KIBÉDI SZABÓ Csaba Zoltán[1], GÁLICZA Judit[2], WEYDA István[3], ANDRÁS Csaba Dezső[4]

[1]„Școala Normală Superioară București", Romanian Academy Institute of Biochemistry, Calea Griviței st. 21, Bucharest, Romania, email: kibedics@yahoo.com
[2]Applied Chemistry and Material Science, Politehnica University of Bucharest, Polizu st. 1-7, Bucharest, Romania
[3] Section for Sustainable Biotechnology, Department of Biotechnology, Chemistry and Environmental Engineering Aalborg University Copenhagen, Lautrupvang st. 15, Ballerup, Denmark
[4]Department of Food Science, Sapientia Hungarian University of Transylvania from Cluj-Napoca, Faculty of Science, Libertății sq. 1, Miercurea Ciuc, Romania

Structural characterization of an endoglucanase from *Serpula lacrymans* var. lacrymans S7.9

Why cellulases?

Today most of the energy consumption worldwide is dependent on fossil fuels, leading to the depletion of limited fossil fuels reserve, and also causing global climate change due to the release of their combustion products into the atmosphere [1]. Since early 2004, oil prices have risen similar to that of the first oil shock (1973-74) and double that of the second oil shock (1979-80). Consequently access to cheap energy has become essential to the functioning of modern economies [2]. Biomass energy is a promising option of renewable energy but the feedstock used for producing biomass energy should come from non-food biomass or agricultural waste [3]. Lignocellulosic biomass is the most abundant sustainable natural resource on the earth available for alternative biofuels [4]. Cellulosic ethanol as a substitute for fossil fuels has attracted the widespread interests [5]. The replacement of petroleum-based auto-motive fuels with bioethanol converted from lignocellulosic materials is anticipated to provide numerous environmental and social benefits [6] High production cost is the main obstacle of hindering the commercialization of cellulosic ethanol [5]. Conversion of plant biomass by the synergistic action of three members of the cellulase complex, to soluble sugars is the primary bottleneck associated with production of economically viable cellulosic fuels [7]. It is generally recognized that one of the problems in cellulose hydrolysis is the slowdown of enzyme activity in time and at high conversions [8]. *Nowadays the efforts to improve the activity and productivity of cellulases are at the research forefront for bioenergy technology* [4]. *Many details of the conversion processes including pretreatment and enzymatic hydrolysis are not completely understood. Without this information on enzymatic hydrolysis the bioethanol production cannot be cost-effective. For cellulosic ethanol production to be competitive with petroleum prices without subventions, further advances in these technologies are required* [9]. Fungal hydrolytic enzymes have a great potential due to the rapid development of enzyme technology and their application in bioethanol production [10]. Dry rot fungus (*Serpula lacrymans*) is known to be the most damaging destroyer of indoor wood construction materials in temperate regions. *S. lacrymans* has optimal temperature range 21-22 °C, however it can survive any temperature from 3 to 26 °C. It requires a humidity between 30-40% and it has a preference for concentrated oxigen [11]. The very recent sequencing of this fungus genome [12] has opened new possibilities. The high cellulase activity and low optimal temperature of *S. lacrymans* makes its cellulases high ranking energy efficient candidate in the biofuel competition. *Since a variety of biomass sources are envisioned for biofuel production (e.g. switchgrass, miscanthus, poplar), a broad spectrum of lignocellulolytic enzymes is required to meet future demands* [13]. *Also deeper understanding of the saccharification of cellulosic biomass could enhance the efficiency of biofuels development* [8]. ***Therefore we performed a structural characterization of the first endoglucanase from Serpula lacrymans. Although, this preliminary results are the first but unavoidable steps in the way of more complex characterization, namely the docking of enzyme with substrate molecules, it contains valuable informations.***

What about endoglucanases?

Endoglucanases (EC 3.2.1.4) initiate random attacks at multiple sites in the amorphous regions of the cellulose fiber and open up sites for subsequent attack by cellobiohydrolases [14]. The cellodextrin molecule or cellulose chain segment interacts with multiple subsites (4-7) situated in the endoglucanase specific cleft. Based on sequence and three dimensional (3D) structure, endoglucanases are grouped together with other enzymes, into ~11 glycoside hydrolase (GH) families, including GH 5, 6, 7, 8, 9, 12, 44, 45, 48, 51, and 74 [15]. As revealed by Carbohydrate Active Enzymes database (CAZy) GH 12 family has only 10 enzymes with known structures and only four of them have fungal origin. ***Thus there is certainly a need for structural characterization of other members also. For this purpose homology modeling may successfully used as it is a powerful tool in 3D structure prediction of the proteins***.

Primary structure analysis

ProtScale (Kyte and Doolittle), HMMTOP, PSORT, Philius, SignalP 4.0, Phobius, PredSI, Sigcleave, ProtParam, EMBOSS Needle, MultAlin, 3D Ligand Site, Q-Site Finder.

The 255 amino acid sequence protein indicated as putative endoglucanase by Eastwood *et. al*. [12], was subjected to hydrophobicity analysis performed by ProtParam (Figure 1). According to HMMTOP, PSORT and Philius the hydrophobic N-terminal region is not a part of the cellular membrane, but in fact is a signal sequence as revealed by SignalP 4.0, Phobius, PredSI and Sigcleave. These programs also indicated that the cleavage site is between the residues 16 and 17, the mature protein thus being 239 residues long.

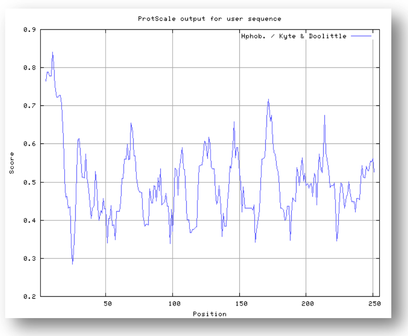

Table 1. Most Important properties of SlCel 12A.

Properties	SlCel12A
Number of amino acids	239
Molecular weight	25108.1
Theoretical pI	3.44
Nr. of negatively charged residues (D+E)	21
Nr. of positively charged residues (R+K)	3
Estimated half life	4.4 hours (mammalian reticulocytes, *in vitro*) >20 hours (yeast, *in vivo*) >10 hours (*E. coli*, in vivo)
Instability index	19.27 (this classifies the protein as stable)
Aliphatic index	74.27

Figure 1. Hydrophobicity chart of SlCel 12A.

The mature proteins most important properties were computed using ProtParam and are summarized in Table 1. The amino acid composition of the protein compared with three endoglucanases with known 3D structures is presented in Figure 2. The homology table confirms that this protein is member of the GH 12 family sharing high identity and similarity with GH 12 family members with solved (Table 2, green field) and yet unsolved (Table 2, blue field) crystal structure. Therefore we follow the annotation used for the endoglucanase from *Trichoderma reesei* (TrCel 12A) and we label the endoglucanse from *S. lacrymans* SlCel 12A.

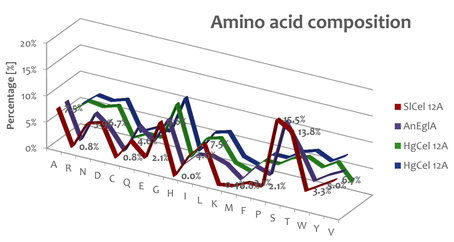

Figure 2. Amino acid composition of *S. lacrymans* SlCel 12A compered with endoglucanases with know structure (for annotation se **Table 2**).

Table 2. Homology table.

	SlCel 12A	GcCel 12A	EdCel 12A	MoEGL	AfEGL	ChbCel 12A	PpEGL	AnEglA	TrCel 12A	HgCel 12A
SlCel 12A		47.3	47.1	42.8	42.4	41.1	40.3	41.7	37.2	34.7
GcCel 12A	67.1		52.0	52.5	52.0	38.3	40.9	52.0	43.6	35.8
EdCel 12A	68.	70.5		46.9	61.7	46.4	44.2	43.2	46.9	41.6
MoEGL	62.1	68.1	65.5		43.0	34.3	36.2	43.8	46.2	39.2
AfEGL	60.9	66.9	74.2	62.4		43.6	40.4	41.0	45.1	37.6
ChbCel 12A	56.6	55.2	62.7	51.1	63.2		40.4	50.7	48.0	53.6
PpEGL	56.7	57.4	62.8	56.3	55.7	60.1		38.6	38.1	39.4
AnEglA	58.3	66.9	61.7	61.5	59.4	68.9	57.2		49.6	43.8
TrCel 12A	56.5	59.1	64.2	61.9	61.9	66.1	57.5	66.5		44.1
HgCel 12A	49.0	49.6	52.4	53.4	54.1	70.1	53.5	57.1	61.7	

SlCel 12 – glycoside hydrolase family 12 protein from *Serpula lacrymans* var. lacrymans S7.9 (EGO19898), 255 amino acid.
GcCel 12A – glycoside hydrolase family 12 endoglucanase from *Grosmannia clavigera* kw1407 (EFX01560), 253 amino acid.
EdCel 12A – glycoside hydrolase family 12 endoglucanase from *Emericella desertorum* (AAM77702), 247 amino acid.
MoEGL – endoglucanase from *Magnaporthe oryzae* 70-15 (XP_361895), 309 amino acid.
AfEGL – endoglucanase from *Aspergillus fumigatus* Af293 (XP_750222), 238 amino acid.
ChbCel 12A - glycoside hydrolase family 12 endoglucanase from *Chaetomium brasiliense* (AAM77701), 247 amino acid.
PpEGL – endo-beta-1,4-glucanase from *Postia placenta* (ADH51731), 225 amino acid.
AnEglA – chain A of endo-beta-1,4-glucanase from *Aspergillus Niger* (1KS5_A), 223 amino acid.
TrCel 12A – glycoside hydrolase family 12, endoglucanase 3 from *Trichoderma Reesei* (1OLQ_A), 218 amino acid.
HgCel 12A - glycoside hydrolase family 12, endoglucanase from *Humicola grisea* (1OLR_A), 224 amino acid.

Sequence alignment of SlCel 12A with three endoglucanases with known structure, performed with MultAlin, reveals high conservation (highly conserved residues in yellow, low conserved region in green) between sequences. As expected the binding sites (colored in red) are also highly conserved in all sequences. In predicting the residues from the binding site in the case of SlCel 12A we used 3D Ligand Site Q-Site Finder. The residues from the active site are indicated by arrows.

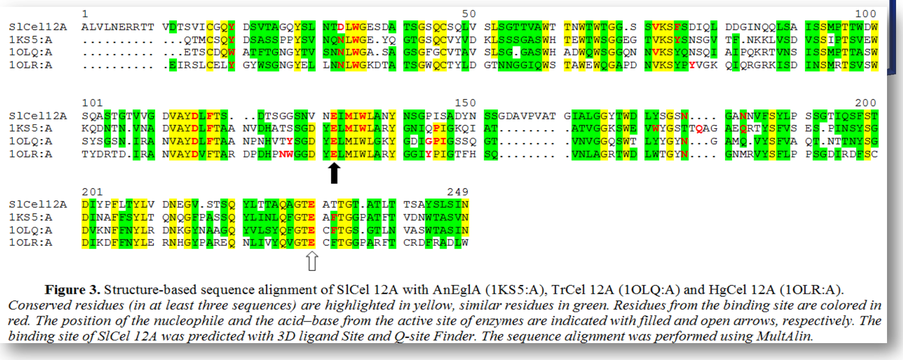

Figure 3. Structure-based sequence alignment of SlCel 12A with AnEglA (1KS5:A), TrCel 12A (1OLQ:A) and HgCel 12A (1OLR:A).
Conserved residues (in at least three sequences) are highlighted in yellow, similar residues in green. Residues from the binding site are colored in red. The position of the nucleophile and the acid–base from the active site of enzymes are indicated with filled and open arrows, respectively. The binding site of SlCel 12A was predicted with 3D ligand Site and Q-site Finder. The sequence alignment was performed using MultAlin.

What next?

We would like to perform a docking study with the obtained structure and molecular dynamics (MD) simulation for determining the enzyme stability.

Secondary structure prediction

SSPRO, SSPRO8, PORTER, PSIPRED, SOPMA, GORIV, HNNC, PREDATOR, JUFO, Disopred, ACCpro, CONpro, Dipro, Discovery Studio 3.1 Visualizer.

The consensus of secondary structure performed after the prediction by nine different methods (SSPRO, SSPRO8, PORTER, PSIPRED, SOPMA, GORIV, HNNC, PREDATOR, JUFO) is presented in Figure 4 line 2, indicating the helixes in red and beta stands is blue. The only disulfide bond in structure predicted by Dipro is presented in Figure 4 as dotted bracket. The N-terminal of the structure is highly disordered as revealed by Disopred (Figure 4, line 3).

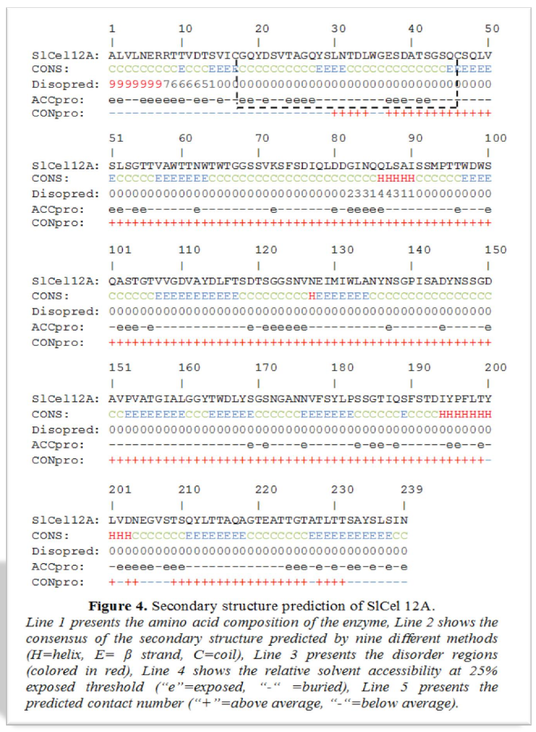

Figure 4. Secondary structure prediction of SlCel 12A.
Line 1 presents the amino acid composition of the enzyme, Line 2 shows the consensus of the secondary structure predicted by nine different methods (H=helix, E= β strand, C=coil), Line 3 presents the disorder regions (colored in red), Line 4 shows the relative solvent accessibility at 25% exposed threshold ("e"=exposed, "-" =buried), Line 5 presents the predicted contact number ("+"=above average, "-"=below average).

The solvent accessibility at 25% exposed threshold is in good agreement with the hydrophobicity chart (Figure 1) indicating that the hydrophobic regions are buried and the hydrophilic ones are mostly exposed (Figure 4, line 4). The contact number analysis realised using CONpro shows that most of the residues in the protein has a higher number of contact then the calculated average in a threshold radius of 12 Å.

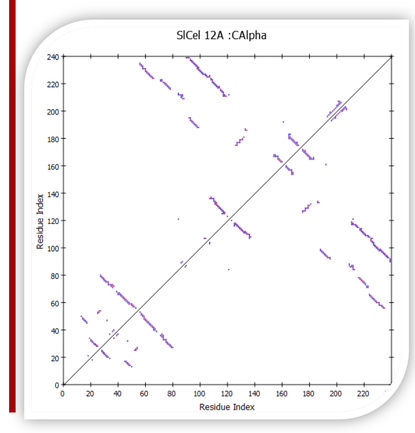

Figure 5. Contact map of SlCel 12A.

The contact map of residues from SlCel 12A realized with Discovery Studio 3.1 Visualizer is presented in Figure 5. This is a particular useful re-presentation of the protein structure. The contact provides useful information about the protein's secondary structure and it also captures non-local interactions in a threshold radius of 6 Å. The clusters of contacts represent certain secondary structures as follows: α-helices appear as bands along the main diagonal (since they involve contacts between one amino acid and its four successors) β-strands are thick bands parallel or anti-parallel to the main diagonal.

What we obtained?

We provide here valuable information about the first endoglucanase from *Serpula lacrymans.*
We realized a 3D model of the enzyme using homology modeling, indicating the residues from the active and binding site respectively.
The comparison of the model with other GH 12 endoglucanase and evaluation with different programs show a good quality of the model.

3D molecular modeling

Swift Modeller (working on Modeller 9.9), GROMOS 96, Swiss PDB Viewer, ModLoop, Discovery Studio 3.1 Visualizer, PROCHECK, PROSA.

For 3D molecular modeling of SlCel 12A eight templates (PBD codes: 1KS5, 1OLQ, 1OLR, 1H8V, 1NLR, 1OA3, 2NLR, 2JEM) where used sharing an identity between 41.7 and 32.5 %. A set of 50 models were build using Swift Modeller. Based on the Discrete Optimized Protein Energy (DOPE), molpdf and GA341 scores the best model was chosen, and further improved by loop optimization realized with ModLoop. The energy minimization was performed by GROMOS 96 implemented in Swis PDB Viewer. The proposed model for SlCel 12A was visualized using Discovery Studio 3.1 Visualizer and is presented in Figure 6.

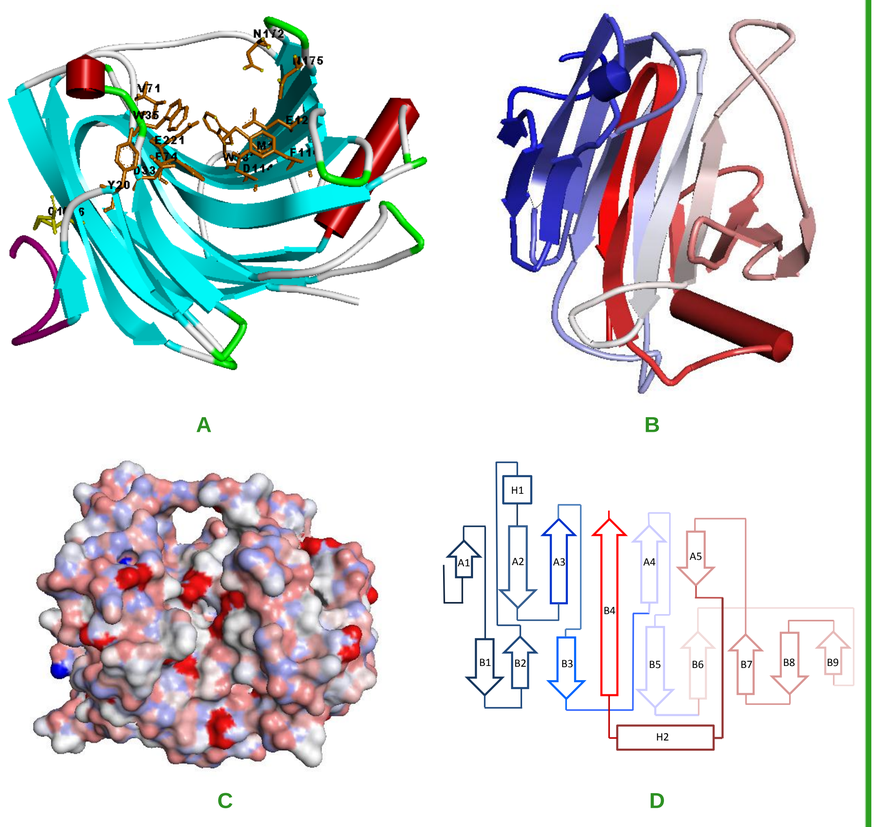

Figure 6. Theoretical model of SlCel 12A.
A- *Ribbon diagram of the model, colored by secondary structure (helicals are in red cylinder, beta strands are cyan arrows, turns are green, and coils are white), indicating the predicted amino acids from the binding and active site colored in gold, predicted disulfide bond in yellow, and the unreliable region in magenta.* ***B***- *Structure colored in a continuous gradient from blue at the N-terminus to red at the C-terminus.* ***C***- *Surface representation of the molecule (negatively charged regions are in red, positively charged ones in blue) showing the substrate binding cleft.* ***D***- *Topology diagram of the secondary-structure elements predicted in SlCel 12A.*

The fold of the protein as expected is similar to the GH 12 family members i. e β-sandwich. The structure contains 14 β-strands folded into two twisted, mainly antiparallel β-sheets (A and B, Figure 6C) that pack on top one of the other. Sheet A contains β-strands whereas sheet B contains 9. There are 5 β-hairpins, 10 β-bulges, 26 β-turns, 1 γ-turn, and also two helices as revealed by our model. The protein contains a disulfide bond (Figure 6A) highly conserved trough this endoglucanase enzyme class. The active site is situated in the crevice formed by the concave surface of sheet B (Figure 6A and C). Although at first glance it seems closed by a "cord" as described in the case of GH family 11 xylanases, the binding site is actually open.

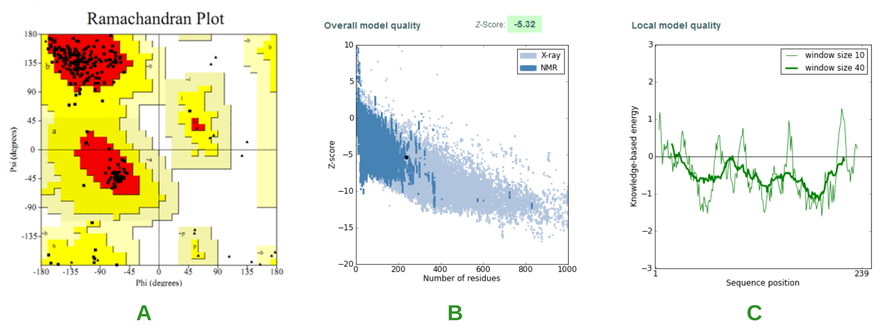

Figure 7. Predicted model validation by PROCHECK (**A**) and ProSa (**B** and **C**).

Table 3. Ramachandran statistics.

	No. of residues	% of residues
Residues in most favoured regions [A,B,L]	191	91.4
Residues in additional allowed regions [a,b,l,p]	18	8.6
Residues in generously allowed regions [~a,~b,~l,~p]	0	0
Residues in disallowed regions	0	0

The model was evaluated using different methods. The stereo chemical quality of the predicted model was evaluated using the Ramachandran plot (Figure 7A.) realized by PROCHECK, which indicated that 91.4 % of the residues phi/psi angle distribution was within core region (Table 3).
The overall quality of the model was determined by ProSa (Figure 7B and C). The program revealed that although the N-terminal region (Figure 7C) consisting of 13 amino acids is unreliable (colored in magenta in Figure 6A), the Z-score (Figure 7B) indicate a good overall quality of the model. The RMSD analysis of the developed model was evaluated by means of deviation from its template sharing the highest identity AnEglA (PDB code 1KS5), as revealed by Table 2. The superimposed Cα and backbone atoms RMSD is 0.69 Å and 0.77 Å respectively.

[1] Lo, Y.C. *et al.*: *Bioresource Technology*, 102, **(2011),** 8384-92.
[2] Hosseini, S.A. & Shah, N.: *Biomass and Bioenergy*, 35, **(2011),** 3830-40.
[3] Cheng, C.L. & Chang, J.S.: *Bioresource Technology*, 102, **(2011),** 8628-34.
[4] Ma, L. *et al.*: *Enzyme and Microbial Technology*, 49, **(2011),** 366-71.
[5] Yang, M. et al: *Biochemical Engineering Journal*, 56, **(2011),** 125-9.
[6] Saitoh, S. *et al.*: *Applied Microbiology and Biotechnology*, 91, **(2011),** 1553-9.
[7] Phillips, C.M. *et al.*: *Journal of Proteome Research*, 10, **(2011)**, 4177-85.
[8] Igarashi, K. *et al. Science*, 333, **(2011),** 1279-82.
[9] Clarke, K. *et al.*: *Biomass and Bioenergy*, 35, **(2011),** 3943-50.
[10] Bordo, E. *et al.*: *Nature Precedings*, doi:10.1038/npre.2011.6217.1.
[11] Palfreyman, J.W. & White, N.A.: *Microbiology Today*, 30, **(2003),** 107-9.
[12] Eastwood, D.C. *et al.*: *Science*, 333, **(2011),** 762-5.
[13] Allgaier, M. *et al.*: *PLoS One*, 5 **(2010),** e8812.
[14] Li, X.H. *et al.*: *Molecular Biology Report*, 38, **(2011),** 3897-902.
[15] Vlasenko, E. *et al.*: *Bioresource Technology*, 101, **(2010)**, 2405-11.