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

Natallia Kulik2,3, Lenka Weignerová4,7, Tomáš Filipi4, Petr Pompach5,7, Petr Novák5,7, Hynek Mrázek6,7, Kristýna Slámová4, Karel Bezouška6,7, Vladimír Křen1,4, and Rüdiger Ettrich1,2,3

1Center for Biocatalysis and Biotransformation, Institute of Systems Biology and Ecology, Academy of Sciences of the Czech Republic, Zámek 136, CZ 373 33 Nové Hrady, Czech Republic; 2Institute of Physical Biology, University of South Bohemia, Zámek 136, CZ 373 33 Nové Hrady, Czech Republic; 3Laboratory of Molecular Structure Characterization, and 5Laboratory of Protein Architecture, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, CZ 142 20 Praha 4, Czech Republic; and 6Laboratory of Biochemistry, Faculty of Science, Charles University in Prague, Albertov 2030, CZ 128 40 Praha 2, Czech Republic

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

Introduction

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

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

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

Results

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

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

1To whom correspondence should be addressed: Tel: +420386361297; Fax: +420386361279; ettrich@nh.usbe.cas.cz (R. Ettrich), kren@biomed.cas.cz (V. Křen).
characterized and, therefore, are assigned to hypothetical proteins with unknown function (gene IDs 4984860, 4978099, and 4987036). The lengths of the hypothetical open reading frames are similar to aglB and vary from 391 to 431 amino acids.

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

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

The purity of the isolated α-N-acetylgalactosaminidase was verified by 2D electrophoresis using narrow pl strips. During this analysis, the enzyme resolved into two spots, differing only in the presence or absence of three N-terminal amino acids (Ser–Ile–Glu) as demonstrated by N-terminal sequencing (Figure 1A). Mass spectrometry techniques were used for additional sequencing that due to their sensitivity used less of the limited amount of protein material available. The results allowed us to conclude that the protein obtained is encoded by
the aglA gene in the A. niger genome. The overall sequence coverage was 67% and thus led to an unambiguous assignment (Figure 1B).

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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3}
\caption{Structure of the $\alpha$-N-acetylgalactosaminidase from \textit{Aspergillus niger}. (A) Overall fold shows a TIM-barrel with the active site at the N-terminus, a small domain of eight antiparallel $\beta$-strands packed in $\beta$-sandwich in the middle, and a ricin-like domain on the right. The generated model (magenta) is overlaid with the crystal structure of the homologs $\alpha$-N-acetylgalactosaminidase from chicken (yellow), with $\alpha$-N-acetylgalactosamine and the ricin-like domain from the xylanase from \textit{Streptomyces olivaceoviridis} E-86 (green). (B) and (C) Molecular surface of the active site of \textit{aglA} enzyme (magenta) and \textit{aglB} enzyme (yellow) with oNP-$\alpha$-GalNAc. The active site of \textit{aglA} enzyme has extra space for accommodating the N-acetyl-group of the substrate, while in \textit{aglB} enzyme this space is occupied by Trp205.}
\end{figure}
recently released crystal structure of the human enzyme would not change the modeling results, nor would it affect the 3D arrangement in the active site. Therefore, at present we can state that the closest probable structure for the enzyme encoded by gene aglB is unambiguously α-galactosidase, whereas for the enzyme encoded by the aglA gene the primary sequence analysis shows a slightly higher probability that its 3D structure is closer to that of α-N-acetylgalactosaminidase (Figure 2).

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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Secondary structure content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α-Helix</td>
</tr>
<tr>
<td>aglA</td>
<td>Before refinement</td>
</tr>
<tr>
<td></td>
<td>After refinement</td>
</tr>
<tr>
<td>aglB</td>
<td>Before refinement</td>
</tr>
<tr>
<td></td>
<td>After refinement</td>
</tr>
</tbody>
</table>

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

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

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

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

**Discussion**

The structural models constructed in this study were indispensable in allowing to explain the inability of \( \alpha \)-galactosidase \( \text{aglB} \) from \( \text{A. niger} \) to accommodate the substrate containing a 2-acetamido group in the active site in a stable and proper position. Two reasons why the \( \alpha \)-galactosidase \( \text{aglB} \) is unable to hydrolyze \( \text{oNP-\alpha-GalNAc} \) have emerged: the lower binding affinity and the sterical hindrance connected with positioning of the C1-O bond close to the catalytic residue that makes it impossible for the hydrolytic reaction to take place. The computer modeling experiments enabled us to assign the protein possessing the dual activity to the enzyme encoded by gene
aglA. This is a completely new view of this gene since a possible \( \alpha-N\)-acetylglactosaminidase activity has never been mentioned in the literature before. Our experimental results show that this protein exhibits dual enzyme activity and is able to hydrolyze 2-NP-\( \alpha \)-GalNAc and 4-NP-\( \alpha \)-Gal with 10 times better cleavage of 2-NP-\( \alpha \)-GalNAc than that of 2-NP-\( \alpha \)-Gal. This contradicts with its original assignment as an \( \alpha \)-D-galactopyranosidase belonging to GH family 27. Substrate docking into the structural model of this enzyme confirmed similar binding energies for both substrates with a clear preference for the \( \alpha \)-N-acetyl-D-galactosaminide over the galactoside (Figure 3B and C), thus providing evidence that this gene from \( A. \) niger does not encode exclusively an \( \alpha \)-galactosidase type A but that rather one combined with a fully functional \( \alpha \)-N-acetylglactosaminidase.

Figure 3B shows the binding pocket of aglA with bound \( \alpha \)-N-acetylgalactosamine, in comparison to the binding pocket of aglB with bound \( \alpha \)-galactose in Figure 3C. The enlarged binding pocket in aglA, which is able to accommodate an C2 acetamido group, is a result of the amino acid sequence Ser170, Ala171, Pro172, Ala173, and Tyr174, which forms a longer loop following β5 and connecting to α5 (Figure 2). Instead of Thr205, which blocks the part of the binding pocket docking the acetamido group in \( \alpha \)-galactosidases, amino acid residues Ser170, Ala173, and Tyr174 (cf. Figure 4A and B) on this “\( \alpha \)-acetyl recognition loop” create an open space that becomes part of the enlarged binding pocket of \( \alpha \)-N-galactosaminidase. The presence of this loop explains the fact that the attempt to model aglA taking \( \alpha \)-galactosidase as a template did not lead to a stable enzyme but induced distortions in the active site during molecular dynamics simulations. The multiple sequence alignment of the \( \alpha \)-N-acetylglactosaminidase/\( \alpha \)-galactosidase homologs in \( A. \) niger (Figure 6) identified conserved residues at positions corresponding to the active site of aglB for all hypothetical proteins, with the exception of a tyrosine and cysteine in AnL4g01800 instead of Thr17 and Asp55 in aglB. AglA is the only homolog having the above described “\( \alpha \)-acetyl recognition loop”. Lacking this loop, all other found potential \( \alpha \)-galactosidases in the \( A. \) niger genome could have \( \alpha \)-galactosidase activity but definitely not \( \alpha \)-N-galactosaminidase activity. This is in accordance with our experimental findings demonstrating that the other \( \alpha \)-galactosidases isolated from \( A. \) niger did not show dual activity but are pure \( \alpha \)-galactosidases.

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

Therefore, we propose that the \( \alpha \)-N-acetylglactosaminidase/\( \alpha \)-galactosidase encoded by aglA utilizes a classical double inversion reaction mechanism (Figure 4C) analogous to homologous \( \alpha \)-galactosidases (Garman et al. 2002; Golubev et al. 2004) and that this mechanism is the same for both of its activities, e.g., \( \alpha \)-N-acetylglactosaminidase and \( \alpha \)-galactosidase.
factors are taken from the General AMBER AutoSMILES approach. The corresponding bond, angle, and SARA, force field parameters were assigned using the 2 ns of MD simulation in water (pH 7, TIP3P water model) with SARA (Krieger et al. 2004) by energy minimization followed by (Wiederstein and Sippl 2007). The models were refined in YASARA package (Sali and Blundell 1993) and checked with ProSA. Hydrogen atoms were built and examined using Modeller 9.1.

The mass spectrometric experiment was performed on a commercial APEX-Qe Fourier Transform Mass Spectrometry (FTMS) Lys/Arg instrument equipped with a 9.4-T superconducting magnet and Combi ESI/MALDI ion source (Bruker Daltonics, Billerica, MA, USA). Mass spectra were obtained by accumulating ions in the collision hexapole and running the quadrupole mass filter in non mass-selective (rf-only) mode so that a broad range of ions (300–2500 m/z) were passed to the FTMS analyzer cell. The accumulation time in the cell was set at 0.5 s, the cell was opened for 400 μs, and 256 experiments were collected for one liquid chromatography run, where one experiment consisted of the accumulation of five spectra. The acquisition time was set to 512 k points at m/z 300 a.m.u. The instrument was externally calibrated using triply and doubly charged ions of angiotensin I and quadruply, quintuply, and sextuply charged ions of bovine insulin. This calibration typically results in a mass accuracy below 2 ppm.

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

**Molecular modeling**

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

Interaction energies were calculated using a similar method to that applied earlier for substrate–hexosaminidase complexes (Ettrich et al. 2007), considering the internal energy obtained with the specified force field, as well as the electrostatic energy and the entropy cost of fixing the ligand implicitly including Van der Waals solvation energy (Bultinck et al. 2003). The solvation energy was calculated using the boundary fast method implemented in YASARA. The boundary between solvent (dielectric constant 80) and solute (dielectric constant 1) was formed by the latter’s molecular surface, constructed with a solvent probe radius of 1.4 Å and the following radii for the solute elements: polar hydrogens 0.32 Å, other hydrogens 1.017 Å, carbon 1.8 Å, oxygen 1.344 Å, nitrogen 1.14 Å, and sulfur 2.0 Å. The solute charges were assigned based on the YASARA 2 force field, using GAFF/AM1BCC for the ligands. An estimate of the entropy cost of exposing 0.01 nm² to solvent was calculated as follows: \( E_{\text{ent}} = S_{\text{SAS}} \cdot \text{solvation entropy} \), where \( S_{\text{SAS}} \) is the solvent accessible area, and the solvation entropy characterizes the entropy cost of exposing 0.01 nm² of surface and can vary. For protein structures, the solvation entropy is usually estimated to be approximately 0.65 kJmol⁻¹Å⁻². It is almost impossible to calculate this entropy cost accurately, but this is fortunately not needed, since it mainly depends on characteristics that are constant during the simulation (ligand and protein size, side chains on the surface, etc.) and thus is a constant factor. Binding energies calculated in this way might therefore be shifted by an unknown amount that depends on the protein; however, their relative values are correct. The more positive interaction energy, the more favorable is the interaction in the context of the chosen force field. Ligand–enzyme complexes after 10 ns of MD were used for measuring binding energies by Autodock.

**Acknowledgments**

Support by the Academy of Sciences of the Czech Republic (AVOZ50200510, AVOZ60870520), the Ministry of Education of the Czech Republic (LC 06010, LC 7017,
aGLA from Aspergillus niger encodes α-N-acetylgalactosaminidase

MSM6007665808, MSM21620808), and by the Czech Science Foundation grants P207/10/0321 to L.W., P207/10/1934 to R.E., P207/10/1040 to P.N., and 303/09/0477 and 305/09/H008 to K.B. is acknowledged. Access to METACentrum supercomputing facilities was provided under the research intent MSM638391701.

Conflict of interest statement

None declared.

Abbreviations

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

References


