Regular paper

A computational approach to structural properties of glycoside hydrolase family 4 from bacteria

Dana Craciun[1], Beatrice Vlad-Oros[2,3,4], Nicoleta Filimon[2,3], Vasile Ostafe[2,3] and Adriana Isvoran[2,3]✉

[1]West University of Timisoara, Teacher Training Department, Timisoara, Romania; [2]West University of Timisoara, Department of Biology and Chemistry, Timisoara, Romania; [3]Laboratory of Advanced Researches in Environmental Protection, Timisoara, Romania; [4]University „Al. I. Cuza" from Lasi, Lasi, Romania (temporally affiliated)

Structural bioinformatics approaches applied to the alpha- and beta-glycosidases from the GH4 enzyme family reveal that, despite low sequence identity, these enzymes possess quite similar global structural characteristics reflecting a common reaction mechanism. Locally, there are a few distinctive structural characteristics of GH4 alpha- and beta-glycosidases, namely, surface cavities with different geometric characteristics and two regions with highly dissimilar structural organizations and distinct physicochemical properties in the alpha- and beta-glucosidases from *Thermotoga maritima*. We suggest that these structurally dissimilar regions may be involved in specific protein-protein interactions and this hypothesis is sustained by the predicted distinct functional partners of the investigated proteins. Also, we predict that alpha- and beta-glycosidases from the GH4 enzyme family interact with difenoconazole, a fungicide, but there are different features of these interactions especially concerning the identified structurally distinct regions of the investigated proteins.

INTRODUCTION

Glycoside hydrolases (GH), also called glycosidases, are enzymes that catalyze the hydrolysis of the glycosidic bonds releasing smaller sugars. They play important roles both in nature and in industrial processes that involve biological conversion of the biomass to fuels, allowing fuel production with reduced costs (Yang *et al.*, 2011). There are two types of glycosidases: alpha-glycosidase that acts on 1-4 linked alpha-glucose residues and beta-glycosidase acting upon beta 1-4 bonds linking two glucose or glucose-substituted molecules (McCarter and Withers, 1994). The CAZy (Carbohydrate-Active enZYmes) database (Cantarel *et al.*, 2009) contains a sequence-based classification for at least 130 families of glycoside hydrolases. Glycoside hydrolase family 4 (GH4) enzymes are called glucosidases. They represent a special group of glycosidases, which includes both alpha- and beta-glucosidases, and displays a reaction mechanism involving NAD+ and divalent metal ion cofactors (Lodge *et al.*, 2003). The group comprises alpha-glucosidases, alpha-galactosidases, alpha-glucuronidases, 6-phospho-alpha-glucosidases, and 6-phospho-beta-glucosidases. The majority of GH4 enzymes are of bacterial origin, the 6-phospho-beta-glucosidase from *Thermotoga maritima* (Yip & Withers, 2006) and 6-phospho-alpha-glucosidase from *Bacillus subtilis* (Yip *et al.*, 2007) being the most studied enzymes of this family.

The genome of thermophilic bacteria, such as *Thermotoga maritima* and *Geobacillus stearothermophilus*, encodes a number of glycosidic enzymes that are involved in sugar and polymer catabolism by anaerobic fermentation. The by-products are carbon dioxide and hydrogen gas, the latter being used as a fuel (Chhabra *et al.*, 2002, Ugwuanyi, 2008). These bacteria grow in hot waters of 35–90°C and their carbohydrate active enzymes are thermostable, which explains their numerous biotechnological applications (Conners *et al.*, 2006). It also explains extensive studies concerning biochemical and structural features of thermophilic bacteria carbohydrate active enzymes.

Recent studies have proven that fungal and bacterial beta-glucosidases show favorable properties to be used in biotechnological applications (Del Pozo *et al.*, 2012; Pei *et al.*, 2012).

Crystallographic data are available for some of bacteria GH4 enzymes. The Protein Data Bank (Berman *et al.*, 2000) contains 9 entries for structures of glycosidic enzymes and their complexes from: *Thermotoga maritima* (5, but 3 of them refer to the same protein in different complexes), *Thermotoga neapolitana (1)*, *Bacillus subtillis* (2) and *Geobacillus stearopthermophilus* (1). Accordingly, we consider only 7 structural files in our study, as explained in the "Material and methods" section.

It is well known that enzymatic activity is influenced by structural features of both the enzyme and the substrate. This is also true for the enzymes hydrolyzing the glycosidic bonds, but due to the complexity of these enzymes and carbohydrate polymers, the mechanisms involved in glycosidic bond hydrolysis are still not fully understood. For example, the comparison of 6-phospho-beta-glucosidase of *Thermotoga maritima* with 6-phospho-alpha-glucosidase from *Bacillus subtilis* reveals a high degree of structural similarity between the two enzymes reflecting a possible common reaction mechanism for

---

Table 1. PDB and UniProt entry codes and EC numbers for GH4 enzymes with known structures.

No.	Structural file description	SwissProt code entry	Amino acids in sequence	PDB code entry	EC number	Observations
1	alpha-glucosidase from *Thermotoga maritima* in complex with maltose and NAD⁺	O33830	480	1OBB	EC 3.2.1.20	the 3D structure is complete
2	alpha-glucoronidase from *Thermotoga maritima* in complex with NAD⁺	Q9WZL1	471	1VJT	EC 3.2.1.139	residues 470-471 are not located in the 3D structure
3	putative alpha-glucosidase from *Thermotoga neapolitana*	B9KAM3	469	3U95	Not available	residues 469-477 are not located in the 3D structure of chain A
4	6-phospho-alpha-glucosidase A from *Bacillus subtilis* in complex with manganese, NAD⁺ and glucose-6-phosphate	P54716	449	1U8X	EC 3.2.1.122	residues 235-241 and 446-449 are not located in the 3D structure
5	putative alpha-glucosidase (alpha-galacturonidase) from *Bacillus subtilis*	P39130	454	3FEF	EC 3.2.1.67	residues 1-6 and 441-454 are not located in the 3D structure
6	6-phospho-beta-glucosidase from *Thermotoga maritima* in complex with manganese, NAD⁺ and glucose-6-phosphate	Q9X108	415	1UP4, 1UP6, 1UP7	EC 3.2.1.86	residues 219-221 are not located in the 3D structure
7	6-phospho-beta-glucosidase from *Geobacillus stearothermophilus*	P84135	450	1S6Y	EC 3.2.1.86	residues 228-239, 303-313 and 446-450 are not located in the 3D structure

the two glucosidases with the specificity assured by small structural differences (Varrot *et al.*, 2005).

Difenoconazole (DFC; 1-({2-[2-Chloro-4-(4-chlorophenoxy)phenyl]-4-methyl-1,3-dioxolan-2-yl}methyl)-1H-1,2,4-triazole) is a broad-spectrum fungicide used on a variety of fruit and vegetables crops (Thom *et al.*, 1997). When it is used, it remains in soil for a considerable period of time. The European Food Safety Authority (EFSA) assessed the toxicity of this substance as low on soil macro-organisms but no long-term data are available, and no studies regarding the toxicity on soil microbiota were made to evaluate the potential effect on soil microbial communities (EFSA report, 2011). Difenoconazole degradation by soil microbial community is important in pollution prevention because the fungicide may be washed from soil, reach the groundwater and pollute the aquatic environment. It is known that DFC is very toxic to aquatic organisms and may cause long term adverse effects in the aquatic environments (EFSA report, 2011). For these reasons it is important to obtain new and valuable information about the interactions of soil microbial communities with DFC in order to prevent the pollution of the aquatic environment and to avoid soil degradation.

The goal of this study is a comparative analysis of the structural and molecular properties of the bacterial alpha- and beta–glucosidases belonging to the GH4 family to obtain a more detailed knowledge and to improve our understanding of their specific interactions. Also, the possible interactions of the GH4 family of alpha- and beta-glucosidases with difenoconazole are investigated.

MATERIAL AND METHODS

Within this paper several bioinformatics tools are used in order to reveal sequence and structure similarities or dissimilarities of bacterial alpha- and beta-glucosidases belonging to the GH4 family. Our analysis is based on the sequences and three dimensional structures avail-

able for the considered proteins. The Protein Data Bank (PDB) entry codes of the structural files for the GH4 enzyme family from bacteria as well as the entry codes for sequences of these enzymes in the UniProt database (Leinonen *et al.*, 2006) and their Enzyme Commission numbers (EC number) are presented in Table 1 .

Sequence similarity between the GH4 enzymes is analyzed by multiple sequence alignment using the ClustalW software (Larkin *et al.*, 2007); global and local physico-chemical properties of the protein chains are retrieved using the PotParam tool (Gaisteiger *et al.*, 2005). We consider in our calculation the following properties computed using ProtParam: theoretical isoelectric point (pI), net charge, the aliphatic index and the grand average of hydropathicity (GRAVY).

The degree of dissimilarity of two three-dimensional protein structures is measured using the root-mean-square distance (RMSD) between equivalent atom pairs (Carugo & Eisenhaber, 1997). A zero value for the RMSD means identical structures and it increases for dissimilar structures. Structural similarity of the considered enzymes is compared using the structure matching tool in the Chimera software (Pettersen *et al.*, 2004) and the surface and volume of each protein are also computed.

In the case of 6-phospho-beta-glucosidase from *Thermotoga maritime* (*BglTm*), there are three structural files: 1UP4 (Varrot *et al.*, 2005) for the protein octamer in the monoclinic form, 1UP6 (Yip *et al.*, 2004) for the protein octamer in the tetragonal form in complex with manganese, NAD and glucose-6-phosphate and 1UP7 (Varrot *et al.*, 2005) for the protein octamer in the tetragonal form in complex with NAD and glucose-6-phosphate. The superimposition of the three determined crystallographic structures of *BglTm* shows RMSD values of 0.246 Å for 1UP6 compared to 1UP4, 0.252 Å for 1UP6 compared to 1UP7 and 0.275 Å for 1UP7 compared to 1UP4. As all the RMSD values are small, we consider that these three structures are highly similar and we take

into account in our further analysis the 1UP6 structural file of chain A because it corresponds to the complex of the enzyme with both the cofactors (NAD and manganese ions) and the product (glucose-6-phosphate).

As the crystallographic structure of alpha-glucosidase from *Thermotoga maritima* (*AglTm*) represents a dimer and the overall root mean square deviation between the two monomers is 2.18 Å (Lodge *et al.*, 2003), we have considered both monomers in our analysis. The biggest difference for the structures of monomers A and B is observed for region 316–355 that includes helix L (333–346) that is rotated in monomer B by 51.70° in comparison to the same helix in monomer A (Lodge *et al.*, 2003).

The structural file of the putative alpha-galactosidase from *Bacillus subtilis* (*pAglBs*), PDB entry code 3FEF, represents a homotetramer. Structure superimposition for the monomers shows RMSDs for 434 aligned Cα atom pairs between 0.197 Å and 0.312 Å. As these values are small, we consider only chain A in our further analysis. Similarly, the structural file of the putative alpha-galactosidase from *Thermotoga neapolitana* (*pAglTn*) represents a homohexamer (3U95, Leisch *et al.*, 2012) but the RSMD values for 462 aligned Cα atom pairs are small (between 0.191 Å and 0.274 Å) and we have considered only chain A in our studies.

Analysis and comparison of protein surface shapes and physicochemical properties, especially electrostatics and hydrophobicity, have provided a valuable contribution to the elucidation of protein function and molecular interactions. Within this study the surface properties of the considered enzymes are expressed in terms of: surface cavities and the global surface roughness which is quantitatively characterized by global surface fractal dimension. This quantity is defined using the fractal geometry concepts and its calculation is based on the method proposed by Lewis and Rees (1985) that considers the scaling law between the surface area (SA) and the radius of a rolling probe molecule (R) on the surface. The surface fractal dimension is determined from the slope of the double logarithmic plot of SA *versus* R. The surface area of the protein is computed using on-line free software GETAREA (Franczkiewicz & Braun, 1998; http://curie.utmb.edu/getarea.html) and probe radii of 1, 1.2, 1.4, 1.6, 1.8 and 2 Å. Also, surface properties of the investigated proteins are analyzed using the CASTp (Dundas *et al*, 2006) and *3Dsurfer* (Li *et al.*, 2008) online tools. These tools allow detection, visualization and characterization of cavities and/or protrusions present at the protein surface and thus characterization of its local geometric properties. Electrostatic properties are investigated using the PyMol software (DeLano, 2002).

Predicted functional partners for proteins may be obtained using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), an on-line free available tool (Snel *et al.*, 2011). We only considered those interacting partners that had high confidence interaction scores, i.e. higher than 0.700.

The possible interactions of the GH4 enzymes with difenoconazole were analyzed using molecular docking performed using the SwissDock server with default parameters and accurate docking option (Grosdidier *et al.*, 2011). The targets were prepared uniformly as an input for docking experiments by eliminating the ligand from the structural file (except the cations where it was the case) and using the Dock Prep tool of the CHIMERA software (Pettersen *et al.*, 2004). The three-dimensional structure of difenoconazole was generated using FROG — Free Online druG conformation generation software

(Leite *et al.*, 2007) starting from its chemical formula in SMILES (Simplified Molecular-Input Line-Entry System) format (Weininger, 1988).

RESULTS AND DISCUSSION

Sequence alignment, presented in Fig. 1, reveals similarity scores between 10 and 50 for the sequences of the considered glucosidases, as presented in Table 2. Figure 1 also presents the elements of secondary structure corresponding to *AglTm*, 1OBB chain A (Lodge *et al.*, 2003).

The similarity score, calculated for every pair of sequences that are aligned is the number of identities between the two sequences divided by the length of the alignment and represented as a percentage (Larkin *et al.*, 2007). For the alpha-glucosidases the similarity scores vary between 10.14 and 50.53 and for the beta-glucosidases the similarity score is 36.39.

The low degree of sequence identity is not reflected in the GH4 enzymes structural properties. Structure matching using the CHIMERA software reveals a high degree of similarity between the analyzed structures that is expressed in terms of RMSD values for pairs of structures and presented in Table 2.

Based on the structural and sequence similarities presented in Table 2, two groups can be distinguished: enzymes 1OBB, 1VJT and 3U95 form the first group and 1U8X, 3FEF, 1UP6 and 1S6Y form the second group. The First group comprises alpha-glucosidases from *Thermotoga maritima* and *Thermotoga neapolitana* and is characterized by high sequence identity and structure similarity. Within this group, the superposition of chain A of *AglTm* with the other enzymes usually reveals smaller RMSD values than those obtained for chain B, so in our further analysis we consider only the A chain.

Alpha-glucosidases from *Bacillus subtilis (AglBs)* show small sequence and structure identity to those found in *Thermotoga maritima* and *Thermotoga neapolitana*; they bear more resemblance to the investigated beta-glucosidases.

The global properties of the investigated proteins are quite similar, as presented in Table 3.

The analyzed proteins have high values of aliphatic indexes revealing high relative volume occupied by the amino acids with aliphatic side chains. This is in good agreement with their known increased thermostability (Conners *et al.*, 2006). Also, it is known that proteins found in thermophilic bacteria are characterized by high values of the aliphatic index (Ikai, 1980). The grand average of hydropaticity (GRAVY) indices have negative values illustrating the hydrophilic character of the investigated proteins. It is in agreement with their net charges and low theoretical isoelectric points indicating their acidic character.

The computed surface fractal dimensions are comparable and no significant differences have been observed between alpha- and beta-glucosidases.

Varrot and coworkers (2005) performed a structural comparison between the A chains of four enzymes belonging to the GH4 family: *AglTm* (PDB code 1OBB), *AgrTm* (PDB code 1VJT), *AglBs* (PDB code 1U8X) and *BglTm* (PDB code 1UP6). They found that these structures were similar, with some structural differences situated in the central region, comprising residues 220–310 of BglTm (PDB code 1UP6).

We extend the structural comparison for the A chains belonging to the other 3 enzymes considered in this study: *pAglTn* (PDB code 3U95), *pAglBs* (PDB code

```
                            eeeee hhhhhhhhhhhhhhhhhhh     eeeee  hh
sp|O33830|AGLA_THEMA    ------MPSVKIGIIGAGSAVFSLRLVSDLCKTPGLSG--STVTLMDIDE  42
sp|Q9X108|BGLT_THEMA    ---------MRIAVIGGGSSYTPELVKGLLDISEDVRIDEVIFYDIDE--  39
tr|P84135|P84135_GEOSE  --SDAXDKRLKIATIGGGSSYTPELVEGLIKRYHELPVGELWLVDIPEGK  48
tr|Q9WZL1|Q9WZL1_THEMA  ---------MKISIIGAGSVRFALQLVGDIAQTEELSREDTHIYMMDVHE  41
tr|B9KAM3|B9KAM3_THENN  ---------MKISIVGAGSVRFALQLVEDIAQTDELSREDTHIYLMDVHE  41
sp|P54716|GLVA_BACSU    ----MKKKSFSIVIAGGGSTFTPGIVLMLLDHLEEFPIRKLKLYDNDK--  44
sp|P39130|LPLD_BACSU    MFHISTLDQIKIAYIGGGSQGWARSLMSDLSIDERMSG-TVALYDLDFEA  49
                        .  *   *.  .* **                      -

                        hhhhhhhhhhhhhhhhhhh   eeeee hhhhhh   eeee      h
sp|O33830|AGLA_THEMA    ERLDAILTIAKKYVEEVGADLKFEKTMNLDDVIIDADFVINTAMVG----  88
sp|Q9X108|BGLT_THEMA    EKQKIVVDFVKR-LVKDRFKVLISDTFEG--AVVDAKYVIFQFRPG----  82
tr|P84135|P84135_GEOSE  EKLEIVGALAKRXVEKAGVPIEIHLTLDRRRALDGADFVTTQFRVG----  94
tr|Q9WZL1|Q9WZL1_THEMA  RRLNASYILARKYVEELNSPVKIVKTSSLDEAIDGADFIINTAYPYDPRY  91
tr|B9KAM3|B9KAM3_THENN  RRLNASYILARKYVEELNSPVKIVKTESLDEAIEGADFIINTAYPYDPRY  91
sp|P54716|GLVA_BACSU    ERQDRIAGACDVFIREKAPDIEFAATTDPEEAFTDVDFVMAHIRVG----  90
sp|P39130|LPLD_BACSU    AQKNEVIGNHSG-----NGRWRYEAVSTLKKALSAADIVIISILPG----  90
                        :  .                  -      -- .-- :.    -

                        hhhhhhhhhhhhhhhh                          hhhhhhhhhh
sp|O33830|AGLA_THEMA    ---GHTYLEKVRQIGEKYGYYRGIDAQEFNMVSDY-YTFSNYNQLKYFVD  134
sp|Q9X108|BGLT_THEMA    ---GLKGRENDEGIPLKYGVIG----QETTGVGGFSAALRAFP----IVEE  122
tr|P84135|P84135_GEOSE  ---GLEARAKDERIPLKYGVIG----QETNGPGGLFKGLRTIP---VILD  134
tr|Q9WZL1|Q9WZL1_THEMA  HDSGSQRWDEVTKVGEKHGYYRGIDSQELNMVSTYTYVLSSYPDMKLALE  141
tr|B9KAM3|B9KAM3_THENN  HDSGSQRWDEVTKVGEKHGYYRGIDSQELNMVSTYTYVLSSYPDVKLALE  141
sp|P54716|GLVA_BACSU    ---KYAMRALDEQIPLKYGVVG----QETCGPGGIAYGMRSIG----GVLE  130
sp|P39130|LPLD_BACSU    ---SLDDMEVDVHLPERCGIYQS--VGDTVGPGGIIRGLRAVP---IFAE  132
                                   :  :  *                   :

                        hhhhhhhhh   eeee   hhhhhhhhhhhhh   eeeee  hhhhh
sp|O33830|AGLA_THEMA    IARKIEKLSPKAWYLQAANPIFEGTTLVTRTVP-IKAVGFCHGHYG----  179
sp|Q9X108|BGLT_THEMA    YVDTVRKTS-NATIVNFTNPSGHITEFVRNYLEYEKFIGLCNVPINF---  168
tr|P84135|P84135_GEOSE  IIRDXEELCPDAWLINFTNPAGXVTEAVLRYTKQEKVVGLCNVPIGX---  181
tr|Q9WZL1|Q9WZL1_THEMA  IAEKMKKMAPKAYLMQTANPVFEITQAVRRWTG-ANIVGFCHGVAG----  186
tr|B9KAM3|B9KAM3_THENN  IAEKMKKMAPKAYLMQTANPVFEITQAVRRWTG-ANIIGFCHGVAG----  186
sp|P54716|GLVA_BACSU    ILDYMEKYSPDAWMLNYSNPAAIVAEATRRLRPNSKILNICDMPVGI---  177
sp|P39130|LPLD_BACSU    IARAIRDYAPESWVINYTNPMSVCTRVLYKVFPGIKAIGCCHEVFGTQKL  182
                        ..   .: .   :: :*  *      .       : :. *.

                        hhh hhhhh  hh   heeeeeeeeee    eeeee  eeeeehhhhhhhh
sp|O33830|AGLA_THEMA    -VMEIVEKLGLEE---EKVDWQVAGVNHGIWLNRFR-YNGGNAYPLLDKW  224
sp|Q9X108|BGLT_THEMA    -IREIAEMFSARL---EDVFLKYYGLNHLSFIEKVF-VKGEDVTEKVFEN  213
tr|P84135|P84135_GEOSE  -RXGVAKLLGVDA---DRVHIDFAGLNHXVFGLHVY-LDGVEVTEKVIDL  226
tr|Q9WZL1|Q9WZL1_THEMA  -VYEVFEKLDLDP---EEVDWQVAGVNHGIWLNRFR-YRGEDAYPLLDEW  231
tr|B9KAM3|B9KAM3_THENN  -VYEVFERLGLDP---EEVDWQVAGVNHGIWLNRFR-YRGKDAYPLLDEW  231
sp|P54716|GLVA_BACSU    -EDRMAQILGLSSR--KEMKVRYYGLNHFGWWTSIQDQGEGNDIMPKLKEH  224
sp|P39130|LPLD_BACSU    LAEMVTERLGIEVPRREDIRVNVLGINHFTWITKAS-YRHIDLLPIFREF  231
                        : :  : :.       :            *: *** .

                        hhh hhhhh          hhhhhhhhhhhhhhhhhhhhhhhhhhhh  eehhhh   hhhhhhh
sp|O33830|AGLA_THEMA    IEE-KSKDWKPENP--FNDQLSPAAIDMYRFYGVMPIGDTVRNSSWRYHR  271
sp|Q9X108|BGLT_THEMA    LK------LKLSNI--PDEDFPTWFYDSVRLIVNP-----------YLR  243
tr|P84135|P84135_GEOSE  VAHPDRSGVTXKNI--VDLGWEPDFLKGLKVLPCP-----------YHR  262
tr|Q9WZL1|Q9WZL1_THEMA  IEK-KLPEWEPKNP--WDTQMSPAAMDMYKFYGMLPIGDTVRNGSWKYHY  278
tr|B9KAM3|B9KAM3_THENN  IEK-ELSKWEPKNP--WDTQMSPAAMDMYRFYGMLPIGDTVRNGTWKYHY  278
sp|P54716|GLVA_BACSU    VSQYGYIPKTEAEA--VEASWNDTFAKARDVQAADP---------DTLPN  263
sp|P39130|LPLD_BACSU    SAHYGESGYELEGECWRDSVFCSAHRVAFDLFET-----------YGA  268
                                  :

                        hhhhhhhhhh    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh       h
sp|O33830|AGLA_THEMA    DLETKKKWYGEPWGGADSEIGWKWYQDTLGKVTEITKKVAKFIKENPSVR  321
sp|Q9X108|BGLT_THEMA    YYLMEKK-----MFKKISTHELRAREVMKIEKELFEKYR--TAVEIPEELT  287
tr|P84135|P84135_GEOSE  YYFQTDKXLAEELEAAKTKGTRAEVVQQLEKELFELYKDPNLAIKPPQLE  312
tr|Q9WZL1|Q9WZL1_THEMA  NLETKKKWFGK-FGGIDNEVERPKFHEQLRRARERLIKLAEEVQQNPGMK  327
tr|B9KAM3|B9KAM3_THENN  NLETKKKWFRR-FGGIDNEVERPKFHEHLRRARERLIKLAEEVQENPHLK  327
sp|P54716|GLVA_BACSU    TYLQYYLFPDDMVKKSNPNHTRANEVMEGREAFIFSQCDMITREQSSENS  313
sp|P39130|LPLD_BACSU    IPAAGDRHLAEFLPGPYLKQP-EVWKFHLTPISFRKQDRAEKRQETERLI  317

                        hhhh      hhhhhhhhhhhhhhhh           hhhhhhhhhhh   eee
sp|O33830|AGLA_THEMA    LSDLGSVLGKDLSEKQFVLEVEKILDPERKSGEQHIPFIDALLNDNKARF  371
sp|Q9X108|BGLT_THEMA    KRG------------------------GSMYSTAAAHLIRDLETDEGKIH  313
tr|P84135|P84135_GEOSE  KRG------------------------GAYYSDAACSLISSIYNDKRDIQ  338
tr|Q9WZL1|Q9WZL1_THEMA  LTEEH----------------PEIFPKGKLSGEQHIPFINAIANNKRVRL  361
tr|B9KAM3|B9KAM3_THENN  ITEKH----------------PEIFPKGRLSGEQHIPFINAIANNKRVRL  361
sp|P54716|GLVA_BACSU    EIKID-------------------DHASYIVDLARAIAYNTGERM  339
sp|P39130|LPLD_BACSU    VQQRG-------------------VAEKASGEEGVNIIAALGLGELVT  347
                                 -

                        eeeee         eeeeeeee   eeee        hhhhhhhhhh
sp|O33830|AGLA_THEMA    VVNIPNKGIIHGIDDDVVLVDKNGIHPEKIEPPLPDRVVKYYLRP  421
sp|Q9X108|BGLT_THEMA    IVNTRNNGSIENLPDDYVLEIPCYVRSGRVH--TLSQGKGDHFALSFIHA  361
tr|P84135|P84135_GEOSE  PVNTRNNGAIASISAESAVEVNCVITKDGPK--PIAVGDLPVAVRGLVQQ  386
tr|Q9WZL1|Q9WZL1_THEMA  FLNVENQGTLKDFPDDVMELPVWVDCCGIHREKVEPDLTHRIKIFYLWP  411
tr|B9KAM3|B9KAM3_THENN  FLNVENQGALKDFPDDLVMELPVWVDSSGIHREKVEPDLTHRIKIFYLWP  411
sp|P54716|GLVA_BACSU    LLIVENNGAIANFDPTAMVEVPCIVGSNGPE--PITVGTIPQFQKGLMEQ  387
sp|P39130|LPLD_BACSU    NVNMPNQGQVLNLPIQAIVETNAFITRNRVQ--PILSGALPKGVEMLAAR  395
                        :  . *:*  .  :.         :*

                        hhhhhhhhhhhhhhhhhhhhhhhhhhh       hhh hhhhhhhhhhhhhhh
sp|O33830|AGLA_THEMA    RIMRMEMALEAFLTGDIRIIKELLYRDPRTKSDE-QVEKVIEEILALPEN  470
sp|Q9X108|BGLT_THEMA    VKMYERLTIEAYLKRSKKLALKALLSHPLGPDVE-DAKDLLEEILEAN-R  409
tr|P84135|P84135_GEOSE  IKSFERVAAEAAVTGDYQTALVAXTINPLVPSDT-IAKQILDEXLEAH-K  434
tr|Q9WZL1|Q9WZL1_THEMA  RILRMEWNLEAYISRDRKVLEEILIRDPRTKSYE-QIVQVLDEIFNLPFN  460
tr|B9KAM3|B9KAM3_THENN  RILRMEWNLEAFISRDRKVLEEILIRDPRTKSYE-QVVKVLDEILSLPFN  460
sp|P54716|GLVA_BACSU    QVSVEKLTVEAWAEKSFQKLWQALILSKTVPNAR-VARLILEDLVEAN-K  435
sp|P39130|LPLD_BACSU    HISNQEAVADAGLTKDTGLAFQAFLNDPLVQIDRSDAEQLFNDMLQCIMQ  445
                        -   :*  .   :*    .:             :::: -  . -
```

Figure 1. Alignment of the sequences of bacterial glucosidases.
Regions with the highest structural dissimilarity are highlighted in grey (see Fig. 2–5). The elements of secondary structure corresponding to *AglTm* (O33830/1OBB chain A) are also presented with notations **h** for helix and **e** for sheet (Lodge et al., 2003).

Table 2. Sequence and structure similarity indices for the GH4 enzymes.

Sequence code entry/ PDB code entry	Sequence code entry/ PDB code entry	Sequence identity score	RMSD (Å) /number of equivalent Cα atom
O33830/1OBB_A	Q9WZL1/1VJT	50.32	0.545/468
O33830/1OBB_B	Q9WZL1/1VJT	50.32	0.885/412
O33830/1OBB_A	B9KAM3/3U95_A	50.53	0.940/406
O33830/1OBB_B	B9KAM3/3U95_A	50.53	0.969/146
O33830/1OBB_A	P54716/1U8X	18.49	1.247/145
O33830/1OBB_B	P54716/1U8X	18.49	1.223/145
O33830/1OBB_A	Q9X108/1UP6_A	18.31	1.017/148
O33830/1OBB_B	Q9X108/1UP6_A	18.31	0.969/146
O33830/1OBB_A	P84135/1S6Y	20.89	1.052/200
O33830/1OBB_B	P84135/1S6Y	20.89	1.025/204
O33830/1OBB_A	P39130/3FEF_A	17.94	1.226/208
O33830/1OBB_B	P39130/3FEF_A	17.94	1.192/216
Q9WZL1/1VJT	Q9X108/1UP6_A	17.11	1.000/162
Q9WZL1/1VJT	P84135/1S6Y	15.56	1.004/176
Q9WZL1/1VJT	P54716/1U8X	10.91	1.277/144
Q9WZL1/1VJT	P39130/3FEF_A	14.80	1.198/196
Q9WZL1/1VJT	B9KAM3/3U95_A	91.90	0.545/468
B9KAM3/3U95_A	P84135/1S6Y	16.67	1.056/191
B9KAM3/3U95_A	P39130/3FEF_A	14.80	1.238/200
P54716/1U8X	Q9X108/1UP6_A	26.99	1.153/258
P54716/1U8X	P84135/1S6Y	29.62	1.113/207
P54716/1U8X	P39130/3FEF_A	15.70	1.324/189
P54716/1U8X	B9KAM3/3U95_A	10.47	1.215/127
Q9X108/1UP6_A	P84135/1S6Y	36.39	0.970/318
Q9X108/1UP6_A	P39130/3FEF_A	22.65	1.133/215
Q9X108/1UP6_A	B9KAM3/3U95_A	16.87	1.049/166
P84135/1S6Y	P39130/3FEF_A	18.16	1.146/241

3FEF) and *BglGs* (PDB code 1S6Y). Comparative structural analysis has been performed using the PyMol software (DeLano, 2002) and reveals the following:

There are no regions with distinct structural organization between alpha-glucosidase (1OBB_A) and alpha-glucoronidase (1VJT) from *Thermotoga maritima*.

The LYS315- LYS351 region of *AglTm* (1OBB_A) has no correspondence in the structure of *BglGs* (1UP6_A) where 25 residues are not located in the structure the sequence is 30 residues shorter; for this reason we cannot analyze their structural similarity.

The ARG321–ASN347 region of *AglTm* (1OBB_A) which contains the ASNS332–ASN347 helix is distinct from the LYS327-LYS338 region of *pAglTn* (3U95_A), as it is revealed in Fig. 2.

The LYS327–LYS338 region of *pAglTn* monomer A contains a short helix LYS327–HIS332 and the rest is unstructured.

The PRO257–SER289 and LYS315–LYS351 regions of *AglTm* (1OBB_A) are distinct from the SER242–SER279 and SER309–SER322 regions, respectively, of *AglBs* (1U8X, Fig. 3A and 3B). Regions PRO257–SER289 of *AglTm* and region SER242–SER279 of *AglBs* are mainly structured in helices, but the orientations of helices are different for the two proteins. The LYS315–LYS351 region of *AglTm* is mainly structured in a long helix, ASNS332–ASN347, whereas the SER309–SER322 region of *AglBs* is unstructured.

Structural dissimilarity between alpha-glucosidase (1OBB_A) and 6-phospho-beta-glucosidase (1UP6_A) from *Thermotoga maritima*, is reflected by two regions with distinct structural organization (Fig. 4A and 4B):

Table 3. Global properties of bacterial alpha- and beta-glucosidases
AA, number of amino acids; MS, molecular surface; V, molecular volume; GRAVY, grand average of hydropathicity

| Protein accession codes UniProt/PDB	AA	MS (Å²)	V (Å³)	theoretical pI	Net charge	Aliphatic index	GRAVY	Surface fractal dimension
O33830/ 1OBB_A	480	20290	61890	5.68	−9	92.15	−0.377	2.41±0.01
Q9WZL1/ 1VJT	471	19890	62780	5.85	−9	83.80	−0.546	2.46±0.01
B9KAM3/ 3U95_A	469	19450	62350	6.12	−7	86.44	−0.552	2.33±0.01
P54716/ 1U8X	449	17550	56150	4.93	−8	84.74	−0.308	2.50±0.01
P39130/ 3FEF_A	446	17310	54770	5.47	−12	97.11	−0.063	2.34±0.01
P84135/ 1S6Y	450	17160	52610	5.66	−8	100.53	−0.147	2.41±0.01
Q9X108/ 1UP6_A	415	17640	54900	5.76	−8	98.12	−0.223	2.39±0.01
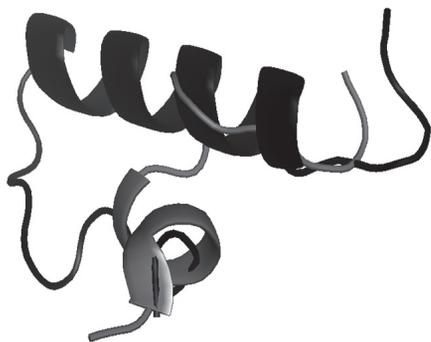
Figure 2. Regions with distinct structural organization inside of alpha-glucosidase monomer A (1OBB_A — black) form *Thermotoga maritima* (ARG321–ASN347) and putative alpha-glucosidase monomer A (3U95_A — grey) of *Thermotoga neapoletana* (LYS327–LYS338).

PRO257–TRP295 (region 1) and LIS315–LYS351 (region 2) for alpha-glucosidase (1OBB_A) compared to PRO240–LEU260 and GLU271–HIS300, respectively, for 6-phospho-beta-glucosidase (1UP6_A). All these regions are mainly structured in helices. Region 1 (PRO257–TRP295) of 1OBB_A contains a short beta sheet PRO257–ILE258 and three helices, GLY259–VAL262, SER266–GLY281 and SER289–TRP295. The PRO240–LEU260 region of 1UP6_A contains two helices, PRO240–TRP245 and GLU247–TYR257, but there is an appreciable distance between the helical segments SER266–GLY281 of 1OBB_A and GLU247–TYR257 of 1UP6_A. As already mentioned, the LYS315–LYS351 region of 1OBB_A contains two helices, ARG321–LYS325 and ASNS332–ASN347. The GLU271–HIS300 region of 1UP6_A contains the LYS286–HIS300 helix that has as a 90° rotated orientation compared to helix ASN332–ASN347 of alpha-glucosidase.

The LYS315–LYS351 region of *AglTm* is also distinct from region ARG315–GLU330 of *pAglBs* (Fig. 5). The long ASNS332–ASN347 helix of *AglTm* has no correspondent in the *pAglBs* structure (3FEF_A), its ARG315–GLU330 region being unstructured.

All presented results reveal that at least one of the regions, PRO257–SER289 and LYS315–LYS351 of *AglTm,* is structurally distinct from the other investigated glucosidases and we focus our attention on these regions and compare their properties with those of the corresponding regions in the other investigated proteins. We must also mention that the structural differences in the two regions of *AglTm* are not unexpected since these regions correspond to the insertion/deletion part of the glucosidase sequences. Moreover, these structurally distinct regions could be associated with substrate specificity.
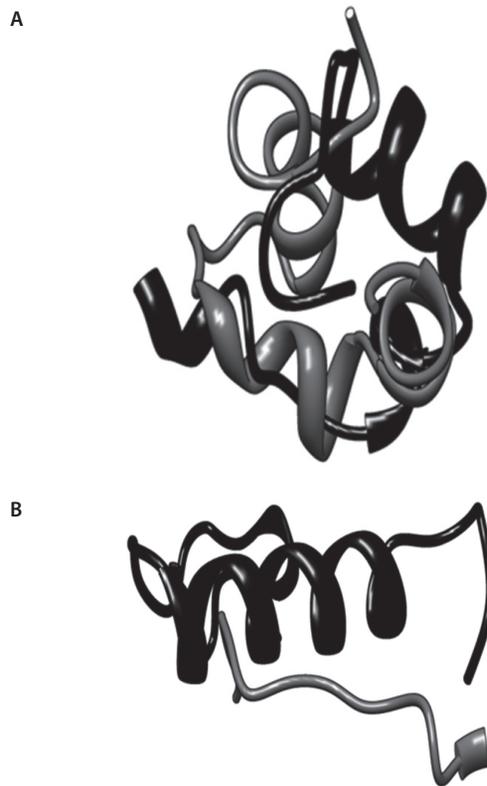

Figure 3. Regions with distinct structural organization inside of alpha-glucosidase monomer A (1OBB_A — black) form *Thermotoga maritima* and alpha-glucosidase A (1U8X — grey) from *Bacillus subtilis*
(**A**) region 1, PRO257–SER289 of 1OBB compared to SER242–SER279 of 1U8X and (**B**) region 2, LYS315–LYS351 for 1OBB compared to SER309–SER322 of 1U8X

The global physicochemical properties of these regions are presented in Table 4.

Different physicochemical properties of the regions with distinct structural organization in the alpha-glucosidases and beta-glucosidases from correlate well with the sequence alignment reflecting sequence dissimilarity. Regions 1 of *AglTm* and *AglBs* are more hydrophilic and unstable than the corresponding regions of *BglTm*. Also, region 1 of *AglTm* has a basic character such as region 1 of *BglTm*, whereas region 1 of *AglBs* has an acidic character. This observation is also sustained by the higher electrostatic potential of region 1 of *BglTm*. Within region 1, residues ASP260 and ARG263 of AglTm are implicated in interactions with the substrate, i.e. maltose (Lodge *et al.*, 2003), residue TYR265 of *AglBs*, is involved in the interaction with alpha-D-glucose-6-phosphate (Rajan *et al.*, 2004) and the corresponding residues of *BglTm* are
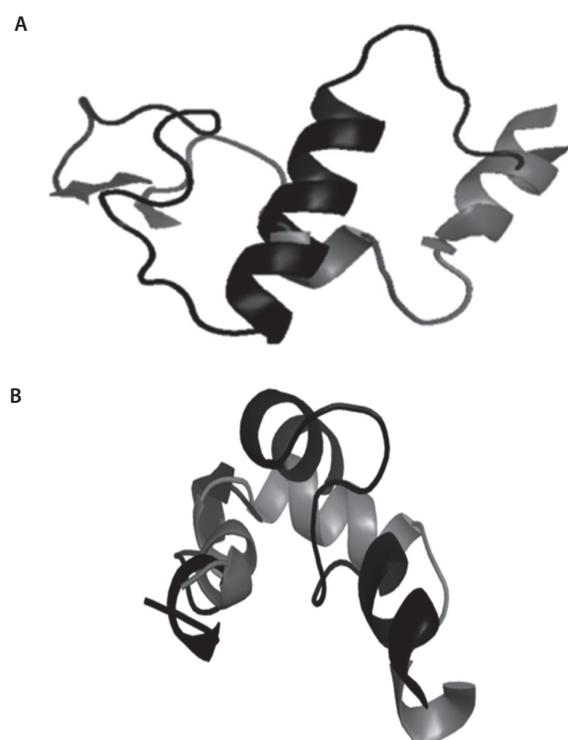
Table 4. The physicochemical properties of regions with distinct structural organization in alpha-glucosidase and 6-phospho-beta-glucosidase from bacteria.

Region/property	Region 1			Region 2				
	PRO257–SER289 of 1OBB	SER242–SER279 of 1U8X	PRO240–LEU260 of 1UP6	LYS315–LYS351 of 1OBB	LYS327–LYS338 of 3U95	ARG315–GLU330 of 3FEF	SER309–SER322 of 1U8X	GLU271–HIS300 of 1UP6
Theoretical pI	8.77	4.39	9.62	5.18	8.51	8.75	4.31	5.63
Aliphatic index	35.45	56.58	74.29	105.14	65.00	97.50	62.86	62.00
GRAVY	−1.485	−0.789	−0.705	−0.643	−1.167	−0.512	−1.214	−0.600
Electrostatic potential	−64.19 ÷ 64.19	−67.34 ÷ 67.34	−90.88 ÷ 90.88	−66.45 ÷ 66.45	−98.21 ÷ 98.21	−78.20 ÷ 78.20	−84.03 ÷ 84.03	−56.48 ÷ 56.48

A

B

Figure 4. Regions with distinct structural organization inside *AglTm* **monomer A (1OBB_A — black) and** *BglTm* **(1UP6 — grey):** (**A**) region 1, PRO257–TRP 295 for *AglTm* compared to PRO240–LEU260 for *BglTm* and (**B**) region 2, LYS315–LYS351 for *AglTm* compared to GLU 271–HIS 300 for BglTm.

not involved in interactions with either substrate or cofactors (Yip *et al*, 2004).

The identified structurally distinct regions 2 also differ in terms of thermostability, hydrophilicity and electrostatic properties. Within region 2 of 1OBB, the 313–334 fragment shows notable disorder reflected by high values for the temperature factors (Lodge *et al*., 2003). Furthermore, the 316–355 region contains the 332–347 helix which has a distinct orientation in the A and B monomers in the dimer. This region not only has differ-

ent structural properties in the two monomers of 1OBB, it is also distinct from the corresponding regions of the other investigated glucosidases. It seems that this region is implicated in the dimerization process (Lodge *et al*., 2003). Except for the GLY290 residue of *BglTm* which is involved in the interaction with alpha-D-glucose-6-phosphate (Yip *et al*., 2004), the other residues belonging to regions 2 of the investigated proteins are not involved in interactions with substrates or cofactors (Lodge *et al*., 2003; Leisch *et al*., 2012; Rajan *et al*., 2004; Yip *et al*., 2004).

Comparison of the surfaces of the investigated proteins reveals distinctive local surface features suggesting potential distinct interacting partners. The number of surface cavities differs for the investigated proteins and also the geometric properties of the first three cavities (considered the biggest) are distinct, as presented in Table 5.

STRING results for the predicted functional partners of the investigated proteins are given in Table 6. STRING predicts the functional association between proteins based on the genomic association of their genes. For the query gene, the program retrieves all the genes that occur in proximity encode functionally interacting proteins that are part of the same protein complex or are members of the same metabolic pathway (Snel *et al*., 2000).

For AglTm, AgrTm and BglTm there is only one predicted common interaction partner, beta-glucosidase. Unfortunately, there are no experimental data to confirm these predicted interactions.

Using molecular docking on the SwissDock server (Grosdidier *et al*., 2011) we have tested the possible interaction between the investigated enzymes and difenoconazole, an fungicide. All considered enzymes are predicted to interact with difenoconazole, but there are some differences especially concerning the identified structurally distinct regions of proteins. The energies of the best scored pose for every enzyme-difenoconazole interaction are presented in Table 7.

The interacting energies for the most favorable interactions with the DFC molecule are comparable for all investigated proteins. In the case of *AglTm*, *AgrTm* and *BglTm*, those structural files also contain ligands (others

Table 5. Surface properties of investigated proteins.

Protein PDB code/ Property	1OBB chain A	1VJT	3U95	1U8X	3FEF	1UP6	1S6Y
Electrostatic potential [kbT/e] (T=300°K)	−72.17÷72.17	−72.55÷72.55	−64.28÷64.28	−78.14÷78.14	−70.52÷70.52	−74.12÷74.12	−71.13÷71.13
Number of surface cavities	54	57	55	68	66	47	50
Surface of 1st cavity (Å²)	3726.7	2266.3	1505.7	2624.1	2668.3	1250.3	1620.1
Volume of 1st cavity (Å³)	6673.9	3698.5	2438.8	5655.9	6209.7	1739.9	2249.5
Surface of 2nd cavity (Å²)	193.7	955.5	388.6	564.6	168.2	496.8	623.3
Volume of 2nd cavity (Å³)	321.5	1129.5	526.8	508.6	207.2	748.1	1220.7
Surface of 3rd cavity (Å²)	273.8	384.4	374.1	409.5	151.7	536.7	614.1
Volume of 3rd cavity (Å³)	216.4	503.8	408.8	390.8	258.6	736	676.2

Table 6. Predicted functional partners of GH4 glucosidases.

Protein (gene name, UniProt accession number)	Predicted interacting partners	Gene name	UniProt accession number
alpha-glucosidase from *Thermotoga maritime* (aglA TM_1834 O33830)	beta-D-galactosidase	**TM_0310**	**Q9WYE6**
	alpha-glucosidase, putative	**TM_0752**	**Q9WZL1**
	alpha-glucosidase, putative	**TM_0434**	**Q9WYR5**
	alpha-glucosidase, putative	**TM_1068**	**Q9S5X4**
	alpha-galactosidase	**galA TM_1192**	**O33835**
	beta-galactosidase	**TM_1195**	**Q9X0S2**
	galactokinase	**galK TM_1190**	**P56838**
	beta-galactosidase	**lacZ TM_1193**	**Q56307**
	beta-fructosidase	**bfrA TM_1414**	**O33833**
	riboflavin-specific deaminase	**TM_1828**	**Q9X2E8**
alpha-glucoronidase from *Thermotoga maritima* (TM_0752, Q9WZL1)	alpha-glucosidase,	aglA TM_1834	O33830
	alpha-galactosidase	galA TM_1192	O33835
	xylosidase (beta-glucosidase)	TM_0076	Q9WXT1
	alpha-glucosidase, putative	TM_1068	Q9S5X4
	beta-glucosidase	TM_0025	Q9WXN2
putative alpha-glucosidase from *Thermotoga neapolitana* (CTN_1830, B9KAM3)	beta-galactosidase	CTN_1379	B9K9C2
	alpha-glucosidase	CTN_0761	B9K7K4
	beta-D-galactosidase	CTN_0377	B9KC07
	alpha-glucosidase	CTN_1501	B9K9P4
	galactokinase	galK CTN_1385	B9K9C8
	alpha-galactosidase	agalA CTN_1383	B9K9C6
	beta-galactosidase	bgalB CTN_1382	B9K9C5
	beta-fructosidase	CTN_1079	B9K8H2
	menaquinone biosynthesis methyl-transferase ubiE	CTN_1829	B9KAM2
	oxidoreductase	CTN_1828	B9KAM1
6-phospho-alpha-glucosidase A from *Bacillus subtilis* (glvA BSU08180, P54716)	phosphotransferase system (PTS) maltose-specific enzyme IICB component	**glvC BSU08200**	**P54715**
	transcriptional activator of the Mal operon; Positive regulator of the glv operon expression	**glvR BSU08190**	**P54717**
	putative phosphotransferase system enzyme IIA component	**ypqE BSU22230**	**P50829**
	aryl-phospho-beta-d-glucosidase	**BSU03410**	**P42403**
	maltose phosphorylase	**yvdK BSU34570**	**O06993**
	trehalose-6-phosphate hydrolase	**treA BSU07810**	**P39795**
	putative phosphotransferase system enzyme IIA component	**yyzE BSU40120**	**O32292**
	phosphotransferase system (PTS) trehalose-specific enzyme IIBC component	**treP BSU07800**	**P39794**
	putative component of transporter	**yvdJ BSU34580**	**O06992**
	beta-phosphoglucomutase; glucose--1-phosphate phosphodismutase	**yvdM BSU34550**	**O06995**

|	beta-galacturonidase	yesZ BSU07080	O31529
	putative oligo-1,6-glucosidase	yugT BSU31290	O05242
	maltose phosphorylase	yvdK BSU34570	O06993
putative alpha-glucosidase (alpha-galacturonidase) from *Bacillus subtilis* (lplD BSU07130 P39130)	putative ABC transporter (permease)	lplC BSU07120	P39129
	putative beta-hexosaminidase, putative lipoprotein	ybbD BSU01660	P40406
	putative ABC transporter (permease)	lplB BSU07110	P39128
	alpha-L-arabinofuranosidase	xsa BSU28510	P94552
	putative endo-1,4-beta-glucanase; Putative aminopeptidase	ysdC BSU28820	P94521
	glucose-6-phosphate isomerase	pgi TM1385	Q9X1A5
6-phospho-beta-glucosidase from *Thermotoga maritima* (bglT TM_1281, Q9X108)	hypothetical protein	TM_1280	Q9X107
	glucokinase	TM_1469	Q9X1I0
	beta-glucosidase	TM_0025	Q9WXN2
6-phospho-beta-glucosidase from *Geobacillus stearothermophilus* (P84135)	STRING was unable to find this protein as it has not an assigned gene name.		
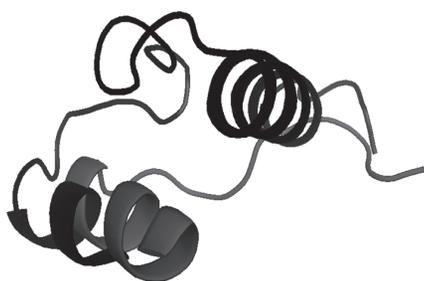
Figure 5. Regions with distinct structural organization inside of alpha-glucosidase monomer A (1OBB_A — black) form *Thermotoga maritima* (LYS315–LYS351) and putative alpha-glucosidase monomer A (3FEF_A) from *Bacillus subtilis* (ARG315–GLU330).

than cations).We can notice that DFC is able to bind to the protein in the same binding region as the ligands. It means that DFC can modulate the catalytic act of these enzymes, although, based only on the information we have collected, we cannot predict if this modulation will be an inhibition or an activation of the catalytic reaction with the natural substrate. We do not exclude the possibility that during the interaction of the enzyme with DFC a catalytic reaction is initiated.

Figure 6 illustrates the identified poses for the predicted interactions of DFC with AglTm (A) and BglTm (B) and Fig. 7 illustrates the poses for the predicted interactions of DFC with the PRO257–SER289 and LYS315–LYS351 regions of *AglTm*.

For region 1 (PRO257–TRP295) of *AglTm*, there are three identified positions for DFC binding to the protein and the corresponding interaction energies are: –2804.92 kcal/mol, –2800.82 kcal/mol and –2799.60 kcal/mol, respectively. For region 2 (LYS315–LYS351) of AglTm, there is only one position for DFC binding and the interaction energy is –2802.36 kcal/mol.

In the case of *BglTm* (1UP6), for region 1 (PO240–LEU260) there are seven identified poses for DFC binding, all of them concerning the TYR242 residue and the interaction energies between –2508.49 kcal/mol and –2504.06 kcal/mol. For region 2 (SER271–HIS300), there are 12 poses for DFC binding with the interaction energy –2511.93 kcal/mol and –2496.30 kcal/mol. Four residues: GLU271, ARG277, ARG289 and TYR294 are
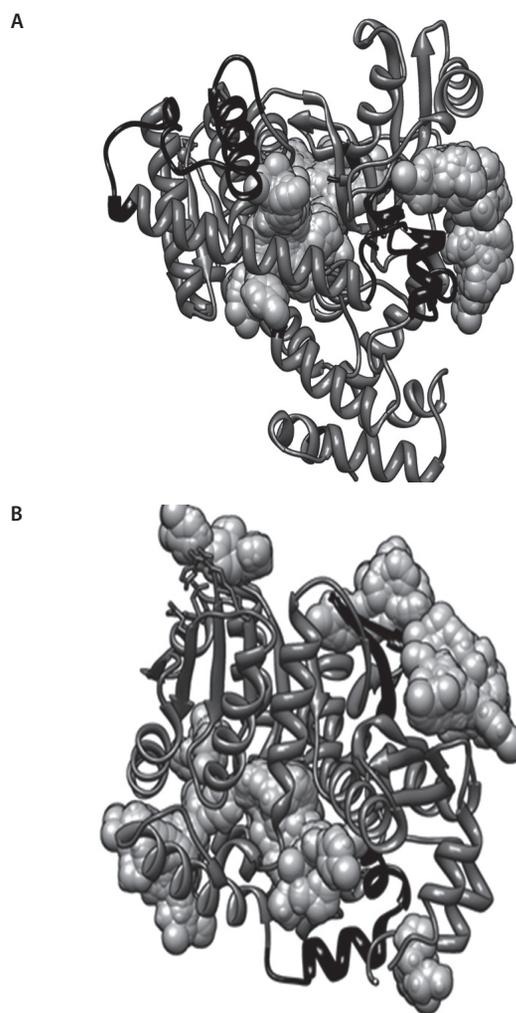

Figure 6. Illustration of the identified poses for the predicted interactions of DFC (grey spheres) with *AglTm* (A) and *BglTm* (B) respectively.
The proteins are presented as ribbon in dark grey and in black are shown the regions with distinct structural organization: PRO257–TRP295 and LYS315–LYS351 for *AglTm* (A), PRO240–LEU260 and GLU271–HIS300 for *BglTm* (B).
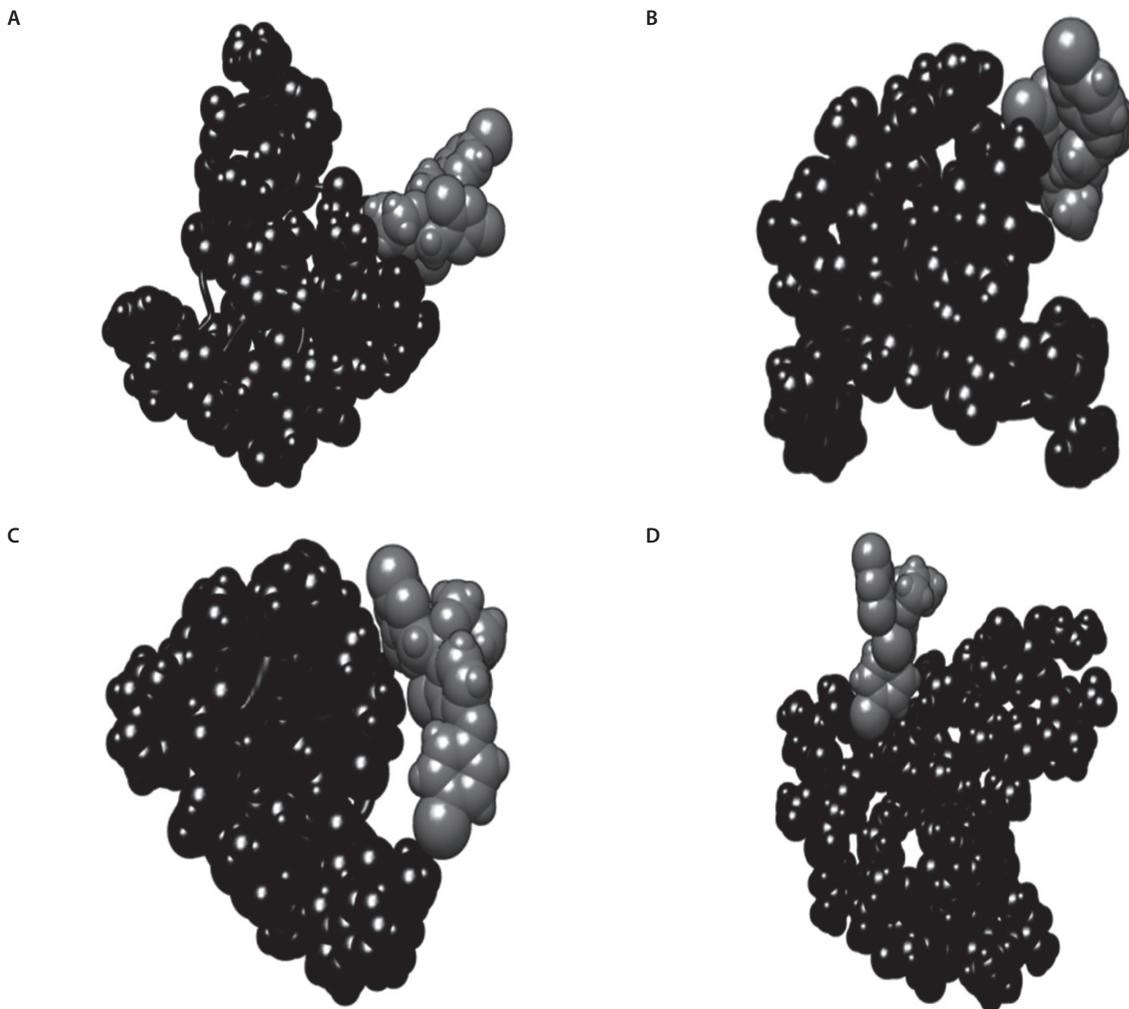
A

B

C

D

Figure 7. Illustration of the identified poses of DFC (grey spheres) located in the regions PRO257–TRP295 (A, B and C) and LYS315–LYS351 (D) of *AglTm* (black spheres).

involved in DFC binding. There are also four poses of DFC binding simultaneously to region 1 and 2 of *BglTm*.

For *pAglBs*, there are four identified poses for DFC binding to the ARG315–GLU330 region, all of them involving the GLU300 residue, with the interacting energies between –2188.35 kcal/mol and –2176.51 kcal/mol.

There are no predicted binding sites for DFC in the identified structurally distinct regions of *AglBs* (1U8X)

and *pAglTn* (3U95_A). Analysis of the predicted poses for DFC interactions with *AglBs* reveals the involvement of residues ARG160, ARD162, GLU344, VAL343 and GLU370. Also, the DFC interactions with *pAglTn* usually involve residues ARG12, TYR85, TYR87, GLU304, ARG307 and GLU311. We notice that both charged and hydrophobic residues are important for DFC binding to GH4 enzymes.

CONCLUSIONS

A major challenge in biotechnological applications is to understand the mechanism of action of carbohydrate active enzymes from thermophilic bacteria in terms of their interactions and activity. As far as we know, it is the first study that compares the structural properties of the GH4 enzyme family using structural bio-

Table 7. The energies for the best scored pose for every enzyme-difenoconazole interaction

Protein accession codes UniProt/PDB	Number of scored poses	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
O33830/ 1OBB_A	48	–2807.65	–7.90
Q9WZL1/ 1VJT	37	–2722.07	–8.49
B9KAM3/ 3U95_A	50	–2405.12	–7.87
P54716/ 1U8X	40	–2461.61	–7.86
P39130/ 3FEF_A	51	–2191.61	–7.99
P84135/ 1S6Y	47	–2329.74	–8.22
Q9X108/ 1UP6_A	41	–2511.93	–7.84

informatics approaches. Our results reveal similar global structural characteristics of these enzymes despite low sequence identity, and are in good agreement with other published data (Rajan *et al.*, 2004). This global structural similarity reflects a common reaction mechanism of these involving NAD$^+$ and divalent metal ions as cofactors. Our findings also agree with the observation that the difference in substrate specificity between alpha- and beta-glucosidases from the GH4 family is due to simple steric factors and subtle modification of the protein conformation (Varrrot *et al.*, 2005) and not to their distinct structural properties.

Our bioinformatics study identified two local regions in the investigated proteins with highly dissimilar structural organization and also with quite distinct physicochemical properties:

— region 1: PRO257–SER289 of 1OBB differs from SER242–SER279 of 1U8X and from PRO240–LEU260 of 1UP6, respectively;

— region 2: LYS315–LYS351 of 1OBB differs from LYS327–LYS338 of 3U95, ARG315–GLU330 of 3FEF, SER309–SER322 of 1U8X and GLU271–HIS300 of 1P6, respectively.

Except for *AglTm* (ASP260 and ARG 263*)* and *AglBs* (TYR265), residues belonging to region1 are situated at the interior of the proteins and are not involved in cofactor binding, catalytic activities or specific interactions with the substrates (Lodge *et al.*, 2003; Leisch *et al.*, 2012; Rajan *et al*, 2004; Varrot *et al.*, 2005; Yip *et al.*, 2004). Also, except for Gly290 of *BglTm*, residues belonging to region 2 are not involved in catalytic activities, but they are exposed to the solvent (Lodge *et al.*, 2003; Leisch *et al.*, 2012; Rajan *et al*, 2004; Varrot *et al.*, 2005; Yip *et al.*, 2004). It suggests that these structurally distinct regions may be involved in oligomerization processes or in other specific protein-protein interactions. This hypothesis is sustained by the presence of cavities with different geometric properties at the protein surfaces and by distinct predicted functional partners. Also, molecular docking studies reveal that all the investigated proteins are able to bind the fungicide, difenoconazole, but there are some differences in difenoconazole binding to the structurally distinct regions of the proteins. For the moment, we are not able to conclude if difenoconazole binding to GH4 enzymes increases or decreases the enzyme activity, or modulates it in an allosteric manner and further experimental data concerning GH4 enzyme family structures and interactions are needed to obtain a detailed knowledge on their reaction and interaction mechanisms with direct implications on their biotechnological applications.

Acknowledgement

REFERENCES

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

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**: D233–D238.

Carugo O, Eisenhaber F (1997) Probabilistic evaluation of similarity between pairs of three-dimensional protein structures utilizing temperature factors. *J Appl Crystallogr* **30**: 547–549.

Chhabra SR, Shockley KR, Ward DE, Kelly RM (2002) Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown on glucan- and mannan-based polysaccharides. *Appl Environ Microbiol* **68**: 545–554.

Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, Kelly RM (2006) Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microbiol Rev* **30**: 872–905.

Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws SA, Golyshina O, Newbold CJ, Polaina J, Ferrer M, Golyshin PN (2012) Microbial beta-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol Biofuels* **5**: 73.

DeLano WL (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos.

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* **34**:W116–W118.

European Food Safety Authority (2011) Conclusion on the peer review of the pesticide risk assessment of the active substance difenoconazole. *EFSA J* **9**:1967. [71 pp.]. doi:10.2903/j.efsa.2011.1967.

Franczkiewicz R, Braun W (1998) A new efficient algorithm for calculating solvent accessible surface areas of macromolecules. *J Comput Chem* **19**: 319–326.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*. Walker JM, ed, pp 571–607. Humana Press, Totowa, New Jersey.

Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res* **39**: W270–W277.

Ikai A (1980) Thermostability and aliphatic index of globular proteins. *J Biochem* **88**: 1895–1898.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. *Bioinformatics* **23**: 2947–2948.

Leinonen R, Nardone F, Zhu W, Apweiler R (2006) UniSave: the UniProtKB Sequence/Annotation Version database. *Bioinformatics* **22**: 1284–1285.

Leisch H, Shi R, Grosse S, Morley K, Bergeron H, Cygler M, Iwaki H, Hasegawa Y, Lau PC (2012) Cloning, Baeyer-Villiger biooxidations, and structures of the camphor pathway 2-oxo-Δ(3)-4,5,5-trimethylcyclopentenylacetyl-coenzyme A monooxygenase of *Pseudomonas putida* ATCC 17453. *Appl Environ Microbiol* **78**: 2200–2212.

Leite TB, Gomes D, Miteva MA, Chomilier J, Villoutreix BO, Tufféry P (2007) Frog: a FRee Online druG 3D conformation generator. *Nucleic Acids Res* **35**:W568–W5672.

Lewis M, Rees DC (1985) Fractal surfaces of proteins. *Science* **230**: 1163–1165.

Li B, Turuvekere S, Agrawal M, La D, Ramani K, Kihara D (2008) Characterization of local geometry of protein surfaces with the Visibility Criterion. *Proteins: Struct Funct Bioinf* **71**: 670–683.

Lodge JA, Maier T, Liebl W, Hoffmann V, Sträter N (2003) Crystal structure of *Thermotoga maritima* alpha-glucosidase AglA defines a new clan of NAD$^+$-dependent glycosidases. *J Biol Chem* **278**: 19151–19158.

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

Pei J, Pang Q, Zhao L, Fan S, Shi H (2012) *Thermoanaerobacterium thermosaccharolyticum* beta-glucosidase: a glucose-tolerant enzyme with high specific activity for cellobiose. *Biotechnol Biofuels* **5**: 31.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612.

Rajan SS, Yang X, Collart F, Yip VYL, Withers SG, Varrot A, Thompson J, Davies GJ, Anderson WF (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD$^+$/Mn^{2+}-dependent phospho-alpha-glucosidase from *Bacillus subtilis*. *Structure* **12**: 1619–1629.

Snel B, Lehmann G, Bork P, Huynen MA (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **28**: 3442–3444.

Thom E, Ottow JCG, Benckiser G (1997) Degradation of the fungicide difenoconazole in a silt loam soil as affected by pretreatment and organic amendment, *Environ Pollut* **96**:409–414.

Ugwuanyi JO (2008) Yield and protein quality of thermophilic *Bacillus spp*. biomass related to thermophilic aerobic digestion of agricul-

tural wastes for animal feed supplementation. *Bioresour Technol* **99**: 3279–3290.

Varrot A, Yip VLY, Li Y, Rajan SS, Yang X, Anderson W, Thompson J, Withers SG, Davies GJ (2005) NAD⁺ and metal-ion dependent hydrolysis by family 4 glycosidases: Structural insight into specificity for phospho-beta-D-glucosides. *J Mol Biol* **346**: 423–435.

Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Model* **28**: 31–36.

Yang B, Dai Z, Ding S-Y, Wyman CE (2011) Enzymatic hydrolysis of cellulosic biomass. *Biofuels* **2**: 421–450.

Yip VL, Varrot A, Davies GJ, Rajan SS, Yang X, Thompson J, Anderson WF, Withers SG (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from *Thermotoga maritima*. *J Am Chem Soc* **26**:8354–8355.

Yip VL, Withers SG (2006) Mechanistic analysis of the unusual redox-elimination sequence employed by *Thermotoga maritima* BglT: a 6-phospho-beta-glucosidase from glycoside hydrolase family 4. *Biochemistry* **45**: 571–580.

Yip VL, Thompson J, Withers SG (2007) Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-alpha-glucosidase from glycoside hydrolase family 4. *Biochemistry* **46**: 9840–9852.