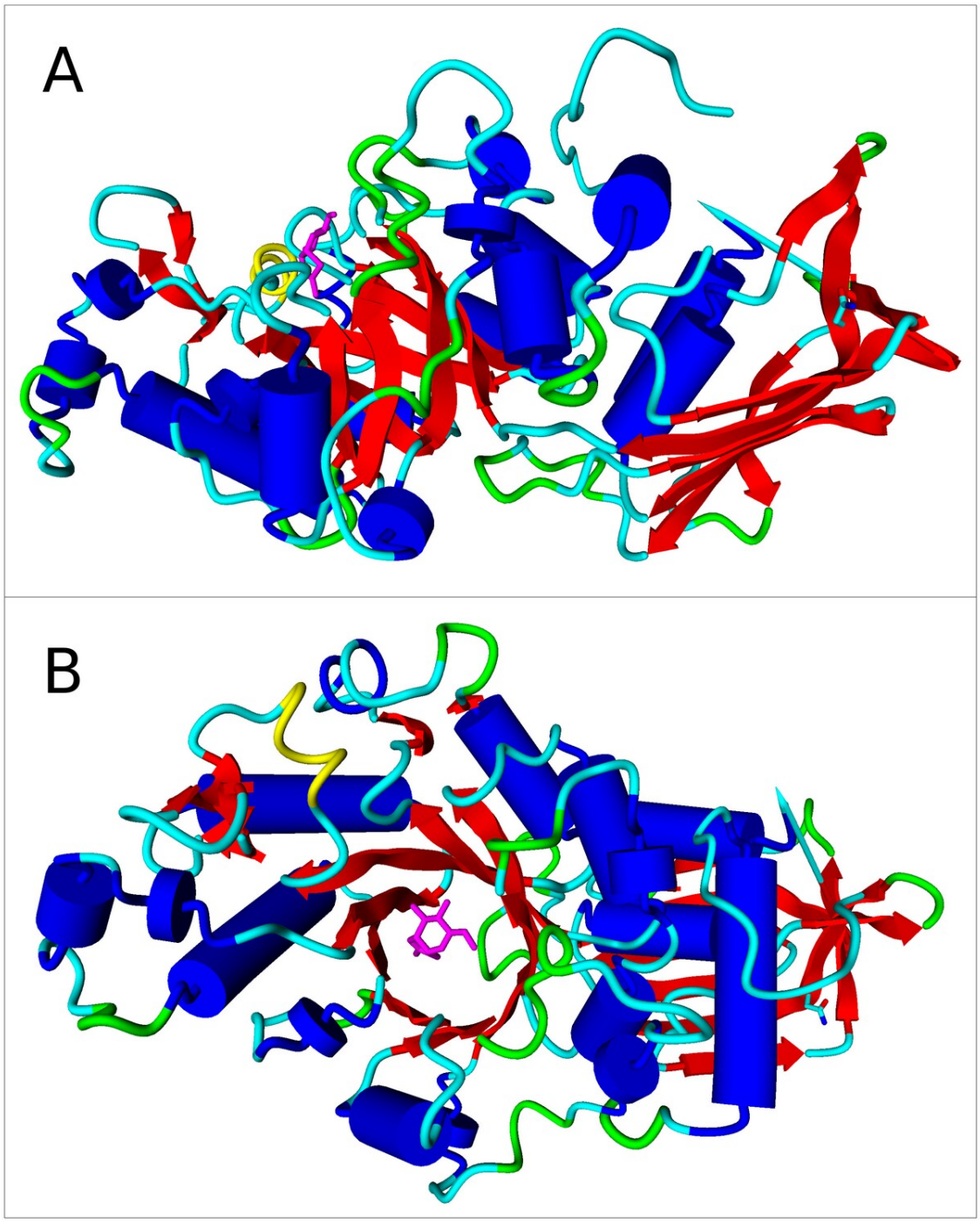

Figure 1.5. Cartoon representation of one monomer of HEX from human (Hex B; PDB-code is 1npo) with inhibitor NAG-thiazoline (magenta) in the active site prepared in Yasara (Krieger *et al.* 2004): A-side view, B- top view.

1.8 ACTIVE SITE

A common feature of all studied HEXs, GALases and NAGALases is the $(\beta/\alpha)_8$-barrel (TIM-barrel) forming the catalytic domain. The topology of the catalytic domains is shown in Figure 1.6.

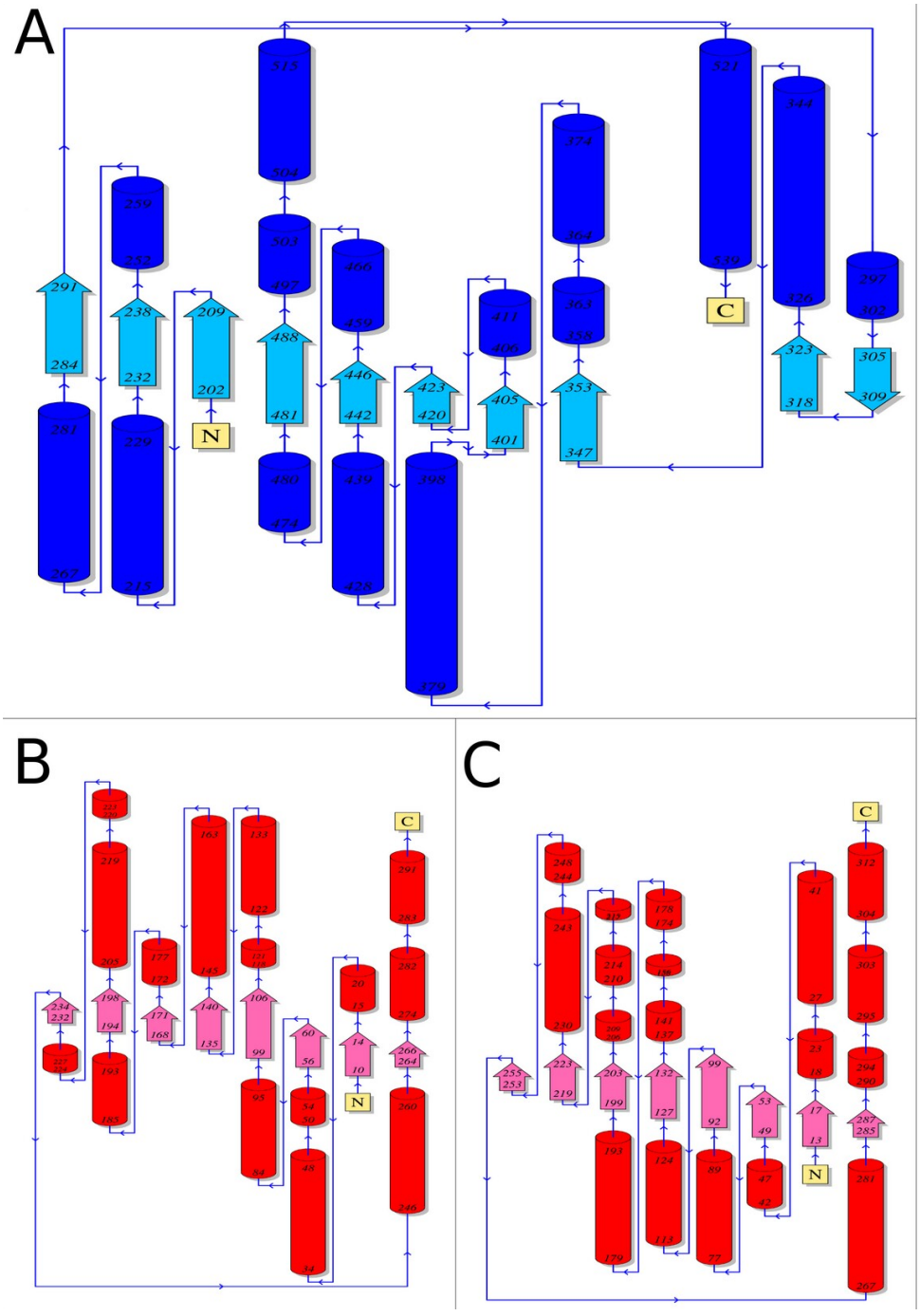

Figure 1.6. Topology of catalytic domain of enzymes downloaded from PDBsum server. Arrows are depicts beta-sheets, cylinders - alpha-helixes: A. HEX from human PDB-code 1now; B. NAGALase from chicken 1ktb: C. GALase from fungi PDB-code 1szn.

1.8.1 Active site of GALase and NAGALase

Active site amino acids of two classes of enzymes are very similar except the region in the vicinity of the N-acetyl group (Garman *et al.* 2002). Ser 172 and Ala 175 in chicken are substituted by Cys 162 and Trp 164 in rice (Fujimoto *et al.* 2003) – Figure 1.7.

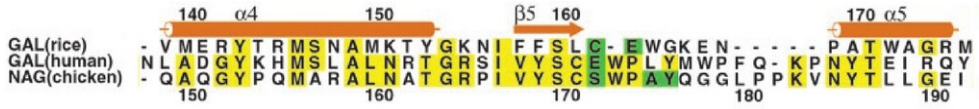

Figure 1.7. Part of sequence alignment of rice and human GALases and chicken NAGALase with important differences in the active site (picture is adjusted from Fujimoto *et al.* 2003).

Active site amino acid residues of rice GALase are shown in figure 1.8 (Fujimoto *et al.* 2003). Asp 185 acts as acid/base and Asp 130 plays a role of nucleophile. Trp 16, Cys 101 and Met 217 make a hydrophobic contact with a-galactose. Trp 16 is situated in parallel plane to the O3, C3-C6 and O6. Other active site residues – Arg 181, Trp 164, Cys 162, Asp 185, Lys 128, Asp 52 and Asp 51 – could form HB with substrate (Fujimoto *et al.* 2003).

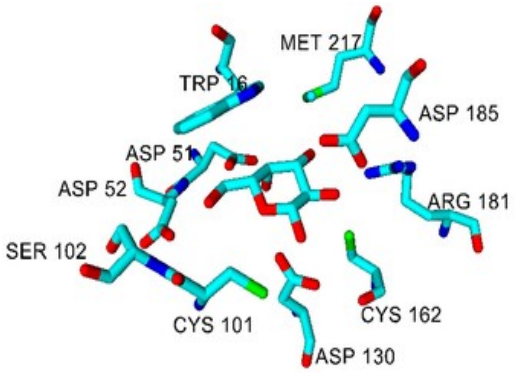

Figure 1.8. Stick representation of active site amino acids of GALase from rice (PDB code is 1uas) with alpha-galactose prepared in Yasara (Krieger *et al.* 2004).

Asp 51 and Trp 16 are responsible for glucose/galactose discrimination in galactose and galactosaminidase (Golubev 2004, Garman 2002).

1.8.2 Active site of HEX

Active site amino acid residues of HEXs identified in the bacterial structure (Tews *et al.* 1996) with chitobiase is shown in the Figure 1.9.

The role of catalytic Asp 539 is in orientation of the C2-acetamido group and stabilization of the oxazolinium ion intermediate (Williams *et al.* 2002), however, the true position of the catalytic residues has not been well described (Mark *et al.* 2003). This role

was confirmed by numerous mutagenesis and kinetic experiments (Mark *et al.* 2003; Williams *et al.* 2002). Asp 539 could directly form hydrogen bonding interaction with the N2 of the non-reducind sugar. The proposed roles of negatively charged Asp 539 are in the stabilization of the positively charged oxazoline ring and electrostatic interaction with Glu (Prag *et al.* 2000). Close connection with catalytic Glu 540 is also not random (Prag *et al.* 2000).

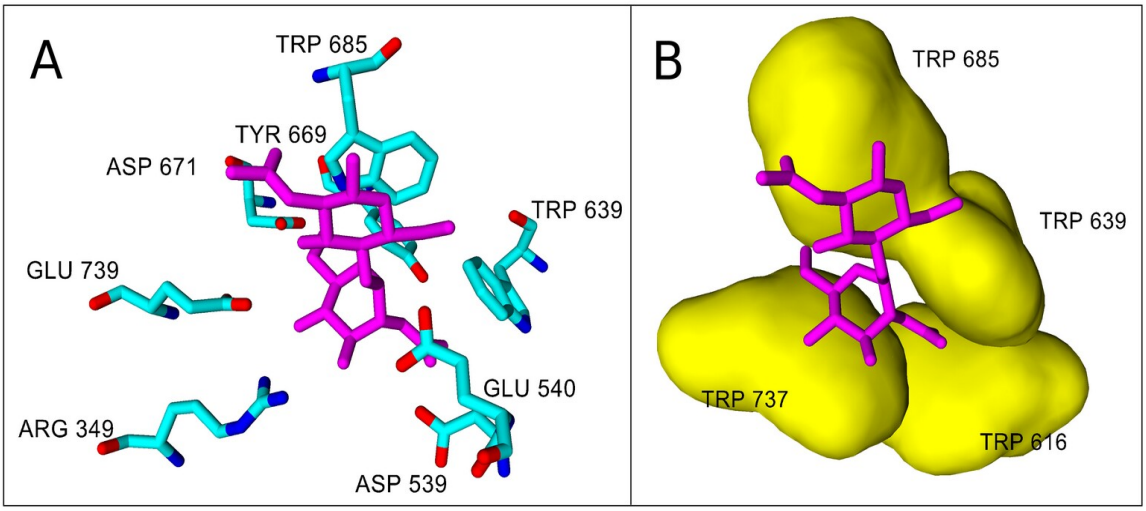

Figure 1.9. Active site amino acids of HEX from *Serratia marcescens* (PDB-code 1qba) with chitobiase from 1c7s (Prag *et al.* 2000) (magenta) prepared in Yasara (Krieger *et al.* 2004): A. Stick representation of important amino acids; B. Hydrophobic pocket formed by Trp residues depicted as molecular surface representation of them.

Trp 616, Trp 639, and Trp 737 form a hydrophobic pocket (yellow color in figure 1.9) for the non-reducing sugar, which results in the acetamido group getting closer to the anomeric carbon and also protecting the intermediate state from unwanted interactions (Mark *et al.* 2001). Enzymes with side-chain mutations of Trp 639 and Trp 737 to Tyr lost catalytic activity (Manuel *et al.* 2007). Tyr 669 supports the oxazolinium ion formation by hydrogen bonding with acetyl group (Tews *et al.* 1997).

Arg 349 stabilizes the substrate in the active site by hydrogen bonding (Hou *et al.* 2000; Mark *et al.* 2001; Matsuzawa *et al.* 2003; Tews *et al.* 1997).

QM/MM modeling of different protonation states of Asp 671 and Glu 739 was applied to study oxazolinium ion formation; this study had mostly informative character (Greig *et al.* 2008).

1.9.1 SUBSTRATES OF GALase/NAGALase

Unnatural and modified substrates of GALase and NAGALase are not extenively studied, and the role of active site amino acids residues is not yet studied in detail.

Crystallization data showed that the beta-anomeric configuration of the non-reducing oligosaccharide could not be cleaved by GALase due to a sterical conflict caused by the orientation of the anomeric oxygen at the bottom of the active site. However, monomer beta-galactose could act as a competitive inhibitor, because it was found in two times higher ratio, than alpha-anomer in crystallographic experiments with rice (Fujimoto *et al.* 2003).

1.9.2 SUBSTRATES OF HEX

Substrates for HEXs are a good alternative for a selective preparation of target products. Based on the idea that glycosidases, unable to cleave certain glycosidic linkages, are often unable to synthesize the same type, hydrolysis of substrates was tested before transglycosylation (Husakova *et al.* 2003, Loft *et al.* 2009).

Substrates modified at C-1, C-2, C-6, sulfated substrates (6-*O*-sulfate, 4-*O*-sulfate and 3-*O*-sulfate) and substrates carrying a ManNAc (N-acetylmannose) moiety were tested with fungal HEXs.

The first report of a transglycosylation reaction catalysed by HEXs using a modified glycosyl donor (acetylated at C-6) was by Husakova (Husakova *et al.* 2001).

The transglycosylation potential of fungal HEX is nicely demonstrated by the usage of *N*-acyl-modified substrates at C-2 carbon. Although the acetamido group is a critical structural feature of the substrates, the enzyme tolerates certain changes at C-2 (shorter or longer acyls, a hydroxyl instead of hydrogen) but not highly electronegative acyls, such as trifluoracetyl, sulfate or a free amino group. The glycosylation using *p*-nitrophenyl 2-deoxy-2-formamido-β-D-glucopyranoside, *p*-nitrophenyl 2-deoxy-2-glycoloylamido-β-D-glucopyranoside and *p*-nitrophenyl 2-deoxy-2-propionamido-β-D-glucopyranoside, catalysed by HEX from *Talaromyces flavus, A. oryzae* and *P. oxalicum*, resulted in *N*-acyl-modified disaccharides exclusively containing $\beta(1\rightarrow4)$ bonds in reasonable yields (Fialova *et al.* 2004).

Despite the interesting biological activity of the ManNAc moiety (ligand for the NK cell activating protein NKR-P1 (Krist *et al.* 2001)) preparation of its derivatives is not studied well and often not successful. Fungal HEXs (for instance *P. oxalicum*) showed a low level of monosaccharide formation from GlcNAc(1->4)ManNAc and GlcNAc(1-

>6)ManNAc, some of them (from *A. oryzae*) – didn't accept it as substrate at all (Husakova *et al.* 2003).

Different successful modifications at C-6 position were reported.

C-6 modified by acetyl group (*p*-nitrophenyl 6-*O*-acetyl-b-D-2-deoxy-2-acetamidoglucopyranoside) is accepted by *P .oxalicum* as substrate for hydrolysis (12.1%) and the transglycosylation reaction, being a poor substrate for *A. oryzae* (Husakova *et al.* 2001).

Aldehyde group at C-6 is an example of a well-accepted modification, that could be used for preparation of immunoactive disaccharides (Fialova *et al.*, 2005a).

Azide substitutes at the C-1 atom glucopyranosides were accepted as good substrates for hydrolysis and transglycosylation, while 2-acetamido-2-deoxy-b-D-galactopyranosyl azide was not hydrolised by any tested fungal enzyme. Such specificity was explained by computational study as the result of unfavorable binding energy due to the lost hydrogen bond with Glu 519 (Fialova *et al.* 2005b).

Regioselectively sulfated *p*-nitrophenyl 2-deoxy-β-D-glucosaminides (6-*O*-sulfate, 4-*O*-sulfate and 3-*O*-sulfate) are relatively bad substrates for sulfatases but good enough for HEXs from various fungal sources (genera *Aspergillus*, *Penicillium*, *Fusarium*, *Talaromyces*, *Trichoderma*) (Loft *et al.* 2009). While being a weak substrate for hydrolysis (Loft *et al.* 2009) 6-O-sulfo-N-acetyl-beta-D-glucosaminide is a good substrate for transglycosylation reactions (Ogata *et al.* 2007; Uzawa *et al.* 2003; Zeng *et al.* 2007). The structural background of different specificities of fungal HEXs has not been described yet.

The sulfated oligosaccharides play an important role in cell-cell adhesion, homing lymphocytes, bacteria binding, cancer metastasis, hormone regulation and some other biological processes. Sulfated sugars have attracted special attention mainly because sulfosugars "construct key recognition structures of natural ligands of selectins, pathogenic bacteria and viruses, and other carbohydrate receptor proteins" (Uzawa *et al.* 2003). The sulfo-group contains an electrophile part and could bind to proteins and carbohydrates tightly; it is not allergic component, as other sulfocomponents (http://home.att.net/~steinert/wwii.htm).

Zeng (Zeng *et al.* 2007) found that in the reaction of transglycosylation catalyzed by HEX from *A. oryzae* the final product depends on the ratio of used substrates. They were able to synthesize sulfo-containing disaccharide with a reasonable yield (13%), but *A. oryzae* prefers *p*-nitrophenyl 2-deoxy-β-D-glucosaminides over sulphated substrates.

A later study (Ogata *et al.* 2007) continued investigations of production of p-nitrophenyl sulfated disaccharides using 6-O-sulfo-N-acetyl-beta-D-glucosaminide as

donor. Modifications at the aglycon part were by hydroxyl (OH), methyl (CH3), propyl (CH2CH2CH3), allyl (CH2CH=CH2) and p-nitrophenyl (pNP; C6H4-NO2) groups.

As known from the work of Kaplan, the ability of human HexA to cleave 6-sulfated hexosamine is mostly determined by alpha AsnArg424 and the inhibition is caused by beta AspLeu453 (Kaplan *et al.* 2001). The structural features in human HEX HexA were studied by site-directed mutagenesis and revealed that the alpha-subunit is responsible for binding different negatively charged substrates through Arg424. Sharma *et al*. however concluded that also other differences in HexA and HexB sequences exist (Sharma *et al.* 2003), determining substrate specificity.

HEXs from *A. oryzae* and *P. oxalicum* showed different activity towards modified substrates. *P. oxalicum* better accept C-2 modified, 6-O-acylated, 4-O-sulfo and 6-O-sulfo substrates for hydrolysis (Fialova *et al.* 2004; Husakova *et al.* 2001; Loft *et al.* 2009).

Recently, the X-ray structure of a prokaryotic HEX capable of hydrolyzing glycosphingolipids was report (Sumida *et al.* 2009).

2. CARBOHYDRATES AS AN OBJECT FOR BIOINFORMATICS STUDY

Carbohydrates play an important role in the nature. They are found as simple sugars (mono and disaccharides), oligosaccharides and polysaccharides. They as an important part of glycoproteins and glycogphingolipids participate in cell-cell recognition, post-translational modifications of proteins.

2.1 STRUCTURE OF CARBOHYDRATES

Structurally carbohydrates are characterized by the presence of hydrophobic carbons and hydrophilic carboxyl-groups. The open-chain form of a monosaccharide often coexists with a closed ring. Monosaccharides forming a five-sided ring, like ribose, are called furanoses. Those forming six-sided rings, like glucose, are called pyranoses (Figure 2.1 with numeration for atoms in the ring). The ring form is favored in aqueous solutions (http://www.scientificpsychic.com/fitness/carbohydrates.html).

Two (or more) saccharides can be connected by a glycosidic bond in a condensation reaction. The glycosidic bond can be formed between any hydroxyl group on the component monosaccharide (Figure 2.1).

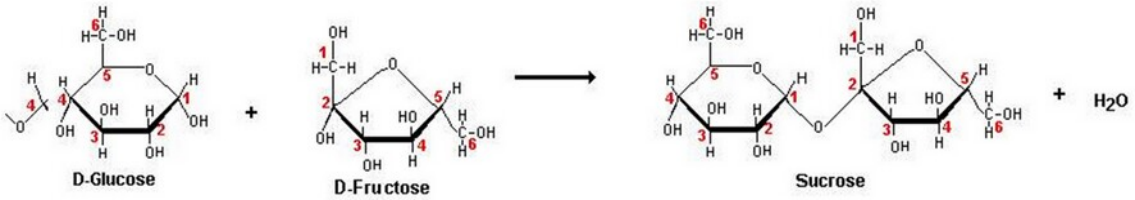

Figure 2.1. Glycosidic bond formation in condensation reaction (picture is from http://www.scientificpsychic.com/fitness/carbohydrates.html). Atom numeration in glucose ring molecule is shown.

This leads to the big number of isomers: regioisomers (different position of bond) and stereoisomers (anomers of glucose: *alpha* - glucose when the hydroxyl group is on the opposite side of the -CH_2OH group, or *beta*-glucose when the hydroxyl group is on the same side as the -CH_2OH group) (Stick and Williams 2008).

The geometry description of the pyranose ring includes 38 ideal conformations (Berces *et al.* 2001; Stoddart 1971). It can be done properly just for the crystal structures or by computational modeling. All pyranose ring conformation can be clustered by boat, twist-boat, twist-chair and chair conformations. The twist-boat conformation exists in equilibrium with boat. One chair conformation is able to interconvert in the alternative.

This proceeds through some intermediate steps (Miljkovic 2009; http://www2.chemistry.msu.edu/faculty/reusch/ VirtTxtJml/sterism2.htm) – Figure 2.2.

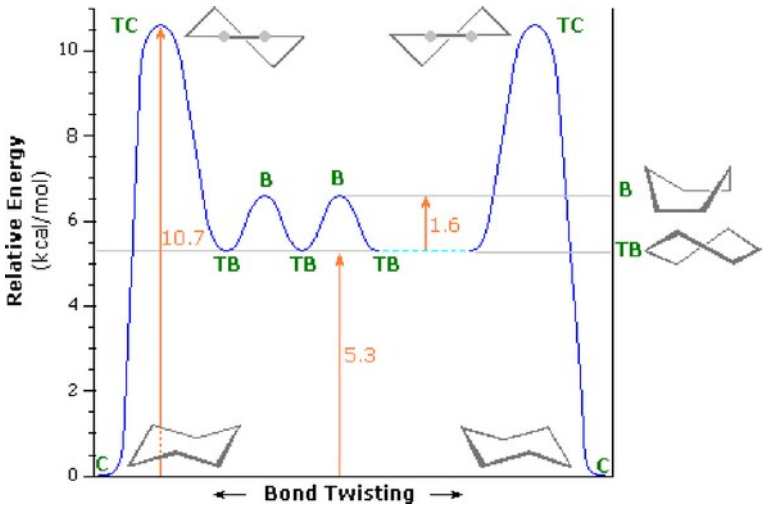

Figure 2.2. Energy diagram of conformation energy in cyclohexane (from http://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/sterism2.htm): twist-chair form (TC), twist-boat (TB), boat (B) and chair (C).

The boat form of the pyranose ring is approximately 17 kJ/mol (~4 kcal/mol) less stable than the chair form (Stick and Williams, 2008). Conformational flexibility of the ring formation leads to difficulties in proper modeling (Laederach and Reilly 2005).

2.2 PROTEIN-CARBOHYDRATE INTERACTION

This interaction is relatively weak. Some important facts about it:

-even though carbohydrates are hydrophilic, hydrophobic interaction plays an important role (Levitt and Perutz 1988; Sigurskjold and Bundle 1992);

-hydrogen bonds determine protein-carbohydrate interaction (Quiocho 1993);

-this binding is often water-mediated (Imberty *et al.* 2000, Lemieux *et al.* 1991, Lütteke 2009).

3. NITRILASES

3.1 NITRILASE SUPERFAMILY: DESCRIPTION, CATALYZED REACTIONS

Nitrilases are enzymes, which catalyze the hydrolysis of nitriles to carboxylic acids and ammonia, without the formation of a "free" amide intermediates (http://en.wikipedia.com). They are found in plants, animals, fungi and certain prokaryotes.

Nitrilases belong to the nitrilase superfamily, which combines 13 branches of enzymes based on sequence and substrates of enzymes (Pace and Brenner 2001): nitrilase (aromatic); aliphatic amidase; amino-terminal amidase; biotinidase; beta-ureidopropionase; carbamylase; 2 groups of glutamine-dependent NAD synthetases; apolipoprotein N-acyltransferase; Nit; branches 11-13.

A proposed earlier catalytic mechanism for nitrilases is shown in the Figure 3.1 (Kobayashi *et al.* 1998), where EXH – is an enzyme with sulfur group (X=S). At route (iii) an enzyme-thioimidate intermediate is created. (I) is a tetrahedral intermediate. Ammonium is then removed from this intermediate (route iv). In the next step the complex is hydrolised by an activated water molecule to produce acid (route v). However, under some specific conditions a reverse reaction (routes ii and i) can take place.

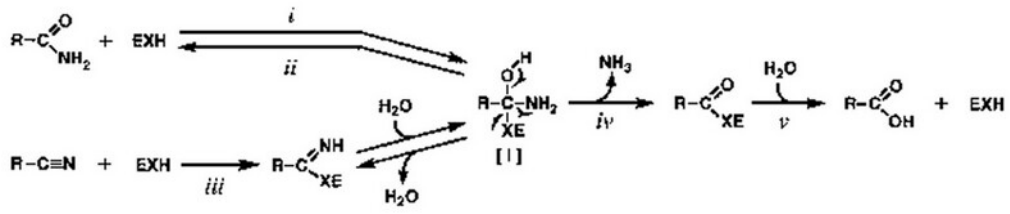

Figure 3.1. Catalytic mechanism of nitrilases (picture from (Kobayashi *et al.* 1998)).

In the later work (Pace and Brenner 2001) four types of reaction for different branches of nitrilase superfamily enzymes are described. Reaction catalyzed by branch 1 nitrilase superfamily (aromatic) and accepted catalytic mechanism are shown in figure 3.2.

Members of this superfamily have a conserved domain, difference in conformation is found in N- and C-termini, which can carry some additional domains (Pace 2001).

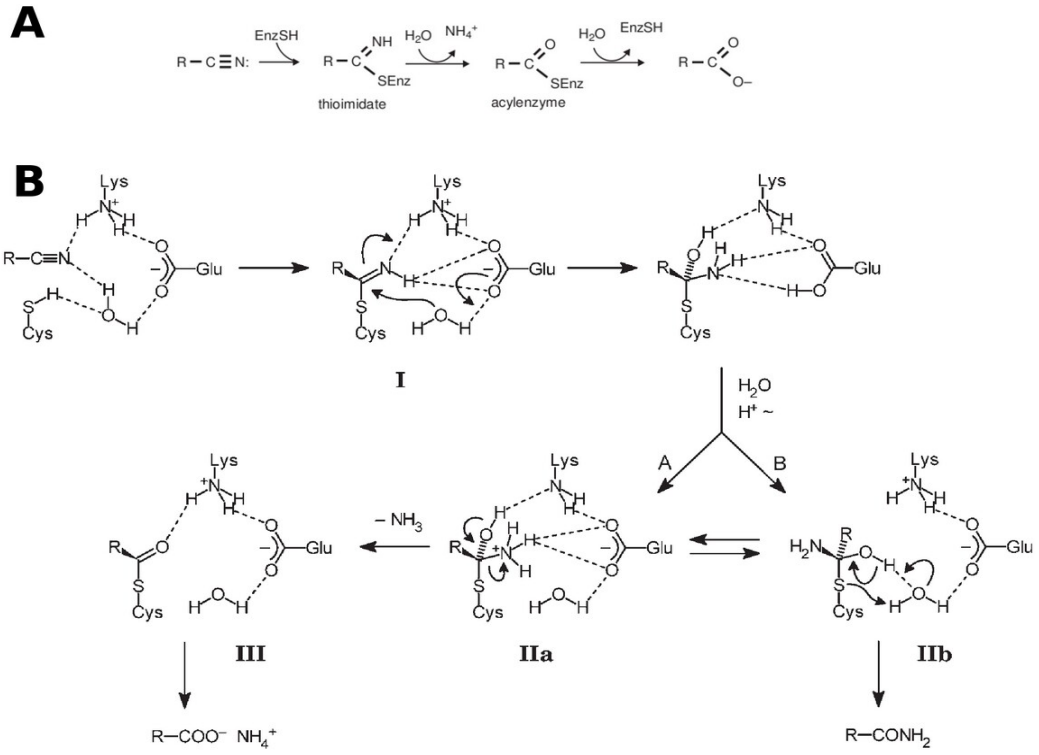

Figure 3.2. A. The nitrilase branch 1 reaction. In plants, the substrate is indole-3-acetonitrile and the product is indole-3-acetic acid (picture from (Pace and Brenner 2001)). B. Proposed catalytic mechanism of aromatic nitrilases (from Fernandes *et al*. 2006).

3.2 AROMATIC NITRILASES – CHARACTERISTIC, ROLE, STRUCTURE

Aromatic nitrilases belong to the branch 1 of nitrilase superfamily, EC 3.5.5.1.

Nitrilases are attractive for organic chemical processing because of the mild conditions, quantitative yields, absence of by-products, and, in some cases, enantio- or regioselectivity that result from their reactions (Kobayashi *et al.* 1990; Polaina and MacCabe 2007). They found great application in industrial production of different amino-containing compounds (acrylamide, nicotinamide) and some chemically important intermediates (Polaina and MacCabe 2007).

Nitrilases are mostly purified and described in bacteria (Harper 1977; Kobayashi *et al.* 1992; O'Reilly and Turner 2003; Yeom *et al.* 2008). Nitrilases can exist as multimers, making so called fibers visible in micrographs.

The is no solved crystal structure of aromatic nitrilases reported to date; however, a model of nitrilase from *Rhodococcus rhodochrous* J1 was proposed (Thuku *et al.* 2007). Four proteins from the nitrilase superfamily were used as templates, PDB codes 1ems (Pace *et al.* 2000), 1erz (Nakai *et al.* 2000), 1f89 (Kumaran *et al.* 2003) and 1j31 (Sakai *et al.* 2004). In bacterial nitrilase from *R. rhodochrous* J1 dimers were reported, that then multimerize to form a large helical multimer (fiber) with 4.9 dimers per turn of the helix

(Thuku et al 2007). The C-terminal region in the multimer model of *R. rhodochrous* J1 is located on the inside of the helix. They characterize some contact surfaces in helix: monomer-monomer surface in dimers (they called it 'A'), dimer-dimer ('C') and surface between turns of helix ('D').

The full-length enzyme is unable to form fibers. Based on this observation Thuku *et al.* proposed that sterical hindrance during multimer formation from full-length enzyme leads to a loss of some interactions in helix, particularly the loss of 'D' surface contacts. The importance of conservation in surface contacts was confirmed by a mutagenesis study, when mutants formed just short 'C'-shape helixes (Thuku *et al.* 2007).

The structure of monomer and active site with catalytic Cys169, Glu54 and Lys127 (catalytic triad in the nitrilase superfamily) amino acids, conserved in the nitrilase superfamily, based on the structure of nucleotide-binding protein Fhit (NitFhit) (PDB code 1ems) (Pace *et al.* 2000) is presented in the Figure 3.3A.

Mutation of Cys 169 by Ser and Ala in *Rhodococcus* gene encoding nitrilase was studied. No significant changes in enzyme conformations were observed, though the mutations are accompanied with a loss of activity (Kobayashi *et al.* 1992).

The residue at position Tyr 142 in nitrilase from *R. rhodochrous* ATCC 33278 was found to be responsible for discrimination between aromatic and aliphatic substrates and acts as an electron acceptor (Yeom *et al.* 2008). This residue is substituted by Ser147 in the pdb structure 1ems (Thuku *et al.* 2007). A model of nitrilase from *R. rhodochrous* ATCC 33278 with bound substrate is shown in figure 3.3B (Yeom *et al.* 2008).

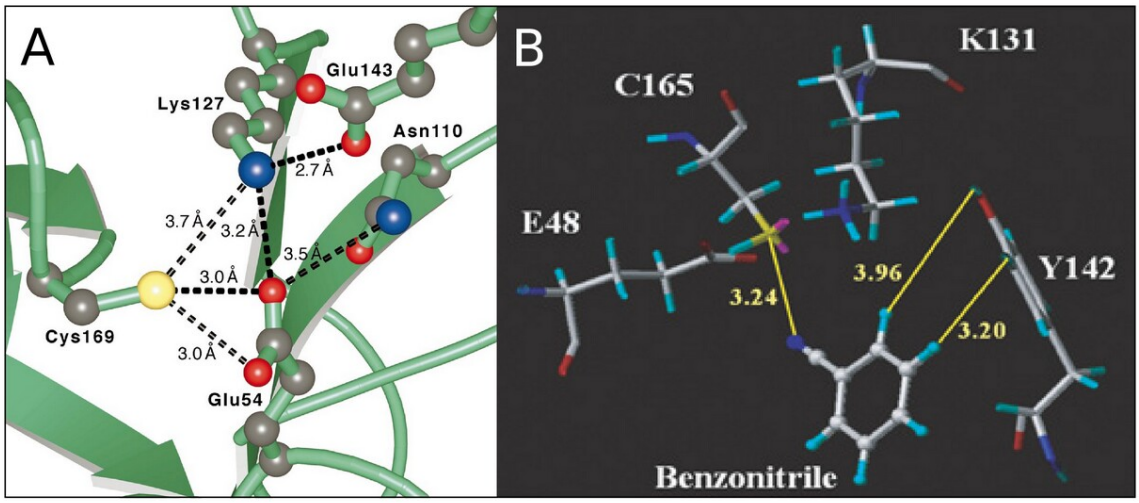

Figure 3.3.A. Active site amino acids of NitFhit protein with catalytic triad - Cys 169, Glu54 and Lys127 (picture from Pace *et al.* 2000). B. Model of the active site of *R.rhodochrous* ATCC 33278 with bound benzonitrile by Surflex docking program (picture from Yeom *et al.* 2008).

3.2.1 Fungal nitrilases

Fungal nitrilases appear promising for biocatalytic applications and biodegradation of nitrile environmental contaminants (Snajdrova *et al.* 2004; Winkler *et al.* 2009).

Only a few fungal aromatic nitrilases are purified and characterized to date – from *Fussarium solani* and *Fusarium oxysporum f. sp. melonis* (David *et al.* 1977; Goldlust and Bohak 1989; Singh *et al.* 2006; Vejvoda *et al.* 2006). Nitrilase from *A. niger* was partly characterized in 2006 and showed different substrate affinity to tested substrates (Kaplan *et al.* 2006). It was found to be a multimer like other fungal nitrilases: *Fusarium oxysporum f.sp. melonis* nitrilase can have 14-15 subunits and *Fusarium solani* IMI196840 – 8 (Harper 1977).

4. AIMS OF THIS WORK

1. Identification of genes responsible for GALase and NAGALase activity in *A. niger*.
2. Three-dimensional modeling of the structure of GALase and NAGALase encoded by genes *aglA* and *aglB* from *A.niger*.
3. Complete structural analysis of HEX from *A. oryzae* including glycosylation and multimerization and their effect on protein function and stability.
4. Docking followed by molecular dynamics simulation of modified substrates, which are biotechnologically important, to assess their potential use in biotransformation; explanation in molecular terms of their properties as substrates or inhibitors of HEX from *A. oryzae.*
5. Adopt modeling procedure used for *A. oryzae* enzyme to other fungal HEX from *P. oxalicum* to explain difference in observed experimental kinetic behavior on a structural basis.
6. Characterization of the enzyme responsible for nitrilase activity from *A. niger*.

5. REFERENCES

Ademark P, Larsson M, Tjerneld F, Stålbrand H. 2001. Multiple alpha-galactosidases from *Aspergillus niger*: purification, characterisation, and substrate specificities. Enzyme and microbial technology. 29:441-448.

Ademark P, de Vries RP, Hägglund P, Stålbrand H, Visser J. 2001. Cloning and characterization of *Aspergillus niger* genes encoding an alpha-galactosidase and a beta-mannosidase involved in galactomannan degradation. Eur J Biochem. 268:2982-2990.

Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. 1998. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. Protein Sci. 7(7):1626-31.

Berces A, Whitfield DM, Nukada T. 2001. Quantitative description of six-membered ring conformations following the IUPAC conformational nomenclature. ChemInform. 32: no.

Bojarova P, Kren V. 2009. Glycosidases: a key to tailored carbohydrates. Trends in Biotechnology. 27(4):199-209.

Bouquelet S, Spik G. 1978. Properties of four molecular forms of N-acetyl-beta-D-hexosaminidase isolated from germinating seeds of fenugreek (*Trigonella foenum graecum*). Eur J Biochem. 84(2):551-559.

Brady D. 2009. Biocatalytic hydrolysis of nitriles. In Anastas PT, Crabtree RH, editors. Handbook of green chemistry: green catalysis. Vol.3. Biocatalysis. Weinheim: Wiley. 27-49.

Caines ME, Zhu H, Vuckovic M, Willis LM, Withers SG, Wakarchuk WW, Strynadka NC. 2008. The structural basis for T-antigen hydrolysis by Streptococcus pneumoniae: a target for structure-based vaccine design. J Biol Chem. 283(46):31279-83.

Calhoun DH, Bishop DF, Bernstein HS, Quinn M, Hantzopoulos P, Desnick RJ. 1985. Fabry disease: isolation of a cDNA clone encoding human alpha-galactosidase A. Proc Natl Acad Sci U S A. 82(21):7364–7368.

Callahan JW, Lassila EL, Tandt WD, Phillippart M. 1973. Alpha-N-acetylgalactosaminidase: isolation, properties and distribution of the human enzyme. Biochem Med. 7:424—431.

Carmona T, Fialova P, Kren V, Ettrich R, Martinkova L, Moreno-Vargas AJ, González C, Robina I. 2006. Cyanodeoxy-Glycosyl derivatives as substrates for enzymatic reactions. Eur. J. Org. Chem. 2006:1876–1885.

Chang CT, Young FP, Chang MH, Sung HY. 1998. Purification and properties of beta-N-Acetylhexosaminidase from cabbage. Biochem Mol Biol Int. 45(2):371-380.

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

Crout DHG, Singh S, Swoboda BEP, Critchley P, Gibson WT. 1992. Biotransformation in carbohydrate synthesis. N-Acetylgalactosaminyl transfer on to methyl N-acetyl-beta-D-glucosaminide (methyl 2-acetamido-2-deoxy-alpha-D-glucopyranoside) catalysed by a beta-N-acetylgalactosaminidase from *Aspergillus oryzae*. J Chem Soc Chem Commun. 704 -705.

Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. Structure. 3:853-859.

Del Campillo E, Shannon LM. 1982. An alpha-Galactosidase with hemagglutinin properties from soybean seeds. Plant Physiol. 69(3):628-31.

Dey PM. 1984. Characteristic features of an alpha-galactosidase from mung beans. Eur J Biochem. 140(2):385-390.

Dey PM, Del Campillo EM, Lezica RP. 1983. Characterization of a glycoprotein alpha-galactosidase from lentil seeds (Lens culinaris). J Biol Chem. 258(2):923-929.

Drouillard S, Armand S, Davies GJ, Vorgias CE, Henrissat B. 1997. *Serratia marcescens* chitobiase is a retaining glycosidase utilizing substrate acetamido group participation. Biochem J. 328:945–949.

Elhalabi JM, Rice KG. 1999. Synthesis and applications for unnatural sugar nucleotides. Curr Med Chem. 6(2):93-116. Review.

Fernandes BC, Mateo C, Kiziak C, Chmura A, Wacker J, van Rantwijk F, Stolz A and Sheldon R. 2006. Nitrile Hydratase Activity of a Recombinant Nitrilase. Advanced Synthesis & Catalysis, 348:2597–2603.

Fialova P, Carmona AT, Robina I, Ettrich R, Sedmera P, Prikrylova V, Husakova L, Kren V. 2005b. Glycosyl azides – Novel substrates for enzymatic transglycosylations. Tetrahedron Letters. 46:8715–8718.

Fialova P, Namdjou D-J, Ettrich R, Prikrylova V, Rauvolfova J, Krenek K, Kuzma M, Elling L, Bezouska K, Kren V. 2005a. Combined application of galactose oxidase

and β-*N*-acetylhexosaminidase in the synthesis of complex immunoactive *N*-acetylgalactosaminides. Advanced Synthesis & Catalysis 347(7–8): 997–1006.

Fialova P, Weignerova L, Rauvolfova J, Prikrylova V, Pisvejcova A, Ettrich R, Kuzma M, Sedmera P, Kren V. 2004. Hydrolytic and transglycosylation reactions of N-acyl modified substrates catalysed by β-N-acetylhexosaminidases. Tetrahedron. 60(3):693–701.

Frank M, Schloissnig S. 2010. Bioinformatics and molecular modeling in glycobiology. Cell Mol Life Sci. 67(16):2749-2772. Review.

Fridjonsson O, Watzlawick H, Gehweiler A, Mattes R. 1999. Thermostable alpha-galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization. FEMS Microbiol Lett. 176(1):147-153.

Fujimoto Z, Kaneko S, Kim WD, Park GG, Momma M, Kobayashi H. 2009. The tetramer structure of the glycoside hydrolase family 27 alpha-galactosidase I from *Umbelopsis vinacea*. Biosci Biotechnol Biochem. 73(10):2360-2364.

Fujimoto Z, Kaneko S, Momma M, Kobayashi H, Mizuno H. 2003. Crystal structure of rice alpha-galactosidase complexed with D-galactose. J Biol Chem. 278:20313-20318.

Garman SC, Garboczi DN. 2004. The molecular defect leading to Fabry disease: structure of human alpha-galactosidase. J Mol Biol. 337(2):319-335.

Garman SC, Hannick L, Zhu A, Garboczi DN. 2002. The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. Structure. 10(3):425-434.

Goldlust A, Bohak Z. 1989. Induction, purification, and characterization of the nitrilase of *Fusarium oxysporum f.sp. melonis*. Biotechnology and Applied Biochemistry. 11:581–601.

Golubev AM, Nagem RA, Brandão Neto JR, Neustroev KN, Eneyskaya EV, Kulminskaya AA, Shabalin KA, Savel'ev AN, Polikarpov I. 2004. Crystal structure of alpha-galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism. J Mol Biol. 339(2):413-22.

Gooday GW, Zhu W-Y, O'Donell RW. 1992. What are the roles of chitinases in the growing fungus? FEMS Microbiol. Lett. 100:387-392.

Goulas T, Goulas A, Tzortzis G, Gibson GR. 2009. A novel alpha-galactosidase from *Bifidobacterium bifidum* with transgalactosylating properties: gene molecular cloning and heterologous expression. Appl Microbiol Biotechnol. 82(3):471-477.

Greig IR, Zahariev F, Withers SG. 2008. Elucidating the nature of the *Streptomyces plicatus* beta-hexosaminidase-bound intermediate using *ab initio* molecular dynamics simulations. J Am Chem Soc. 130(51):17620-17628.

Grossmann GA, Terra WR. 2001. Alpha-galactosidases from the larval midgut of Tenebrio molitor (Coleoptera) and Spodoptera frugiperda (Lepidoptera). Comp Biochem Physiol B Biochem Mol Biol. 128(1):109-122.

Harper DB. 1977. Fungal degradation of aromatic nitriles. Enzymology of C–N cleavage by *Fusarium solani*. Biochem J. 167(3):685–692.

Harper DB. 1977. Microbial metabolism of aromatic nitriles. Enzymology of C-N cleavage by *Nocardia sp*. (*Rhodochrous* group) N.C.I.B. 11216. Biochem J. 165(2):309-319.

Henrissat B, Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino- acid sequence similarities. Biochem J. 293:781-788.

Hoskins LC, Boulding ET, Larson G. 1997. Purification and characterization of blood group A-degrading isoforms of alpha-N-acetylgalactosaminidase from *Ruminococcus torques* strain IX-70. J Biol Chem. 272:7932-7939.

Hou Y, Vocadlo D, Withers S, Mahuran D. 2000. Role of beta Arg211 in the active site of human beta-hexosaminidase B. Biochemistry. 39(20):6219-6227.

Husakova L, Herkommerova-Rajnochova E, Semenuk T, Kuzma M, Rauvolfovs J, Ettrich E, Bezouska K, Kren V. 2003. Enzymatic discrimination of GlcNAc and ManNAc-containing disaccharides using β-*N*-acetylhexosaminidase. Advanced Synthesis and Catalysis. 345(6–7):735–742.

Husakova L, Riva S, Casali M, Nicotra S, Kuzma M, Hunkova Z, Kren V. 2001. Enzymatic glycosylation using 6-O-acylated sugar donors and acceptors: ß-N-acetylhexosaminidase catalysed synthesis of 6-O,N,N`-triacetyl-chitobiose and 6`-O,N,N`-triacetylchitobiose. Carbohydr Res. 331 (2):143-148.

Imberty A, Gautier C, Lescar J, Perez S, Wyns L, Loris R. 2000. An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. J Biol Chem. 275:17541-17548.

Jain AN. 2003. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem. 46:499–511.

Kanzaki T, Yokota M, Mizuno N, Matsumoto Y, Hirabayashi Y. 1989. Novel lysosomal glycoaminoacid storage disease with angiokeratoma corporis diffusum. Lancet. 1(8643):875–877.

Kaplan F, Boulay B, Cordeiro P, Hechtman P. 2001. Identification of two amino acid residues which determine the substrate specificity of human β-D-N-acetylhexosaminidase A. Gene Function & Disease, 2:38–45.

Kaplan O, Vejvoda V, Plihal O, Pompach P, Kavan D, Bojarova P, Bezouska K, Mackova M, Cantarella M, Jirků V, Kren V, Martinkova L. 2006. Purification and characterization of a nitrilase from *Aspergillus niger* K10. Appl Microbiol Biotechnol. 73:567-575.

Knapp S, Vocadlo D, Gao Z, Kirk B, Lou J, Withers SG. 1996. NAG-thiazoline, an N-acetylbeta-hexosaminidase inhibitor that implicates acetamido participation. J Am Chem Soc. 118:6804–6805.

Kobayashi M, Goda M, Shimizu S. 1998. Nitrilase catalyzes amide hydrolysis as well as nitrile hydrolysis. Biochemical and Biophysical Research Communications, 253(3):662-666.

Kobayashi M, Komeda H, Yanaka N, Nagasawa T, Yamada H. 1992. Nitrilase from *Rhodococcus rhodochrous* J1. Sequencing and overexpression of the gene and identification of an essential cysteine residue. J Biol Chem. 267(29):20746-20751.

Kobayashi M, Yanaka N, Nagasawa T, Yamada H. 1990. Purification and characterization of a novel nitrilase of *Rhodococcus rhodochrous* K22 that acts on aliphatic nitriles. J Bacteriol. 172(9):4807-4815.

Koshland D. 1953. Stereochemistry and the mechanism of enzymatic reactions. Biol Rev 28:416–436.

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

Krist P, Herkommerova-Rajnochova E, Rauvolfova J, Semenuk T, Vavruskova P, Pavlicek J, Bezouska K, Petrus L, Kren V. 2001. Towards an optimal oligosaccharide ligand for rat natural killer cell activation receptor NKR-P1. Biochem Biophys Res Commun. 287(1):11-20.

Kumaran D, Eswaramoorthy S, Gerchman SE, Kycia H, Studier FW, Swaminathan S. 2003. Crystal structure of a putative CN hydrolase from yeast. Proteins. 52(2):283-291.

Kusiak JW, Quirk JM, Brady RO. 1978. Purification and properties of the two major isozymes of alpha-galactosidase from human placenta. J Biol Chem. 253(1):184-90.

Laederach A, Reilly PJ. 2005. Modeling protein recognition of carbohydrates. Proteins: Structure, Function, and Bioinformatics. 60:591–597.

Lai EC, Withers SG. 1994. Stereochemistry and kinetics of the hydration of 2-acetamido-D-glucal by beta-N-acetylhexosaminidases. Biochemistry. 33(49):14743–14749.

Langley DB, Harty DW, Jacques NA, Hunter N, Guss JM, Collyer CA. 2008. Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen *Streptococcus gordonii* and its complex with the mechanism-based inhibitor NAG-thiazoline. J Mol Biol. 377(1):104-16.

Legler G, Bollhagen R. 1992. (+/-)-6-Acetamido-1,2-anhydro-6-deoxy-myo-inositol: a tight-binding inhibitor and pseudosubstrate for N-acetyl-beta-glucosaminidases. Carbohydr Res. 233:113–123.

Lemieux MJ, Mark BL, Cherney MM, Withers SG, Mahuran DJ, James MN. 2006. Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. J Mol Biol. 359(4):913-29.

Lemieux RU, Delbaere LT, Beierbeck H, Spohr U. 1991. Involvement of water in host-guest interactions. Ciba Found Symp. 158:231-245; discussion 245-8. Review.

Lenny L, Hurst R, Zhu A, Goldstein J, Galbraith R. 1995. Multiple-unit and second transfusions of red cells enzymatically converted from group B to group O: report on the end of phase 1 trials. Transfusion. 35:899–902.

Levitt M, Perutz MF. 1988. Aromatic rings act as hydrogen bond acceptors. J Mol Biol. 201(4):751-754.

Levy GN, Aminoff D. 1980. Purification and properties of alpha-N-acetylgalactosaminidase from *Clostridium perfringens*. J. Biol. Chem. 255:11737-11742.

Lieberman RL, D`aquino JA, Ringe D, Petsko GA. 2009. Effects of pH and iminosugar pharmacological chaperones on lysosomal glycosidase structure and stability. Biochemistry. 48:4816-4827.

Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, Spence J, Nudelman E, Levery SB, White T, Neveu JM, Lane WS, Bourne Y, Olsson ML, Henrissat B, Clausen H. 2007. Bacterial glycosidases for the production of universal red blood cells. Nat Biotechnol. 25(4):454-464.

Loft KJ, Bojarova P, Slamova K, Kren V, Williams SJ. 2009. Synthesis of sulfated glucosaminides for profiling substrate specificities of sulfatases and fungal beta-N-acetylhexosaminidases. Chembiochem. 10(3):565-576.

Lütteke T. 2009. Analysis and validation of carbohydrate three-dimensional structures. Acta Crystallogr D Biol Crystallogr. 65(Pt 2):156-68.

Ly HD, Withers SG. 1999. Mutagenesis of glycosidases. Annu Rev Biochem. 68:487-522. Review.

Maier T, Strater N, Schuette CG, Klingenstein R, Sandhoff K, Saenger W. 2003. The X-ray crystal structure of human beta-hexosaminidase B provides new insights into Sandhoff disease. J Mol Biol. 328(3):669-681.

Manuel SGA, Ragunath C, Sait HBR, Izano EA, Kaplan JB, Ramasubbu N. 2007. Role of active-site residues of dispersin B, a biofilm-releasing β-hexosaminidase from a periodontal pathogen, in substrate hydrolysis. FEBS Journal. 274:5987–5999.

Manzanares P, de Graaff LH, Visser J. 1998. Characterization of galactosidases from *Aspergillus niger*: purification of a novel -galactosidase activity. Enzyme Microb. Technol. 22 (5):383-390.

Margolles-Clark E, Tenkanen M, Luonteri E, Penttilä M. 1996. Three -galactosidase genes of *Trichoderma reesei* cloned by expression in yeast. Eur. J. Biochem. 240**:**104-111.

Mark BL, Mahuran DJ, Cherney MM, Zhao D, Knapp S, James MN. 2003. Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. J Mol Biol. 327(5):1093-1109.

Mark BL, Vocadlo DJ, Triggs-Raine BL, Withers SG, James MNG. 2001. Insights into the mechanisms of retaining hexosaminidases: crystal structure of *Streptomyces plicatus* b-hexosaminidase in complex with the cyclic intermediate analogue NAG-thiazoline at 2.1 Å. J Biol Chem. 276:10330-10337.

Matsuo I, Kim S, Yamamoto Y, Ajisaka K, Maruyama JI, Nakajima H, Kitamoto K. 2003. Cloning and overexpression of beta-N-acetylglucosaminidase encoding gene *nagA* from *Aspergillus oryzae* and enzyme-catalyzed synthesis of human milk oligosaccharide. Biosci Biotechnol Biochem. 67(3):646-50.

McCarter JD, Withers SG. 1994. Mechanisms of enzymatic glycoside hydrolysis. Curr Opin Struct Biol 4:885–892.

Mega T, Ikenaka T, Matsushima Y. 1970. Studies on *N*-Acetyl-*β*-D-glucosaminidase of *Aspergillus oryzae*. J Biochem. 68:109-117.

Miljkovic M, editor. 2009. Carbohydrates: Synthesis, Mechanisms, and Stereoelectronic Effects. New York:Springer.

Miller ES Jr, Kimberley, Parker N, Liebl W, Lam D, Callen W, Snead MA, Mathur EJ, Short JM, Kelly RM. 2001. Alpha-D-galactosidases from *Thermotoga species*. Methods Enzymol. 330:246-260.

Mohamad SB, Nagasawa H, Uto Y, Hori H. 2002. Tumor cell alpha-N-acetylgalactosaminidase activity and its involvement in GcMAF-related macrophage activation. Comparative Biochemistry and Physiology Part A. 132:1–8.

Nakai T, Hasegawa T, Yamashita E, Yamamoto M, Kumasaka T, Ueki T, Nanba H, Ikenaka Y, Takahashi S, Sato M, Tsukihara T. 2000. Crystal structure of N-carbamyld-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. Structure 8:729–737.

Nieder V, Kutzer M, Kren V, Gallego RG, Kamerling JP, Elling L. 2004. Screening and characterization of beta-N-acetylhexosaminidases for the synthesis of nucleotide-activated disaccharides. Enzyme and Microbial Technology. 34(5):407-414.

Ogata M, Zeng X, Usui T, Uzawa H. 2007. Substrate specificity of N-acetylhexosaminidase from *Aspergillus oryzae* to artificial glycosyl acceptors having various substituents at the reducing ends. Carbohydrate research. 342:23-30.

O'Reilly C, Turner PD. 2003. The nitrilase family of CN hydrolysing enzymes – a comparative study. J Appl Microbiol. 95(6):1161-1174. Review.

Pace HC, Brenner C. 2001. The nitrilase superfamily: classification, structure and function. Genome Biol. 2:REVIEWS0001.

Pace HC, Hodawadekar SC, Draganescu A, Huang J, Bieganowski P, Pekarsky Y, Croce CM, Brenner C. 2000. Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers. Curr Biol. 10:907–917.

Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JA, van den Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutinho PM, Danchin EG, Debets AJ, Dekker P, van Dijck PW, van Dijk A, Dijkhuizen L, Driessen AJ, d'Enfert C, Geysens S, Goosen C, Groot GS, de Groot PW, Guillemette T, Henrissat B, Herweijer M, van den Hombergh JP, van den Hondel CA, van der Heijden RT, van der Kaaij RM, Klis FM, Kools HJ, Kubicek CP, van Kuyk PA, Lauber J, Lu X, van der Maarel MJ, Meulenberg R, Menke H, Mortimer MA, Nielsen J, Oliver SG, Olsthoorn M, Pal K, van Peij NN, Ram AF, Rinas U, Roubos JA, Sagt CM, Schmoll M, Sun J, Ussery D, Varga J, Vervecken W, van de

Vondervoort PJ, Wedler H, Wösten HA, Zeng AP, van Ooyen AJ, Visser J, Stam H. 2007. Genome sequencing and analysis of the versatile cell factory Aspe*rgillus niger* CBS 513.88. Nat Biotechnol. 25(2):221-231.

Plihal O, Sklenar J, Kmonickova J, Man P, Pompach P, Havlicek V, Kren V, Bezouska K. 2004. *N*-glycosylated catalytic unit meets *O*-glycosylated propeptide: complex protein architecture in a fungal hexosaminidase. Biochem Soc Trans 32:764-765.

Polaina J, MacCabe AP, editors. 2007. Industrial enzymes: structure, function and applications. Dordrecht: Springer.

Post DA, Luebke VE. 2005. Purification, cloning, and properties of alpha-galactosidase from *Saccharopolyspora erythraea* and its use as a reporter system. Appl Microbiol Biotechnol. 67:91-96.

Prag G, Papanikolau Y, Tavlas G, Vorgias CE, Petratos K, Oppenheim AB. 2000. Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. J Mol Biol. 300(3):611-7.

Quiocho FA. 1993. Probing the atomic interactions between proteins and carbohydrates. Biochem Soc Trans. 21(2):442-8. Review.

Rajnochova E, Dvorakova J, Hunkova Z, Kren V. 1997. Reverse hydrolysis catalysed by β- N-acetylhexosaminidase from *Aspergillus oryzae*. Biotechnology Letters. 19(9):869-872.

Ramasubbu N, Thomas LM, Ragunath C, Kaplan JB. 2005. Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen *Actinobacillus actinomycetemcomitans*. J Mol Biol. 349(3):475-86.

Ríos S, Pedregosa AM, Fernandez Monistrol I, Laborda F. 1993. Purification and molecular properties of an alpha-galactosidase synthesized and secreted by *Aspergillus nidulans*. FEMS Microbiol Lett. 112(1):35-41.

Sakai N, Tajika Y, Yao M, Watanabe N, Tanaka I. 2004. Crystal structure of hypothetical protein PH0642 from *Pyrococcus horikoshii* at 1.6 A resolution. Proteins: Struct Funct Bioinform. 57:869–873.

Sakuraba H, Matsuzawa F, Aikawa S, Doi H, Kotani M, Lin H, Ohno K, Tanaka A, Yamada H, Uyama E. 2002. Molecular and structural studies of the GM2 gangliosidosis 0 variant. J Hum Genet. 47(4):176-183.

Sandhoff K, Andreae U, Jatzkewitz H . 1968. Deficient hexozaminidase activity in an exceptional case of Tay-Sachs disease with additional storage of kidney globoside in visceral organs. Life Sci. 7(6):283–288.

Schindler D, Bishop DF, Wolfe DE, Wang AM, Egge H, Lemieux RU, Desnick RJ. 1989. A neuroaxonal dystrophy due to lysosomal alpha-N-acetylgalactosaminidase deficiency. N. Engl. J. Med. 320:1735-1740.

Sharma R, Bukovac S, Callahan J, Mahuran D. 2003. A single site in human beta-hexosaminidase A binds both 6-sulfate-groups on hexosamines and the sialic acid moiety of GM2 ganglioside. Biochim Biophys Acta. 1637(1):113-118.

Sigurskjold BW, Bundle DR. 1992, Thermodynamics of oligosaccharide binding to a monoclonal antibody specific for a salmonella O-antigen point to hydrophobic interactions in the binding site. Journal of Biological Chemistry. 267:8371-8376.

Singh R, Sharma R, Tewari N, Geetanjali, Rawat DS. 2006. Nitrilase and its application as a 'green' catalyst. Chem Biodivers. 3(12):1279-1287. Review.

Singh S, Packwood J, Samuel CJ, Critchley P, Crout DH. 1995. Glycosidase-catalysed oligosaccharide synthesis: preparation of N-acetylchitooligosaccharides using the beta-N-acetylhexosaminidase of *Aspergillus oryzae*. Carbohydr Res. 279:293-305.

Sinnott ML. 1990. Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev. 90:1171–1202.

Skov LK, Mirza O, Henriksen A, De Montalk GP, Remaud-Simeon M, Sarçabal P, Willemot RM, Monsan P, Gajhede M. 2001. Amylosucrase, a glucan-synthesizing enzyme from the alpha-amylase family. J Biol Chem. 276(27):25273-25278.

Snajdrova R, Mylerova-Kristova V, Crestia D, Nicolaou K, Kuzma M, Lemaire M, Galienne E, Bolte J, Kren V, Martinkova L. 2004. Nitrile biotransformation by Aspergillus niger. J. Mol. Catal. B: Enzymatic. 29 (1-6):227–232.

Spangenberg P, Andre C, Dion M, Rabiller C, Mattes R. 2000. Comparative study of new alpha-galactosidases in transglycosylation reactions. Carbohydr Res. 329(1):65-73.

Sripuan T, Aoki K, Yamamoto K, Tongkao D, Kumagai H. 2003. Purification and characterization of thermostable alpha-galactosidase from *Ganoderma lucidum*. Biosci Biotechnol Biochem. 67(7):1485-1491.

Stick RV, Williams S, editors. 2008. Carbohydrates: the essential molecules of life. 2nd edition. Oxford:Elsevier.

Stoddart JF. 1971. Stereochemistry of Carbohydrates. New York:Wiley-Interscience.

Sulzenbacher G, Driguez H, Henrissat B, Schülein M, Davies GJ. 1996. Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group. Biochemistry. 35:15280–15287.

Sumida T, Ishii R, Yanagisawa T, Yokoyama S, Ito M. 2009. Molecular cloning and crystal structural analysis of a novel b-N-acetylhexosaminidase from *Paenibacillus sp*. TS12 capable of degrading glycosphingoliids. J Mol Biol. 392(1):87-99.

Sung SSJ, Sweely CC. 1980. Purification and partial characterization of porcine liver alpha-N-acetylgalactosaminidase. J Biol Chem. 255:6589—6594.

Suzuki R, Katayama T, Kitaoka M, Kumagai H, Wakagi T, Shoun H, Ashida H, Yamamoto K, Fushinobu S. 2009. Crystallographic and mutational analyses of substrate recognition of endo-alpha-N-acetylgalactosaminidase from Bifidobacterium longum. J Biochem. 146(3):389-398.

Tay W. 1881. Symmetrical changes in the region of the yellow spot in each eye of an infant. Trans.Ophthalmol. Soc. UK, 1:155–157.

Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. 1996. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. Nature Struct Biol. 3:638-648.

Tews I, Terwisscha van Scheltinga A, Perrakis A, Wilson KS, Dijkstra B. 1997. Substrate-assisted catalysis unifies two families of chitinolytic enzymes. J Amm Chem Soc. 119:7954-7959.

Thuku RN, Weber BW, Varsani A, Sewell BT. 2007, Post-translational cleavage of recombinantly expressed nitrilase from *Rhodococcus rhodochrous* J1 yields a stable, active helical form. FEBS Journal, 274:2099–2108.

Tramice A, Pagnotta E, Romano I, Gambacorta A, Trincone A. 2007. Transglycosylation reactions using glycosyl hydrolases from *Thermotoga neapolitana*, a marine hydrogen-producing bacterium. Journal of Molecular Catalysis B: Enzymatic. 47:21–27.

Uzawa H, Zeng X, Minoura N. 2003. Synthesis of 6'-sulfodisaccharides by beta-N-acetylhexosaminidase-catalyzed transglycosylation. Chem Commun (Camb). 1:100-101.

Varki A, Cummings R, Esko J, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, editors. 1999. Essentials of Glycobiology. New York: Cold Spring Harbor Laboratory Press.

Vejvoda V, Kaplan O, Klozova J, Masak J, Cejkova A, Jirků V, Stloukal R, Martinkova L. 2006. Mild hydrolysis of nitriles by *Fusarium solani* strain O1. Folia Microbiol (Praha). 51(4):251-256.

Wang LX, Huang W. 2009. Enzymatic transglycosylation for glycoconjugate synthesis. Curr Opin Chem Biol. 13(5-6):92-600. Review.

Webb, Edwin C. 1992. Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes. San Diego: Published for the International Union of Biochemistry and Molecular Biology by Academic Press. Available from: http://www.chem.qmul.ac.uk/iubmb/enzyme/.

Weignerova L, Filipi T, Manglova D, Kren V. 2008. Induction, purification and characterization of alpha-N-acetylgalactosaminidase from *Aspergillus Niger*. Appl Microbiol Biotechnol. 79(5):769-774.

Williams SJ, Mark BL, Vocadlo DJ, James MNG and Withers SG. 2002. The catalytic role of aspartate 313 in the *Streptomyces plicatus* hexosaminidase: Implications from kinetic and structural analysis. J Biol Chem. 277(42):40055-40065.

Winkler M, Kaplan O, Vejvoda V, Klempier N, Martinkova L. 2009. Biocatalytic application of nitrilases from *Fusarium solani* O1 and *Aspergillus niger* K10. Journal of Molecular Catalysis B: Enzymatic. 59(4):243-247.

Xiao M, Tanaka K, Quian XM, Yamamoto K, Kumagai H. 2000. High-yield production and characterization of alpha-galactosidase from *Bifidobacterium breve* grown on raffinose. Biotechnol. Lett. 22:747-751.

Yamamoto K, Lee KM, Kumagai H, Tochikura T. 1985. Purification and characterization of beta-N-acetylhexosaminidase from *Penicillium oxalicum*. Agric Biol Chem. 49:611-619.

Yamamoto N. 2006. Pathogenic significance of a-N-acetylgalactosaminidase found in the envelope glycoprotein gp160 of human immunodeficiency virus type 1. AIDS Res Human Retroviruses. 22:262–271.

Yamamoto N, Naraparaju VR, Srinivasula SM. 1995. Structural modification of serum vitamin D3-binding protein and immunosuppression in HIV-infected patients. AIDS Res Human Retrovirus. 11:1373–1378.

Yamamoto N, Naraparaju VR, Urade M. 1997. Prognostic utility of serum alpha-N-acetylgalactosaminidase and immunosuppression resulted from deglycosylation of serum Gc protein in oral cancer patients. Cancer Res. 57:295–299.

Yamamoto N, Ushijima N, Koga Y. 2009. Immunotherapy of HIV-infected patients with Gc protein-derived macrophage activating factor (GcMAF) Journal of Medical Virology. 81:16–26.

Yamashita A, Hashimoto H, Fujita K, Okada M, Mori S, Kitahata S. 2005. Reverse reaction of *Aspergillus niger* APC-9319 alpha-galactosidase in a supersaturated

substrate solution: production of alpha-linked galactooligosaccharide (alpha-GOS). Biosci Biotechnol Biochem. 69(7):1381-1388.

Yeom SJ, Kim HJ, Lee JK, Kim DE, Oh DK. 2008. An amino acid at position 142 in nitrilase from *Rhodococcus rhodochrous* ATCC 33278 determines the substrate specificity for aliphatic and aromatic nitriles. Biochem J. 415(3):401-7.

York WS, Hawkins R. 2000. Preparation of oligomeric β-glycosides from cellulose and hemicellulosic polysaccharides via the glycosyl transferase activity of a *Trichoderma reesei* cellulose. Glycobiology. 10(2):193-201.

Zeilinger S, Kristufek D, Arisan-Atac I, Hodits R, Kubicek CP. 1993. .Conditions of formation, purification, and characterization of an alpha-galactosidase of *Trichoderma reesei* RUT C-30. Appl Environ Microbiol. 59(5):1347-1353.

Zeng X, Sun Y, Ye H, Liu J, Uzawa H. 2007. Synthesis of *p* -nitrophenyl sulfated disaccharides with β-D-(6-sulfo)-GlcNAc units using β- *N* -acetylhexosaminidase from *Aspergillus oryzae* in a transglycosylation reaction. Biotechnol Lett. 29(7):1105-1110.

PART II METHODS

6.1 PREPARATION OF PROTEIN: HOMOLOGY MODELING, EQUILIBRATION, N-GLYCOSYLATION

6.1.1 Homology modeling

Homology modeling is a class of methods for constructing an atomic-resolution model of a protein from its amino acid sequence. Generally, there are two main groups of protein modeling methods that take the primary structure as the starting point: *ab initio* or *de novo* modeling (Lee *et al.* 2009) and comparative modeling (Zhang 2008).

Ab initio or *de novo* methods produces a structure based on physical principles underlying protein folding; however, currently it is limited by the size of protein and even for small proteins the results are not necessarily convincing, as protein folding is still not understood completely. Homology modeling is the most successful method at the moment for the prediction of three-dimensional protein structures (Zhang 2008). Despite numerous on-line services, so called “black box” approaches, a “handish” step by step modeling reflecting each modeling cycle , improving the constraints, alignments and parameters, is still the best way, as it enables not only to reflect the performance of the algorithms but lets one include constrains coming from additional experimental information as the biological role of certain residues or data from different low resolution techniques that constrain the behavior of certain amino acids (as for example gauche-trans-gauche conformation of disulfide bridge from Raman spectroscopy).

The final quality of model is influenced by the appropriate selection of templates and correct alignment. A sufficient number of templates is critical, especially in regions that are in the so-called “twilight zone”, in which simple pairwise alignment are not reliable; on the other hand, an unreasonably high number of templates increases the computational time and generates noise.

The approach we used for generating homology models is depicted in figure 6.1. It is inspired by the iterative genetic algorithm (John and Sali 2003). The advantage of this approach can be seen in the iterative generation and validation of alternative models.

Homology modeling includes several steps: template search and selection, alignment of the target sequence with the templates, building several homology models, validation of models and refinement.

6.1.2 Template selection

The widely employed basic local alignment search algorithm BLAST (Altschul *et al.* 1990) was used for identifying possible templates, which means sequences with high primary sequence identity, so called homologs. BLAST is faster than the other popular heuristic algorithm FASTA (Lipman and Pearson 1985). For performing searches of remote homology, a method based on the hidden Markov models is the optimal choice (Söding *et al.* 2005). One shortcoming of BLAST that one has to have in mind when performing a search, consists in the local character of the alignment produced. Especially in case of low sequence similarity, one has to check if the algorithm did not "eliminate" the biological role of the query sequence. High scoring templates are then extracted from the PDB database (http://www.pdb.org), a database for X-ray and NMR protein structures.

6.1.3 Alignment

For pairwise and multiple alignments of amino acid sequences, a number of approaches, methods and programs can be employed. We used the T-Coffee server (Poirot *et al.* 2003), that allows the use of a combination of different algorithms (ClustalW, ProbCons, MUSCLE, DALIGN, PCMA, POA, MAFFT), is able to perform global and/or local alignments, and operates with various data types (Notredame *et al.* 2000).

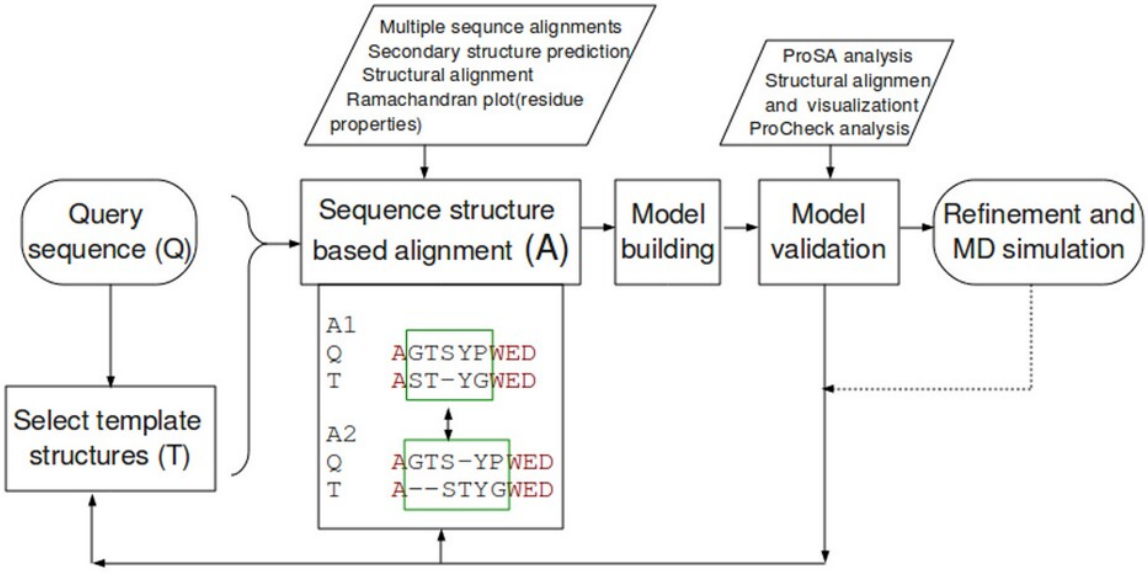

Figure 6.1. General scheme of used homology modeling procedure

T-Coffee uses the progressive algorithm (Feng and Doolittle 1987; Hogeweg and Hesper 1984) for aligning sequences. The main difference from the widely used ClustalW algorithm, as the standard alignment tool, and similar programs is in the way T-Coffee scores and evaluates alignments.

Advantages of T-Coffee include the possibility to combine information obtained from local and global alignments, the optimization method, position dependent score, combination of sequence alignment with structural information (Sullivan *et al.* 2004). The major disadvantages of the heuristic method used in T-Coffee (in contrast to the classical Needleman Wunsch algorithm) are the facts that it cannot guarantee to find the optimal solution, it is *N* times slower (CPU time) than Clustal W, however, its strength is the improved accuracy (Notredame and Abergel 2003).

T-Coffee scoring function

The alignment strategy of T-Coffee is depicted in figure 6.2 (Notredame *et al.* 2000). Square blocks mean procedure and rounded blocks – data structures.

Library of the list of weighted pairwise residue matches extracted from pairwise alignments, made of provided sequences. The selection of alignments is arbitrary and can contain some alternative alignments of the same pair of sequences.

The combination of global (by ClustalW method) and local (usually Lalign program) methods, as depicted in figure 6.2, gives good results (Notredame *et al.* 2000).

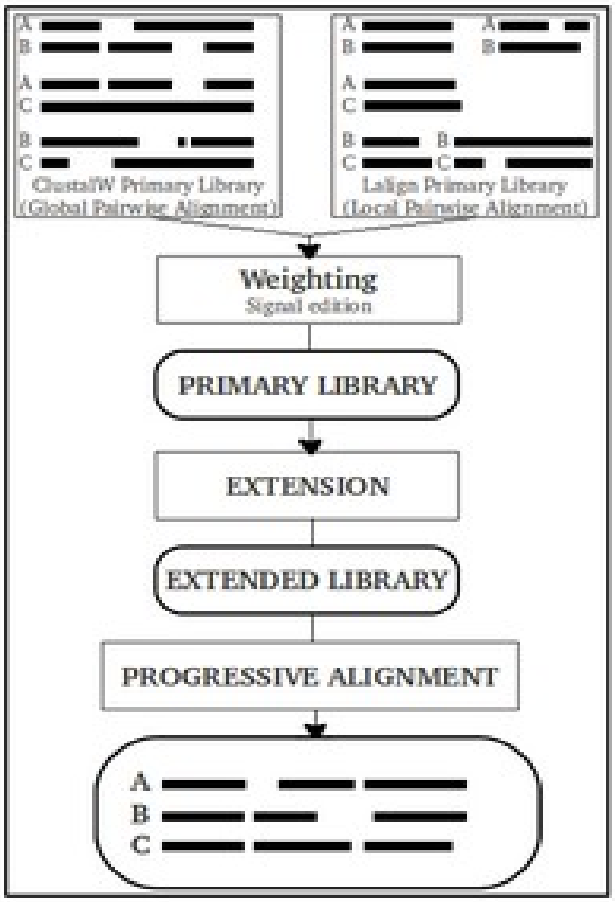

Figure 6.2. Algorithm of T-Coffee alignment (picture is adjusted from Notredame *et al.* 2000).

Obtained residue pairs are not equal because they are parts of different alignments, so their information input to the final alignment is also different. A weighting scheme is used to range residue substitutions (Sander and Schneider 1991), the multiple alignment hereby is stronger influenced by closest sequences.

Weights extended by heuristic algorithms are used instead of substitution matrices for multiple alignment construction, which reflects the degree to which residues from primary library align consistently with residues from all other sequences.

The progressive alignment step exploits dynamic programming, the so-called Needleman-Wunsch algorithm, to produce the final alignment (Needleman and Wunsch 1970).

To align two prealigned sequence pairs average library scores in each column are used.

Alignment strategy

We used a combination of different algorithms to generate alignments with T-Coffee. Conserved blocks of aligned sequences that are common for all alignments are selected. The one alignment with the best score is chosen for production of the first set of models. Structural alignments of the selected templates were additionally generated with Sheba (plug-in version in Yasara, Jung and Lee 2000) and visualized with program Yasara. These structural alignments and the consensus secondary structure prediction (http://npsa-pbil.idcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html) were further used for validation and correction of the T-Coffee output according to secondary structure elements and conserved residues in known structures.

6.1.4 Model building

Homology modeling is applied to generate a three-dimensional model based on the multiple alignment resulting from the previous step. We have chosen the restraint-based approach implemented in the modeling program Modeller (Sali and Blundell 1993), that generates positions for all non-hydrogen atoms in the target structure by satisfaction of spatial restraints. The method takes its inspiration from the way calculations are performed to construct a three-dimensional structure from data generated by NMR spectroscopy. The work-flow of the program includes distance geometry, local energy minimization, and molecular dynamics. The multiple alignment is used to construct a set of geometrical criteria like dihedral angles and atom distances, that are then converted to probability density functions for each restraint. The distance geometry algorithm then applies these restraints to the main internal coordinates to construct rough models for the sequence. These are refined using energy minimization and simulated annealing, by combining the constraints obtained from the template structures with a simple molecular mechanics energy function.

6.1.5 Model validation and iteration of alignment cycle

Estimation of the general quality of the generated models is essential and not only validates the theoretical probability of the structure/conformation, but also allows to rank

the structures and to select the best. We used different approaches based on the statistical data (Figure 6.1), extracted from the structures in PDB: geometric parameters (Laskowski *et al.* 1993) and calculated energy values (Sippl 1993). ProCheck analyses geometrical parameters of three-dimensional structures: distances between atoms, angles, torsions. Agreement of dihedral angles with allowed values in the modelled residues is represented by a Ramachandran plot. Amino acids in generously allowed and disallowed regions are candidates for realignment in the next step.

ProSA employs a method based on knowledge based mean fields, which basically is the function of the position of a residue within the chain and its conformation. On the contrary, this method does not allow any analysis of possible close contacts or violations of other basic steric principles.

The validation is either followed by acceptance of the best model or a new modeling cycle, including either additional restraints coming from experimental data or realignment of badly modeled regions. After some iterations of the modeling cycle we thus receive a model, which can be accepted for further loop modeling and refinement.

6.1.6 Loop modeling

Misalignments can appear in the loop parts and the start/end positions of the secondary structure elements; the N- and C-terminal parts of proteins are also challenging for modeling (Fiser *et al.* 2000, Rost 1999). modeling of loops has been done by MODLOOP (Fiser *et al.* 2000).

6.1.7 Refinement

Equilibration of the models in the solvent by molecular dynamics is a way to guide investigation of potential modeling problems, and it either leads to a clear improvement or demonstrates model robustness, and gives a refined model as a result, or it points to critical, unstable regions in the structure drifting from their initial positions, and thus gives information for the generation of better alignments.

A good refinement (equilibration) is improves the hydrogen bonds network in the solvated protein, corrects rarely populated local conformations or relaxes high potential energy regions, as rotameric conformations and internal packing of residue side chains could be incorrectly modeled by homology modeling.

Refinement of models was done by MD. The three-dimensional structure was placed in a box applying periodic boundary conditions; the cell was filled with TIP3P water, and neutralized by placing counter ions. To remove steric overlaps and correct the covalent

geometry, the energy of the complex was minimized with the Yasara2 force field and default parameters, followed by a short simulated annealing protocol (atom velocities scaled down by 0.9 every 10th step) until a convergence was reached (Krieger *et al.* 2004). It followed by short MD with Yamber2 force field in explicit water solution.

Results of refinement by MD were analyzed: RMSD (root mean square deviation), secondary structure changes and solvent accessible surface.

6.1.8 Glycosylation

Glycosylation was performed on-line at http://www.glycosciences.de/ with GlyProt using carbohydrates from SWEET, a program useful for constructing 3D models of saccharides from their sequences using standard nomenclature **(**Bohne-Lang and von der Lieth 2005). Selected glycan antennas were checked for clashes and in some cases cut.

6.2 PREPARATION OF SUBSTRATES: QUANTUM MECHANICS, OPTIMIZATION

Initial geometries of molecules were either obtained from PDB and extracted from already solved structure and modified or built from scratch. Correct geometry of molecule is achieved by optimization. For geometry optimization of molecules used for MD simulation it is often enough to use semi-empirical methods, as AM1 (Austin model 1) (Dewar *et al.* 1985), implemented in Yasara.

Yasara can assign force field parameters by the AutoSmile approach (Jakalian *et al.* 2002). In a first step semi-empirical AM1 Mulliken point charges are calculates that are corrected by assignment of AM1BCC atom types and improved AM1BCC charges by fragments of molecules with known RESP charges, to closer resemble RESP charges. Corresponding bond, angle and torsion potential parameters are taken from the General AMBER force field.

However, charged molecule or radicals should be treated by *ab initio* methods. Parameters for molecule in this case should be added to the force field.

6.2.1 Partial charge derivation and force field parameterization

Partial charges and force field parameters were calculated in Gaussian for the Yamber2 force field. Molecule geometries were preoptimized with basis set 6-31G* by restricted Hartree-Fock method, partial charges were calculated by population analysis. We calculated ESP (electrostatic potential) charges by Gaussian and then transformed

them in RESP (Restrained Electrostatic Potential) with Antechamber (Wang *et al.* 2004, Wang *et al.* 2006).

The basic idea of electrostatic potential fit charges is that a least squares fitting algorithm is used to derive a set of atom-centered point charges which best reproduce the electrostatic potential of the molecule (http://ambermd.org/doc6/html/AMBER-sh-19.4.html).

The 6-31G* basis set tends to result in dipole moments which are 10-20% larger than the gas phase, which is better for MD in water. DFT (Density function theory) charges are smaller more appropriate in gas phase (St-Amant *et al.* 1995). ESP fit charges reproduce interaction energies well. RESP charges are modification of the original ESP method (Bayly *et al.* 1993).

One problem with electrostatic potential fit charges in general is that they reproduce the molecular potential and the dipole moment very well for the conformation of the molecule employed in the fit. However, when those charges are applied to other conformations, the agreement is not as good.

6.3 DOCKING

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. The method one could choose for docking substrates depends on the information available about substrate-enzyme interactions.

The final results of docking very often depend on initial position, cavity properties (Kontoyianni *et al.* 2004), quality of structure (Nissink *et al.* 2002). Docking programs show different accuracy for different classes of proteins (Cole *et al.* 2005). This implies some limitations for the use of homology models.

Docking programs are often successful in reproduction of experimentally observed docking position. Scoring functions are able to distinguish between good substrates and “not accepted”, however the best scoring model is not always corresponding to the correct (like in crystal) pose (Warren *et al.* 2006). The K_i values obtained by AutoDock 3 in 50 % of tested inhibitors are in good agreement with experiment, in other cases are underestimated (Toprakci and Yelekci 2005).

AutoDock is one of the most often reported and used methods (Sousa *et al.* 2006). Recently Glide is reported as the best program (among Glide, GOLD, AutoDock and FlexX) for carbohydrate-antibodies docking, AutoDock produce less accurate results (Agostino *et al.* 2009).

Above methods are in good accordance with the ‘key-lock’ theory. However, concerning the induced fit model by Koshland, 1958, or the today generally accepted new paradigm of conformational selection, in which the protein is described as a dynamic conformational ensemble and the ligand recognizes one specific conformation out of the full ensemble in a lock-key manner — MD of the protein-ligand complex and the protein itself can give us much more information about possible conformational changes at the active site. Thus docking combined with MD is powerful tool for ligand geometry and binding prediction (Sousa *et al.* 2006).

We used different strategies for docking substrates: derivation of substrate coordinates from previously determined homologous protein-ligand complexes; blind docking by docking program (AutoDock). The first approach is applied when the position of the substrate is unambiguously determined with high resolution in a conserved binding site, like in HEXs. Initial placement hereby is done by overlaying the model with the crystal structure co-crystallized with substrate in Yasara and further modification of it.

6.4 MD SIMULATION

“Molecular dynamics (MD) is the computational simulation of physical movements of atoms and molecules” (www.en.wikipedia.com), iteratively solving the Newtonian equations of motion for a collection of particles

The advantage of this simulation technique is that it can model the dynamic behavior of a system in its equilibrium state on an atomistic scale, which is mostly impossible to do on an all-atom level experimentally. MD is widely used to study of atomic and local fluctuations, conformational changes, enzyme-substrate interaction (Schlick 2002), protein folding and stability (http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html). MD appeared to be an appropriate method for correct modeling of carbohydrate flexibility in carbohydrate-protein complexes (Woods 1998). MD uses some approximations in its calculations, which are depending on the goal of simulation.

Movement of atoms and molecules in MD simulation follow the Newton’s law of motion (6.4a). In terms of energy it could be represented through potential energy (6.4b) (http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html):

$$F_i = m_i \times a_i \qquad (6.4a),$$

where $\mathbf{F_i}$ is a sum of forces acting on a particle, $\mathbf{m_i}$ is a mass of particle and $\mathbf{a_i}$ is an acceleration.

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (6.4b)$$

V in the equation (6.4b) is the potential energy of the particle with mass $\mathbf{m_i}$ and $\mathbf{r_i}$ - is a position of a particle in time t. Definition of potential energy (or more correct potential energy function) is a basic notion for different conceptions of MD.

6.4.1 Potential energy function

The potential energy function is an equation, including forces acting on particles in the particular condition. Potential energy is a function of the position of particle in 3D space. There are some approaches how to determine the potential energy field: *ab initio*, semi-empirical and empirical methods.

"The basic idea underlying *ab initio* MD is to compute the forces acting on the nuclei from electronic structure calculations that are performed “on–the–fly” as the molecular dynamics trajectory is generated” – described by Marx and Hunter. The time dependent Schrodinger equation is used. This group of method is useful for dynamics of small systems and short simulation times, but it is very sensitive to defined initial conditions (Marx and Hutter 2000).

Semi-empirical methods use several parameters that were derived from empirical data. They have limited applications.

Empirical energy functions (classical MD) use empirically determined parameters - by NMR, from crystal structure or by spectroscopy. Some of them could be calculated by quantum mechanics (http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html). Yasara approximates atoms as balls connected via springs.

Potential energy in classical MD is calculated as the sum of bonded and non-bonded terms. Interaction between the particles and behavior of particles is modeled using descriptions based on the underlying physical laws. Different force fields parameters are developed by different groups, the most widely used for proteins are AMBER (Cornell et al. 1995), CHARMM (MacKerell *et al.* 1998), GROMACS (Van Der Spoel *et al.* 2005) and OPLS (Jorgensen and Tirado-Rives 1988). There is no consensus on the best force field for carbohydrate (Forster 2002).

MD in Yasara can be performed with different force fields: AMBER, NOVA, YAMBER and YASARA. Generally all of them are derived from AMBER and use equation 6.4c for potential energy calculation (Cornell *et al.* 1995):

$$V(r) = V_{intramolecular} + V_{intermolecular} = \sum_{nonb,ij} \left(A_{ij}/r_{ij}^{12}\right) - \left(B_{ij}/r_{ij}^{6}\right) + \left(q_i q_j / r_{ij}\right) + \\ + \sum_{bonds} k_b \left(b - b_0\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{dihedrals} \left(V_n/2\right)\left(1 + \cos\left[n\varphi - \delta\right]\right) \quad (6.4c)$$

Intermolecular forces ($\mathbf{V_{intermolecular}}$) act between bonded atoms within 4 atoms connected by 3 bonds. Intramolecular forces ($\mathbf{V_{intramolecular}}$) include Van der Waals interaction and electrostatic Coulomb interaction. In the equation (6.4c) $\mathbf{A_{ij}}$ and $\mathbf{B_{ij}}$ are parameters proportional to the distance between particles i and j at which the potential energy is minimum, $\mathbf{r_{ij}}$ - is a distance between particles i and j; $\mathbf{q_i}$ and $\mathbf{q_j}$ - are charges, $\mathbf{k_b}$ and $\mathbf{k_\theta}$ are constant parameters (~ 'spring constant'); $\mathbf{b - b_0}$ and $\mathbf{\theta - \theta_0}$ are displacement of bond length and angle from equilibrium geometry; last term is the energy for twisting a bond or lone pairs of electrons.

Yasara uses no polarization in force field calculation.

6.4.2 Propagation of the motion in time

Equation (6.4b) describes the position of atom in each selected time-interval (so called time-step in bioinformatics). Several integration algorithms for equation 6.4b exist. The most common are Verlet algorithms (Schlink 2002), Gear integration (Nezbeda 2007), Beeman's algorithm (http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html).

The Verlet algorithm has some alternatives - position, leap-frog and velocity algorithm. They differ by the way velocity is treated. While position Verlet doesn't include it explicitly, other two calculate it in different time intervals (Schlink 2002). Despite of its simplicity Verlet algorithm showed a good conservation of energy (Schlink 2002).

The mathematical formulation of the leap frog algorithm is given in ((6.4d) and (6.4e)). The name leap frog is derived from the fact that velocity is calculated each half step and the coordinate is determined each step. This scheme is widely used in multistep simulations (Schlink 2002).

$$r(t+\delta t) = r(t) + v\left(t + \frac{1}{2}\delta t\right)\delta t \quad (6.4d)$$

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + a(t)\delta t \quad (6.4e),$$

where $\mathbf{r(t+j\delta t)}$ is a new position of particle (in the time $\mathbf{t+j\delta t}$), changed from **r(t)** by movement with velocity **v**; and **a** is an acceleration in a corresponding time. The steps of trajectory propagation used by Yasara (Krieger 2002) is presented in figure 6.3. In the

beginning of simulation Yasara assigns random velocities using a Maxwell-Boltzmann distribution.

Step 1. Calculates of potential energy of each atom using initial positions of atoms. Acceleration is calculated from the potential energy.

Step 2. New velocities are calculated.

Step 3. New coordinates are calculated and we go back to the step 1.

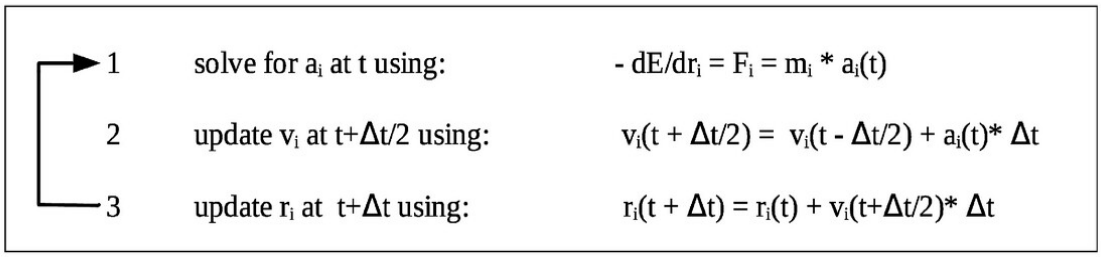

Figure 6.3. Algorithm of trajectory propagation (figure from http://swift.cmbi.kun.nl/teach/B1SEM/B1SEM_10.html).

6.4.3 Initial state and simulation parameters

There are several initial parameters that need to be specified for the system used in MD. Coordinates are mostly obtained from experimentally determined three-dimensional structures in PDB file-format or from energetically minimized, optimized structures built from scratch using a molecular builder.

To specify the size of the system, we need to limit it by defining a simulation cell. Cell boundary conditions can be periodic (transparent wall surrounded by copies of itself) or the simulation cell is really surrounded by a wall.

The macrostate of the system in MD is described by N (number of particles), P (pressure), T (temperature), V (volume). The conditions under which MD is specified in our case conserve the total number of particles N, total pressure P and temperature T – the isothermal-isobaric (NPT) ensemble (number of particles, temperature and pressure are constant). In addition to a thermostat, a barostat is needed. The NPT ensemble resembles most closely laboratory conditions with a flask open to ambient temperature and pressure. The pressure is kept constant by rescaling the simulation cell (Yasara documentation). The temperature is adjusted using a Berendsen thermostat (Berendsen *et al.* 1984) based on the time-averaged temperature and at regular intervals is rescaled to keep temperature constant.

6.4.4 Timestep and energy calculation

The correct timestep is an important issue (Frenkel and Smit 2002). When chosen comparatively too large, it easily can lead to inaccurate simulation followed by a simulation crash, if chosen too small, the MD gets very slow and random errors can play a significant role for the results of MD.

Yasara use multistep algorithm which means that it calculates intermolecular and intramolecular forces in different time-intervals (timesteps) (Krieger *et al.* 2004, Krieger *et al.* 2009). Intramolecular forces have high vibration frequencies and should be treated in shorter time-intervals than intermolecular (Frenkel and Smit 2002, Allen 2004).

6.4.5 Water models

The possibility to model of the dynamic behavior of a system in different solvents is a big advantage of MD. The most often used solvent for water soluble proteins is … water. There are two types of solvent: explicit or implicit.

In explicit water, the solvent is represented by interacting molecules, concerning molecular parameters. Several water models exist, which differ mostly in geometry, dielectric constant, liquid density, heat of vaporization, diffusion coefficient, partial charges and polarizability; the most common are - SPC, TIP (Guillot 2002) - Figure 6.4.

TIP3/TIP3P (Jorgensen *et al.* 1982), 3-site water is used by Yasara. It uses rigid water molecules, interactions in water are modeled by intermolecular forces.

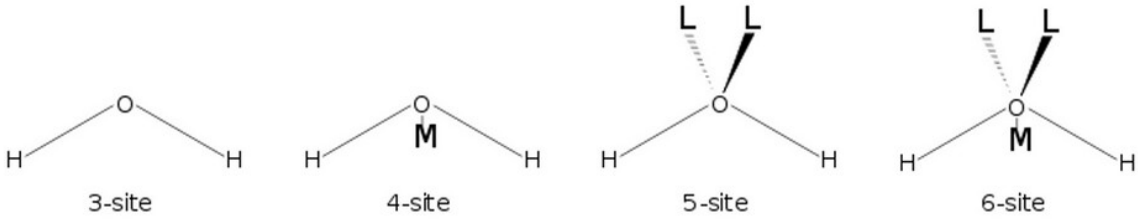

Figure 6.4. Schematic representation of different water models (picture from http://en.wikipedia.org/wiki/Water_model).

In the implicit water model, the solvent is assumed to be a continuum without separation into molecules. The model implemented in Yasara is the COSMO solvation method.

6.4.6 Cut-off

For large systems, as are proteins, not all intramolecular interactions can be included in the calculations, as their calculation is most time-consuming in MD. Therefore a cut off

parameter is used,with interaction between the atoms at the distance more than cut off assumed as zero (van Gunsteren and Berendsen 1990), not contributing to the interactions. All forces within the cutoff distance are calculated according to the equation used for potential energy determination.

However, as long range electrostatics can play a role and might contribute to certain interactions. We can work around that by applying the Particle Mesh Ewald algorithm for the calculation of the Coulomb energy (Essman *et al.* 1995). In this case then electrostatic forces are calculated without cut off. To perform PME calculation simulation cell should be neutralized. The general equation is 6.4f.

$$E_{tot} = \sum_{i,j} \phi(r_j - r_i) = E_{sr} + E_{lr} \qquad (6.4f)$$

General terms in the Ewald summation are common with crystallography (Schlink 2002). The first term in (6.4f) represents the summation in a real space while the second term is the summation in Fourier space, using a discrete lattice for charge density evaluation; **φ** - is an interaction potential. Yasara uses grid spacing less than 1 Å, forth order B-splines, tolerance for a direct space sum depends on the simulation speed.

6.5 ANALYSIS OF RESULTS

To predict the stability of homology model, the complex “substrate-enzyme” and to examine the potential inhibition ability of certain substrates we calculated several parameters: energies, RMSD, interatomic distances, displacement of critical amino acids.

There are several energy types of which can be calculated by Yasara: energy of formation, potential energy (include some components), binding energy, solvation energy.

“The **energy of formation** is the heat released or absorbed (enthalpy change) during the formation of a molecule from its elements, at constant pressure” (Yasara documentation). Energy of formation can be calculated to compare different conformations of one molecule or different molecules built from the same atoms. Lower energy of molecular formation is more favorable.

Potential energy is calculated according to the selected force field (described in chapter 6.4.1).

Solvation energy calculates the interaction energy between the solvent and the solute. This is done by treating the solvent as a continuum without explicit solvent molecules. Yasara uses the boundary element method for the calculation of solvation energy. The electrostatic field is generated by point charges of the solute atoms at all molecular surface

vertices. The actually calculated induced charges at all surface vertices are proportional to the scalar product of the electrostatic field vectors and the surface normals. The induced charges are normalized according to equation (6.5a).

$$q = \sum q_{solute}(e_{solute} - e_{solvent})/(e_{solute} + e_{solvent}) \qquad (6.5a)$$

where $\mathbf{e}_{\mathbf{solute}}$ and $\mathbf{e}_{\mathbf{solvent}}$ are determined dielectric constants, .$\mathbf{q}_{\mathbf{solute}}$ is a solute charge.

Finally, the electrostatic interaction energy between the induced charges and the solute charges is calculated.

Solvation energy strongly depends on the surface, used for normal determination. Older versions of Yasara (version 6) use a molecular surface, while in newer ones it is determined from the solvent accessible surface which leads to the shift of solvation energy to more negative numbers when compared with older calculations,which limits comparison of energies obtained from different versions .

The above equation for the solvation energy doesn't include Van der Waals energy and entropic cost. Van der Waals term can be calculated as (6.5b).

$$E_{VdW}(kcal/mol) = 0.085 + 0.005 \times S_{SolventAccessibleSurface} \qquad (6.5b),$$

where $\mathbf{S}_{\mathbf{SolventAccessibleSurface}}$ is a solvent accessible surface area. The absolute value for the Van der Waals energy ($\mathbf{E}_{\mathbf{VdW}}$) is significant, but its contribution to binding energy is negligible.

A correction taking account for the entropy cost of exposure of atoms to the solvent can be also included in solvation energy. This calculation is based on a simple model for calculation of non electrostatic contribution (Leach 2001) – (6.5c).

$$E_{Entropy}(kcal/mol) = Surf\cos t \times S_{SolventAccessibleSurface} \qquad (6.5c)$$

where Surfcost is a value for the exposing of 0.01nm^2 of surface.

The entropic contribution to the binding energy is practically a constant value, but it can shift the absolute value of the calculated binding energy, mostly to more positive numbers. Value of surfcost (also called atomic solvation parameter) depends on the hydrophobicity of atoms in the solvation surface (Deep and Ahluwalia 2003, Wesson and Eisenberg 1992). In the case of many atoms (over 50000 like in MD) some approximate value is used.

Binding energy by Yasara is a binding energy of the selected object. The more positive the binding energy, the more favorable the interaction for the chosen force field (6.5d):

$$E_{binding} = (E_{Internal} + E_{solvation})_{components} - (E_{Internal} + E_{solvation})_{complex} \qquad (6.5d)$$

where $\mathbf{E_{internal}}$ is a potential energy and $\mathbf{E_{solvation}}$ - solvation, calculated by Boundary fast method (Yasara documentation).

Free energy of binding in AutoDock

AutoDock use free energy of binding for binding evaluation (equation 6.5e). The designation of energy in AutoDock is closer to experimental results.

$$\Delta G_{binding} = \Delta G_{VdW} + \Delta G_{hbond} + \Delta G_{elect} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol} \qquad (6.5e)$$

ΔG_{VdW} - include Van der Waals energy, ΔG_{hbonds} calculates hydrogen bonding change in energy, ΔG_{elect} include electrostatic term. $\Delta G_{conform}$ - deviations from covalent geometry, ΔG_{tor} estimates change in torsional free energy and ΔG_{sol} – includes hydrophobic effect.

All ΔG values on the right side use empirically derived from a set of protein-ligand complexes constant parameters. Some problems with correct calculation of solvation area are reported also for AutoDock (Morris *et al.* 1998).

7. REFERENCES

Agostino M, Jene C, Boyle T, Ramsland PA, Yuriev E. 2009. Molecular docking of carbohydrate ligands to antibodies: structural validation against crystal structures. J Chem Inf Model. 49(12):2749-60.

Allen MP. 2004. Introduction to molecular dynamics simulation. In: Computer soft matter: From Synthetic Polymers to Proteins, Lecture Notes, editors Attig N, Binder K, Grubmuller H, Kremer K. Julich:John von Neumann Institute for Computing. 23:1–28.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403-410.

Bayly CI, Cieplak P, Cornell W, Kollman PA. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. The Journal of Physical Chemistry. 97(40):10269-10280.

Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. 1984. Molecular-dynamics with coupling to an external bath. J Chem Phys. 81(8):3684–3690.

Bohne-Lang A, von der Lieth CW. 2005. GlyProt: in silico glycosylation of proteins. Nucleic Acids Res. 33(Web Server issue):W214-219.

Cole JC, Murray CW, Nissink JW, Taylor RD, Taylor R. 2005. Comparing protein-ligand docking programs is difficult. Proteins. 60(3):325-332. Review.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 117: 5179–5197.

Deep S, Ahluwalia JC. 2003. Theoretical studies on solvation contribution to the thermodynamic stability of mutants of lysozyme T4. Protein Eng. 16(6):415-422.

Dewar MJS, Zoebisch EG, Healey EF, Stewart JJP. 1985. AMI: A new general purpose quantum mechanical molecular model. J Am Chem Soc. 107:3902-3909.

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

Feng D-F, Doolittle RF. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol. 25:351-360.

Fiser A, Do RK, Sali A. 2000. Modeling of loops in protein structures. Protein Sci. 9(9):1753-1773.

Forster MJ. 2002. Molecular modelling in structural biology. Micron 33:365-384.

Frenkel D, Smit B. 2002. Understanding molecular simulation - from algorithms to applications. Academic Press:NewYork.

Guillot B. 2002. A reappraisal of what we have learnt during three decades of computer simulations on water. J Molecular Liquids. 101:219-260.

Hogeweg P, Hesper B. 1984. The alignment of sets of sequences and the construction of phylogenetic trees. An integrated method. J. Mol. Evol. 20:175-186.

Jakalian A, Jack DB, Bayly CI. 2002. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem. 23, 1623-1641.

John B, Sali A. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res. 31(14):3982-3992.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 79:926-935.

Jorgensen WL, Tirado-Rives J. 1988. The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc. 110:1657–1666.

Jung J, Lee B. 2000. Protein structure alignment using environmental profiles. Protein Eng. 13(8):535-43.

Kontoyianni M, McClellan LM, Sokol GS. 2004. Evaluation of docking performance: comparative data on docking algorithms. J Med Chem. 47(3):558-565.

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

Krieger E, Koraimann G, Vriend G. 2002. Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. Proteins. 47:393-402.

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 77(Suppl 9):114-122.

Laskowski RA, McArthur MW, Moss DS. Thorntan JM. 1993. *PROCHECK*: a program to check the stereochemical quality of protein structures. J Appl Crystallog. 26:283.

Leach A. 2001. Molecular modelling: principles and applications. 2nd edition. Dorchested, GB:Prentice Hall.

Lee J, Wu S, Zhang Y. 2009. Ab initio protein structure prediction. From Protein Structure to Function with Bioinformatics, ch.1. Springer: London. 1-26.

Lipman DJ, Pearson WR. 1985. Rapid and sensitive protein similarity searches. Science. 227(4693):1435-1441.

MacKerell Jr AD, Bashford D, Bellott M, Dunbrack Jr RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem. 102:3586-3616.

Marx D, Hutter J. 2000. *Ab initio* molecular dynamics: Theory and implementation. In: Modern methods and algorithms of quantum chemistry. Editor Grotendorst J. Julich:John von Neumann Institute for Computing. 1: 301-449.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. 1998. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J. Computational Chemistry. 19:1639-1662.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 48:443-453.

Nezbeda I. 2007. The Art of Molecular Simulations. From Principles to Applications. Usti nad Labem:Edicni stredicko PF UJEP.

Nissink JW, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R. 2002. A new test set for validating predictions of protein–ligand interaction. Proteins: Structure, Function, and Bioinformatics, 49(4):457–471.

Notredame C, Abergel C. 2003. Using multiple alignment methods to assess the quality of genomic data analysis. In Andrade, M. editor. Bioinformatics and genomics: Current Perspectives. Wymondham, UK:Horizon Scientific Press. 30-55.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302(1):205-17.

O'Sullivan O, Suhre K, Abergel C, Higgins DG, Notredame C. 2004. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. J Mol Biol. 340(2):385-95.

Poirot O, O'Toole E, Notredame C. 2003. Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res. 31(13):3503-3506.

Rost B. 1999. Twilight zone of protein sequence alignments. Protein Engineering. 12(2):85-94.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 234(3):779-815.

Sander S, Schneider R. 1991. Database of homology derived protein structures and the structural meaning of sequence alignment. Proteins: Struct Funct Genet. 9:56-68.

Schlick T. 2002. Molecular Modeling and Simulation: An Interdisciplinary Guide, Edition 1. NY: Springer.

Sippl MJ. 1993. Recognition of errors in three-dimensional structures of proteins. Proteins. 17:355–362.

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33(Web Server issue):W244-248.

Sousa SF, Fernandes PA, Ramos MJ. 2006. Protein-ligand docking: current status and future challenges. Proteins. 65(1):15-26. Review.

St.-Amant A, Cornell WD, Kollman PA, Halgren TA. 1995. Calculation of molecular geometries, relative conformational energies, dipole moments, and molecular electrostatic potential fitted charges of small organic molecules of biochemical interest by density functional theory. J Comput Chem. 16:1483–1506.

Toprakci M, Yelekci K. 2005. Docking studies on monoamine oxidase-B inhibitors: Estimation of inhibition constants (Ki) of a series of experimentally tested compounds. Bioorganic & Medicinal Chemistry Letters. 15(20):4438-4446.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. 2005. GROMACS: fast, flexible, and free. J Comput Chem. 26(16):1701–1718.

Van Gunsteren WE, Berendsen HJC. 1990. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angew Chem lnt Ed Engl. 29:992-1023.

Wang J, Wang W, Kollman PA, Case DA. 2006. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model. 25(2):247-260.

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004. Development and testing of a general AMBER force field. J Comput Chem. 25:1157-1174.

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. 2006. A critical assessment of docking programs and scoring functions. J Med Chem. 49(20):5912-31.

Wesson L, Eisenberg D. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Science. 1:227–235.

Woods RJ. 1998. Computational carbohydrate chemistry: what theoretical methods can tell us. Glycoconjugate Journal. 15**:** 209-216.

Zhang Y. 2008. Progress and challenges in protein structure prediction. Curr Opin Struct Biol. 18(3):342-348.

PART III RESULTS AND CONCLUSIONS

8. SUMMARY OF ARTICLES

The idea for our work came from a screening study that showed NAGALase activity for an enzyme from *A. niger,* and further purification hinted that this activity comes from one enzyme that must be present in a mixture of several GALases from *A. niger.* The hunt for the origin of the NAGALase activity, the identification of the corresponding gene, and the structural and functional characterization of the corresponding enzyme caught our our attention, and so in the next step this NAGALase was biochemically characterized, and in a concerted effort with the experimentalists we started to elucidate the exact structure-functional relationship.

N-terminally sequencing, however, made the story much more complicated than we first thought, as it identified the gene that was in the *A. niger* genome assigned as encoding GALase gene variant A, but we clearly found a NAGALase. Screening the non-redundant primary sequence database of *A. niger* for similar sequences found four other genes with high sequence identity to *AglA*. One of them was assigned as encoding GALase gene variant B, the three other genes had no assigned specific function and were assigned as encoding hypothetical proteins. Sequence alignment of the corresponding amino acid sequences revealed that *aglA* differs from the four others and shows some striking differences in the assumed active site amino acids, common for NAGALases. The amino acid sequence of GALase variant B (*aglB*) instead has more than 50% identity to GALases with known 3D-structure and only 33% identity with the only known 3D structure of NAGALase from chicken by then. However, fully in agreement with the experimental fact that we see NAGALase activity, the enzyme encoded by *aglA* has only 28% identity to the available crystal structure of GALase from *Trichoderma reesei*, but 34% identity to NAGALase from chicken.

Our modeled structure for *aglA* shows one additional domain – ricin-like. The concrete function of this domain is not described for hydrolases, but generally ricin-like domains are identified to bind small sugars. Analysis of our 3D models and comparison to solved crystal structures of characterized enzymes revealed the significant difference in the size of the active centers in *aglA* and *aglB* and lead to an explanation of the specificity for hydrolyzed carbohydrates. Substrate docking clearly demonstrated the preference of *aglA* for α-D-N-acetylgalactosamine over galactose due to the active site extension in the vicinity of the substrates N-acetyl group that we call ‘N-acetyl recognition loop’.

Therefore we can state that the GALase type A gene from *A. niger* does not encode a GALase, and therefore the gene was wrongly assigned, but encodes *A. niger* NAGALase that we were able to characterize structurally and functionally. This whole story is reported in **Article 1.**

Prediction of the ability of HEX to accept modified substrates, that might be potentially interesting for the usage in biosynthesis leading to artificial speciality carbohydrates, is not possible without having structural information at least of the active site. In the lack of crystal structure of any fungal HEX, docking of modified substrates into the active site of HEX from *A. oryzae* can only be performed using a high-quality homology model. The quality of model hereby was not only checked for theoretical correctness, fulfilling steriochemical parameters or Ramachandran distributions, but was additionally checked by computational and experimental methods for stability in solution. Although simple docking of some modified substrates into an active site model of HEX from *A. oryzae* has been reported before, the enzyme lacked a complete structural description. Our modeling attempt included the active dimeric form and all posttranslational modifications – glycosilation and disulfide bridges. **Article 2** reports the structural model, the enzyme-substrate complex for HEX with its natural substrate chitobiose together with the experimental verification.

The basic fold of our model of HEX shows the typical TIM-barrel conserved for GH 20, with one large loop being an exclusive feature not seen before. This loop appeared to be specific for fungi (**Article 2**, **Article 3**) and is situated close to the active site including even one active site amino acid - Trp 482, responsible for aglycon part fixation. As this loop belongs to dimer interface and is located right above the second dimers active site, we proposed the hypothesis that it might play the role of a ‘lid’ regulating substrate access to the active site in the dimer.

Dimerization also revealed a couple of charged residues (2 Arg, 3 His, 3 Glu and 4 Asp) involved in dimer formation. These interactions could explain the reversible pH dependent dimer formation. Raman spectroscopy, together with cross-linking and mass spectrometry, confirmed the 3 modeled disulfide bond formations and their conformations. Hereby, Cys448 and Cys483 could be important for active site stabilization. The secondary structure content of model is in excellent agreement with least-squares analysis of infrared spectra (**Article 2**).

Modeled glycosyl antennas do not participate in the substrate binding or active site protection, however experiments confirmed the role of them in protein stabilization at low pH.

Docking of the natural substrate N, N'-diacetylchitobiose into the active site guided the description of important interactions between substrate and fungal enzyme (**Article 2**): C4 and C5 carboxyl groups of the non-reducing sugar of the substrate form hydrogen bonds with Arg 193, Asp 447, Trp 482 and Glu 519; Asp 345 and Tyr 445 take part in hydrogen bonding with the acetamino group of the non-reducing GlcNAc of chitobiose. The position of the reducing sugar is stabilized by π-π interaction with the aromatic ring of Trp482. Binding is stabilized after 1.6 ns with an average energy during the stable MD of 447.1 kJ/mol calculated by Yasara version 6.30 (without entropic cost).

Our model of HEX from *A. oryzae* was further used in a molecular dynamics study of substrates with a modification at the fourth carbon position C-4 (**Article 4**), as hexosaminidases tolerate the hydroxyl at this position in galacto and gluco conformations (this is why they are called hexosaminidases). The affinity of HEX towards 4-deoxy substrates was studied experimentally and compared with the results of computational study. Our main interest was the prediction and study of the binding behaviour of possible substrates and inhibitors of the mentioned enzyme with computer simulations methods. Hereby we assume, that the stability of the complex “substrate-enzyme” (and with that the possible inhibition ability of certain substrates) is important for predicting the tolerance of this enzyme towards different modification at carbohydrates.

As the initial structure the model with docked natural substrate was used, then the substrate was modified and 4-deoxy substrates were built and optimized: phenyl 2-acetamido-2,4-dideoxy-β-D-xylo-hexopyranoside (1); phenyl 2-acetamido-2,4-dideoxy-β-D-xylo-1,5-dialdo-hexopyranoside (2); p-Nitrophenyl 2-acetamido-2,4-dideoxy-α-L-threo-hex-4-enopyranoside (3) and p-Nitrophenyl 2-acetamido-2,4-dideoxy-α-L-threo-hex-4-enodialdo-1,5-pyranoside (4). For direct comparion with the experimentally used substrates, we used also pNP-GlcNac (p-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside) as standard for the comparison of modified substrates and also compared docking with results for pNP-GalNac (p-nitrophenyl 2-acetamido-2-deoxy-β-D-galactopyranoside). The binding energy depends on the potential energy of the substrate which is also a function of substrate geometry, and experimentally the p-nitrophenyl group is used as an indicator of hydrolysis as after cleavage the product, p-nitrophenol, exhibits a yellow color and can be easily determined using the spectrophotometer. Molecular dynamics simulations were run until full equilibration of the system exhibiting a stabilized enzyme-substrate complex.

Kinetic experiments showed that substrates (3) and (4) were not cleaved at all by any HEX, enzyme from *A. oryzae* was able to cleave just substrate (1). These results are in

agreement with calculated binding energies, which are significantly lower than for the standard and substrate (1). Abrupt changes in binding energy were observed during the MD simulation for substrates (3) and (4), that can be the result of changing of mutual orientation of substrates and active site amino acids accompanied by overcoming of energy barrier.

While some contacts were found to be less critical, and therefore can either be successfully substituted by other interactive pairs, or even might not be present, but influence just the rate of cleavage of the substrate, other are essential and of high importance for hydrolysis proceed, and this is first of all concerning catalytic amino acids.

Substrate (2) and (4) lost an interaction of C3 with Arg 193, π-π stacking of substrates with Trp 482 and Trp 517, which correspond to the correct orientation of pyranose ring into the active site as reported in the **Article 2**.

Alternation of position of substrates (3) and (4) into the active site leads to redistribution of hydrogen bonding contacts with active site amino acids. Distance of glycosidic oxygen from Glu, which acts as acid at the first step of hydrolysis, significantly increased. So deoxydation and dehydration at C-4 position produces an inactive conformation of carbohydrate unable to create the oxazolinium ion intermediate.

Fixation of C3 atom by Arg 193 or Asp 345 appeared to be important for hydrolysis. Interaction with Arg 193 could be successfully substituted by interaction with Asn 309 like in the case of pNP-GalNAc (**Article 4**).

Experimentally, higher yield for the hydrolysis of 4-deoxy substrates by *Penicillinium. oxalicum* were obtained , with a rate of hydrolysis of (1) substrate ~ 25–50% versus 4-9% by *A. oryzae*. Therefore, enzymes from collection strains of *Penicillium oxalicum* (CCF 1959 and CCF 3438) were purified, sequenced and described in **Article 3**. Sequence comparison of both enzymes revealed a high sequence identity, with a similar length of the catalytic domain. To identify the location and coordinates of differing residues and give possible reasons for different activities for modified substrates, a 3D model of hex from *P. oxalicum* was generated and both enzymes were compared at the atomic level.

Structurally *P. oxalicum* is characterized by three disulfide-bonds, confirmed by mass-spectrometry, similar as in fungal HEX from *A. oryzae*, nevertheless with certain differences from human HEX. The catalytic domain has a size and topology similar to *A. oryzae*, with active site amino acids strongly conserved. Five *N*-glycosilation sites were confirmed and structurally four are similar to *A. oryzae*.

Also the large loop that is common for fungal HEXs is present. The general structural analysis and comparison of monomeric models of HEXs from *P. oxalicum* an *A. oryzae* by itself is not enough to explain differences in enzyme kinetics.

However, dimerization, MD and docking experiments by AutoDock unmasked the actual key differences and lead to a structural explanation for completed homology modelling. A possible 'secondary' binding place in the vicinity of the catalytic Tyr 446 residue was found by 'blind' docking for *P. oxalicum*, explaining noncompetitive inhibition by pNP-GlcNAc and GlcNAc. HEX from *A. oryzae* didn't show inhibition due to a different sequence in *P. oxalicum* in this 'secondary' binding pocket.

The kinetics and differences in the rate of hydrolysis of of pNP-GalNAc and pNP-GlcNAc was explained based on the substrate-enzyme interaction analysis during MD. It showed higher binding energy for pNP-GlcNAc that corresponds to the very high yield of product, however at high concentrations of pNP-GlcNAc inhibition by excess of the substrate occurred, which is not observed for pNP-GalNAc.

Changes in the position of docked N-acyl modified substrates during MD influencing they low activity were analyzed for both enzymes. Distances from glycosidic oxygen and catalytic Glu, binding energy and changes in the position of substrates were analyzed. Changes in the binding site in both enzymes initiated by corresponding substrates were similar. This means that the different activity in case of N-acyl modified substrates cannot be explained by sequence/structural differences in binding site.

Dimer building and glycosylation of model of *P .oxalicum* gave us other hints to understanding enzyme kinetics. We compared position of 'lid'-loop of the active site and found it to be more flexible in *P. oxalicum* thanks to the Gly amino acids, explaining easier diffusion of substrates into the binding pocket (**Article 3**). Another major difference between both enzymes is the second loop close to the active site, the 301-312 loop, with the whole loop being in a different conformation in both enzymes differs despite differing by one amino acid only (Figure 8). So how can be the loop conformation and the position of Val 306 be so different in the corresponding loop in the *P.oxalicum* enzyme. The only substitution we find is K302 in *P. oxalicum* to L302 in *A. oryzae*. However, this is not a conservative substitution, and in *P.oxalicum* the positively charged lysine interacts with E307, and this internal hydrogen bond stabilizes the whole loop in a very defined conformation, having a short beta-branched segment in its middle.

```
Poxal  301  PKHTAVEPNPGQ   312
Aoryz  301  PLHTAVEPNPGQ   312
```

Figure 8. Part of multiple sequence alignment

As a result in *P.oxalicum* this stabilized loop keeps a structure far above the binding pocket and in the dimeric structure it gets in contact with the 'lid' loop from the opposite monomer. Together with the lid loop its then forms a large open space right above the active site (**Article 5**).

In *A.oryzae* the position of this loop is different from the beginning, as without the additional stabilization it tends to find a position closer to the protein surface to get additional stabilizing contacts. Thus the differences in the 3D structure of loop 301-312, as a result of the positively charged Lysine302, determine the ability of *P.oxalicum* enzyme to accept bulky charged substrates into the active site. **Article 5** presents a thorough study on the ability of six C-6 modified β-*N*-acetylhexosaminides (aldehyde, uronate, 6-*O*-sulfate, 6-*O*-phosphate) to serve as substrates of β-*N*-acetylhexosaminidases from various sources. In this structure-activity relationship study we investigate the influence of the nature and amount of negative charge in carbohydrate compounds on the affinity of β-*N*-acetylhexosaminidases and on the stimulation of NK cells.

The knowledge about HEXs, gained by *in vitro* experiments and computational work was reviewed in **Article 6** that describes the application of different bioinformatical tools for enzyme studies, and its correlation with experiments. AutoDock, Yasara and Sybyl results of docking and MD of HEX from *A. oryzae* are compared.

Homology modeling can be still helpful even in the case of low overall sequence identity, if the catalytic domain is structurally conserved, like in the case of nitrilases, as we demonstrate on the example of nitrilase from *A. niger* in the **Article 7**. The procedure for model building is similar to the HEXs and GAL/NAGALases (**Article 1**). The multiple sequence structure based alignment was corrected based on secondary structure predictions due to the low identity. Two different template sets were used for model building. 3D arrangement of residues 316-356 is lost in templates. *A. niger* nitrilase has three loop regions that are longer than found in the template structures, similar to nitrilase from *Rhodococcus rhodochrous* J1.

Positioning of benzonitrile in the assumed active site was done using AutoDock, however the study of this substrate-enzyme complex by MD is difficult because such as small substrate has a weak interaction.

This enzyme is found to exist in helical-like structures. Electronic micrographs of them helped us to model multimer if nitrilase from *A. niger*, formed by approximately 16 subunits, which appeared to be smaller than other nitrilases. Similar open ‘C-like’ structure was seen in micrographs of mutants of nitrilase from *Rhodococcus rhodochrous* J1.

9. CONCLUSIONS

1) Our study gives a clear structural explanation of the NAGALase activity found the enzyme that originally was assigned as GALase encoded by gene *aglA* from *A. niger.* Its ability to accept an N-acetyl group at the C-2 position of alpha-galactose is due to an extension of the binding pocket in the active site. This theoretical study leads to the correct classification of the gene *aglA.*

2) We propose a full model for HEX from *A. oryzae* including dimerization and glycosylation, and describe the dynamic interaction of active site amino acids with the natural substrate chitobiose.

3) The importance of the C3-carbon fixation in the non-reducing sugar within the active site was demonstrated. Deoxydation and dehydration of the C-4 carbon of the pyranosyl ring leads to an inactive substrate conformation.

4) The three-dimensional structure of HEX from *P. oxalicum* was described and compared with *A. oryzae*. Generally, 3D structures of both enzymes are similar. Observed differences in the substrate affinity to N-acetyl modified substrates can be explained by easier penetration of substrates into the active site of *P. oxalicum* due to differences in the large loop of *P. oxalicum* that is significantly more flexible.

5) Another significant difference between both enzymes are the two types of inhibition observed experimentally in *P. oxalicum*, by product (GlcNAc and GalNAc) and substrate pNP-GlcNAc access. A secondary binding place was identified for *P. oxalicum*, with the specific seqence that differs from *A. oryzae* in this place enabling GlcNAc binding. A single lysine residue in the second loop close to the active site changes and stabilizes the conformation of this loop, and therefore *P. oxalicum* accepts bulky charged substrates into the active site.

6) Fungal nitrilase from *A. niger* was modeled. This model was used to fit into into low resolution electron micrographs of multimeric structures formed by this enzyme. A possible multimerization mode was described, with the enzyme forming multimer with 16 units.

10. LIST OF ARTICLES IN IMPACTED JOURNALS

Article 1

Kulik N, Weignerova L, Filipi T, Pompach P, Novak P, Mrazek H, Slamova K, Bezouska K, Kren V, Ettrich R. 2010. The α-galactosidase type A gene *aglA* from *Aspergillus niger* encodes a fully functional α-N-acetylgalactosaminidase. Glycobiology. 20(11):1410-1419.

Article 2

Ettrich R, Kopecky V Jr, Hofbauerova K, Baumruk V, Novak P, Pompach P, Man P, Plihal O, Kuty M, Kulik N, Sklenar J, Ryslava H, Kren V, Bezouska K. 2007. Structure of the dimeric N-glycosylated form of fungal beta-N-acetylhexosaminidase revealed by computer modeling, vibrational spectroscopy, and biochemical studies. BMC Struct Biol. 7:32.

Article 3

Ryslava H, Kalendova A, Doubnerova V, Skocdopol P, Hodkova A, Kukacka Z, Slamova K, Bojarova P, Pompach P, Kumar V, Kulik N, Ettrich R, Kren V and Bezouska K. Enzymatic characterization and molecular modeling of an evolutionary interesting fungal beta-N-acetylhexosaminidase. FEBS Journal, accepted, 5.5.2011, *in press.*

Article 4

Slamova K, Gazak R, Bojarova P, Kulik N, Ettrich R, Pelantova H, Sedmera P, Kren V. 2010. 4-Deoxy-substrates for beta-N-acetylhexosaminidases: how to make use of their loose specificity. Glycobiology. 20(8):1002-1009.

Article 5

Bojarová P., Slámová K., Křenek K., Gažák R., Kulik N., Ettrich R., Pelantová H., Kuzma M., Riva S., Adámek D., Bezouška K., and Křen V. 2011. Charged hexosaminides as ligands of NK cell activation receptors – concerted chemo-enzymatic synthesis, Advanced Synthesis and Catalysis, *invited paper, submitted May2011*

Article 6

Kulik N, Slamova K. Computational modelling of catalytic properties and modified substrates of fungal β-*N*-acetylhexosaminidases. Mini-Reviews in Organic Chemistry, *in press* (special issue scheduled for 9/2011).

Article 7

Kaplan O, Bezouska K, Plihal O, Ettrich R, Kulik N, Vanek O, Kavan D, Benada O, Malandra A, Sveda O, Vesela AB, Rinagelova A, Slamova K, Cantarella M, Felsberg J, Duskova J, Dohnalek J, Kotik M, Kren V, Martinkova L. 2011. Heterologous expression, purification and characterization of nitrilase from Aspergillus niger K10. BMC Biotechnol. 11(1):2.

APPENDIX

List of papers with abstracts

Article 1: The α-galactosidase type A gene *aglA* from *Aspergillus niger* encodes a fully functional α-N-acetylgalactosaminidase. Kulik N, Weignerova L, Filipi T, Pompach P, Novak P, Mrazek H, Slamova K, Bezouska K, Kren V, Ettrich R. 2010. Glycobiology. 20(11):1410-1419.

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

Article 2: **Structure of the dimeric N-glycosylated form of fungal beta-N-acetylhexosaminidase revealed by computer modeling, vibrational spectroscopy, and biochemical studies.** Ettrich R, Kopecky V Jr, Hofbauerova K, Baumruk V, Novak P, Pompach P, Man P, Plihal O, Kuty M, Kulik N, Sklenar J, Ryslava H, Kren V, Bezouska K. 2007. BMC Struct Biol. 7:32.

Abstract: BACKGROUND: Fungal beta-N-acetylhexosaminidases catalyze the hydrolysis of chitobiose into its constituent monosaccharides. These enzymes are physiologically important during the life cycle of the fungus for the formation of septa, germ tubes and fruit-bodies. Crystal structures are known for two monomeric bacterial enzymes and the dimeric human lysosomal beta-N-acetylhexosaminidase. The fungal beta-N-acetylhexosaminidases are robust enzymes commonly used in chemoenzymatic syntheses of oligosaccharides. The enzyme from *Aspergillus oryzae* was purified and its sequence was determined.

RESULTS: The complete primary structure of the fungal beta-N-acetylhexosaminidase from *Aspergillus oryzae* CCF1066 was used to construct molecular models of the catalytic subunit of the enzyme, the enzyme dimer, and the N-glycosylated dimer. Experimental data were obtained from infrared and Raman spectroscopy, and biochemical studies of the native and deglycosylated enzyme, and are in good agreement with the models. Enzyme deglycosylated under native conditions displays identical kinetic parameters but is significantly less stable in acidic conditions, consistent with model predictions. The molecular model of the

deglycosylated enzyme was solvated and a molecular dynamics simulation was run over 20 ns. The molecular model is able to bind the natural substrate - chitobiose with a stable value of binding energy during the molecular dynamics simulatio.

CONCLUSION: Whereas the intracellular bacterial beta-N-acetylhexosaminidases are monomeric, the extracellular secreted enzymes of fungi and humans occur as dimers. Dimerization of the fungal beta-N-acetylhexosaminidase appears to be a reversible process that is strictly pH dependent. Oligosaccharide moieties may also participate in the dimerization process that might represent a unique feature of the exclusively extracellular enzymes. Deglycosylation had only limited effect on enzyme activity, but it significantly affected enzyme stability in acidic conditions. Dimerization and N-glycosylation are the enzyme's strategy for catalytic subunit stabilization. The disulfide bridge that connects Cys448 with Cys483 stabilizes a hinge region in a flexible loop close to the active site, which is an exclusive feature of the fungal enzymes, neither present in bacterial nor mammalian structures. This loop may play the role of a substrate binding site lid, anchored by a disulphide bridge that prevents the substrate binding site from being influenced by the flexible motion of the loop

Article 3: **Enzymatic characterization and molecular modeling of an evolutionary interesting fungal beta-N-acetylhexosaminidase.** Ryslava H, Kalendova A, Doubnerova V, Skocdopol P, Hodkova A, Kukacka Z, Slamova K, Bojarova P, Pompach P, Kumar V, Kulik N, Ettrich R, Kren V and Bezouska K. FEBS Journal, accepted, 5.5.2011, *in press.*

SUMMARY: Fungal β-N-acetylhexosaminidases are inducible extracellular enzymes with many biotechnological applications. The enzyme from *Penicillium oxalicum* has unique enzymatic properties despite its close evolutionary relationship with other fungal hexosaminidases. It has high GalNAcase activity, better tolerates substrates with the modified N-acyl group, and has some other unusual catalytic properties. In order to understand these features, we performed isolation, biochemical and enzymological characterization, molecular cloning, and molecular modeling. The native enzyme is composed of two catalytic units (65 kDa each) and two propeptides (15 kDa each), yielding a molecular weight of 160 kDa. Enzyme deglycosylated by endoglycosidase H had comparable activity, but reduced stability. We have cloned and sequenced the gene coding for the entire hexosaminidase from *P. oxalicum*. Sufficient sequence identity of this hexosaminidase with the structurally solved enzymes from bacteria and humans with complete conservation of all catalytical residues allowed us to construct a molecular model of the enzyme. Results from molecular dynamics simulations and substrate docking supported the experimental kinetic and substrate specificity data and provided a molecular explanation for why the hexosaminidase from *P. oxalicum* is unique among the family of fungal hexosaminidases.

Article 4: 4-Deoxy-substrates for beta-N-acetylhexosaminidases: how to make use of their loose specificity. Slamova K, Gazak R, Bojarova P, Kulik N, Ettrich R, Pelantova H, Sedmera P, Kren V. 2010. Glycobiology. 20(8):1002-1009.

Abstract: Beta-N-Acetylhexosaminidases feature so-called wobbling specificity, which means that they cleave substrates both in gluco- and galacto- configurations, with the activity ratio depending on the enzyme source. Here we present the new finding that fungal beta-N-acetylhexosaminidases are able to hydrolyze and transfer 4-deoxy-N-acetylhexosaminides with high yields. This clearly demonstrates that the 4-hydroxy moiety at the substrate pyranose ring is not essential for substrate binding to the enzyme active site, which was also confirmed by molecular docking of the tested compounds into the model of the active site of beta-N-acetylhexosaminidase from *Aspergillus oryzae*. A set of four 4-deoxy-N-acetylhexosaminides was synthesized and screened against a panel of beta-N-acetylhexosaminidases (extracellular and intracellular) from various sources (fungal, human, animal, plant and bacterial) for hydrolysis. The results of this screening are reported here, as well as the structures of three novel 4'-deoxy-disaccharides prepared by transglycosylation reaction with high yields (52% total disaccharide fraction) using beta-N-acetylhexosaminidase from *Talaromyces flavus*.

Article 5: Charged hexosaminides as ligands of NK cell activation receptors – concerted chemo-enzymatic synthesis. Bojarová P., Slámová K., Křenek K., Gažák R., Kulik N., Ettrich R., Pelantová H., Kuzma M., Riva S., Adámek D., Bezouška K., and Křen V. 2011. Advanced Synthesis and Catalysis, *invited paper, submitted May 2011*

Abstract. This work is a structure-activity relationship study that investigates the influence of the nature and amount of negative charge in carbohydrate substrates on the affinity of beta-*N*-acetylhexosaminidases and on the stimulation of natural killer cells. It describes synthetic procedures yielding novel glycosides useful in immunoactivation. Specifically, we present a thorough study on the ability of six C-6 modified beta-*N*-acetylhexosaminides (aldehyde, uronate, 6-*O*-sulfate, 6-*O*-phosphate) to serve as substrates for cleavage and glycosylation by a library of beta-*N*-acetylhexosaminidases from various sources. Four novel disaccharides with one or two (negatively) charged groups were prepared in synthetic reactions in good yields. Surprisingly, 6-*O*-phosphate, though cleaved by a number of enzymes from the series, worked neither as a donor nor as an acceptor in transglycosylation reactions.

The results of wet experiments were supported by molecular modelling of substrates in the active site of two representative enzymes from the screening. All ten prepared compounds were examined from the viewpoint of their immunoactivity, namely as ligands of two activation receptors of natural killer (NK) cells, NKR-P1 and CD69, both with isolated proteins and whole cells. Furthermore, especially sulfated disaccharides acted as very efficient protectants of NK cells against activation-induced apoptosis and as stimulants of natural killing of resistant tumor cells, which makes them good candidates for potential clinical use in anti-cancer treatment.

Article 6: Computational modelling of catalytic properties and modified substrates of fungal β-*N*-acetylhexosaminidases. Kulik N, Slamova K. Mini-Reviews in Organic Chemistry, *in press* (special issue scheduled for 9/2011).

Abstract: Besides their implication in human physiology and disease, β-N-acetylhexosaminidases (EC 3.2.1.52, CAZy GH 20) have recently gained a lot of attention thanks to their great potential in the enzymatic synthesis of carbohydrates and glycomimetics. Extracellular β-N-acetylhexosaminidases from filamentous fungi proved to be a powerful synthetic tool for the preparation of both natural andmodified glycosides under mild conditions with good yields. A homology model of β-N-acetylhexosaminidase from the filamentous fungus *Aspergillus oryzae* has recently been reported, and its quality was corroborated by vibrational spectroscopy and biochemical studies. Computational modelling and analysis helped to identify active site amino acids and other basic structural features of this enzyme important in the catalytic process; moreover, the surface interactions of the subunits of the glycosylated enzyme were identified. The model of β-N-acetylhexosaminidase from *Aspergillus oryzae* prepared the ground for further *in silico* studies of enzyme-substrate complexes including prediction and explanation of its substrate specificity.

Article 7: Heterologous expression, purification and characterization of nitrilase from *Aspergillus niger* K10. Kaplan O, Bezouska K, Plihal O, Ettrich R, Kulik N, Vanek O, Kavan D, Benada O, Malandra A, Sveda O, Vesela AB, Rinagelova A, Slamova K, Cantarella M, Felsberg J, Duskova J, Dohnalek J, Kotik M, Kren V, Martinkova L. 2011. BMC Biotechnol. 11(1):2.

Abstract

BACKGROUND: Nitrilases attract increasing attention due to their utility in the mild hydrolysis of nitriles. According to activity and gene screening, filamentous fungi are a rich source of nitrilases distinct in evolution from their widely examined bacterial counterparts. However, fungal nitrilases have been less explored than the bacterial ones. Nitrilases are typically heterogeneous in their quaternary structures, forming short spirals and extended filaments, these features making their structural studies difficult.

RESULTS: A nitrilase gene was amplified by PCR from the cDNA library of *Aspergillus niger* K10. The PCR product was ligated into expression vectors pET-30(+) and pRSET B to construct plasmids pOK101 and pOK102, respectively. The recombinant nitrilase (Nit-ANigRec) expressed in Escherichia coli BL21-Gold(DE3)(pOK101/pTf16) was purified with an about 2-fold increase in specific activity and 35% yield. The apparent subunit size was 42.7 kDa, which is approx. 4 kDa higher than that of the enzyme isolated from the native organism (Nit-ANigWT), indicating post-translational cleavage in the enzyme's native environment. Mass spectrometry analysis showed that a C-terminal peptide (Val327 - Asn_{356}) was present in Nit-ANigRec but missing in Nit-ANigWT and Asp_{298}-Val_{313} peptide was shortened to Asp_{298}-Arg_{310} in Nit-ANigWT. The latter enzyme was thus truncated by 46 amino acids. Enzymes Nit-ANigRec and Nit-ANigWT differed in substrate specificity, acid/amide ratio, reaction optima and stability. Refolded

recombinant enzyme stored for one month at 4°C was fractionated by gel filtration, and fractions were examined by electron microscopy. The late fractions were further analyzed by analytical centrifugation and dynamic light scattering, and shown to consist of a rather homogeneous protein species composed of 12-16 subunits. This hypothesis was consistent with electron microscopy and our modelling of the multimeric nitrilase, which supports an arrangement of dimers into helical segments as a plausible structural solution.

CONCLUSIONS: The nitrilase from *Aspergillus niger* K10 is highly homologous (≥86%) with proteins deduced from gene sequencing in *Aspergillus* and *Penicillium* genera. As the first of these proteins, it was shown to exhibit nitrilase activity towards organic nitriles. The comparison of the Nit-ANigRec and Nit-ANigWT suggested that the catalytic properties of nitrilases may be changed due to missing posttranslational cleavage of the former enzyme. Nit-ANigRec exhibits a lower tendency to form filaments and, moreover, the sample homogeneity can be further improved by in vitro protein refolding. The homogeneous protein species consisting of short spirals