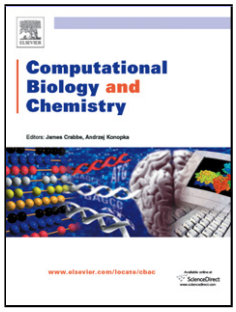


Title: Structural insights of a cellobiose dehydrogenase enzyme from the basidiomycetes fungus *Termitomyces clypeatus*

Authors: Sanchita Banerjee, Ankit Roy, M S Madhusudhan, Hridoy R Bairagya, Amit Roy

PII: S1476-9271(18)30529-2
DOI: https://doi.org/10.1016/j.compbiolchem.2019.05.013
Reference: CBAC 7077

To appear in: *Computational Biology and Chemistry*

Received date: 10 August 2018
Revised date: 18 February 2019
Accepted date: 28 May 2019

Please cite this article as: Banerjee S, Roy A, Madhusudhan MS, R Bairagya H, Roy A, Structural insights of a cellobiose dehydrogenase enzyme from the basidiomycetes fungus *Termitomyces clypeatus*, *Computational Biology and Chemistry* (2019), https://doi.org/10.1016/j.compbiolchem.2019.05.013

Original Research Article:

Structural insights of a cellobiose dehydrogenase enzyme from the basidiomycetes fungus *Termitomyces clypeatus*

Sanchita Banerjee[1], Ankit Roy[2], M S Madhusudhan[2], Hridoy R Bairagya[1#] and Amit Roy[1*]

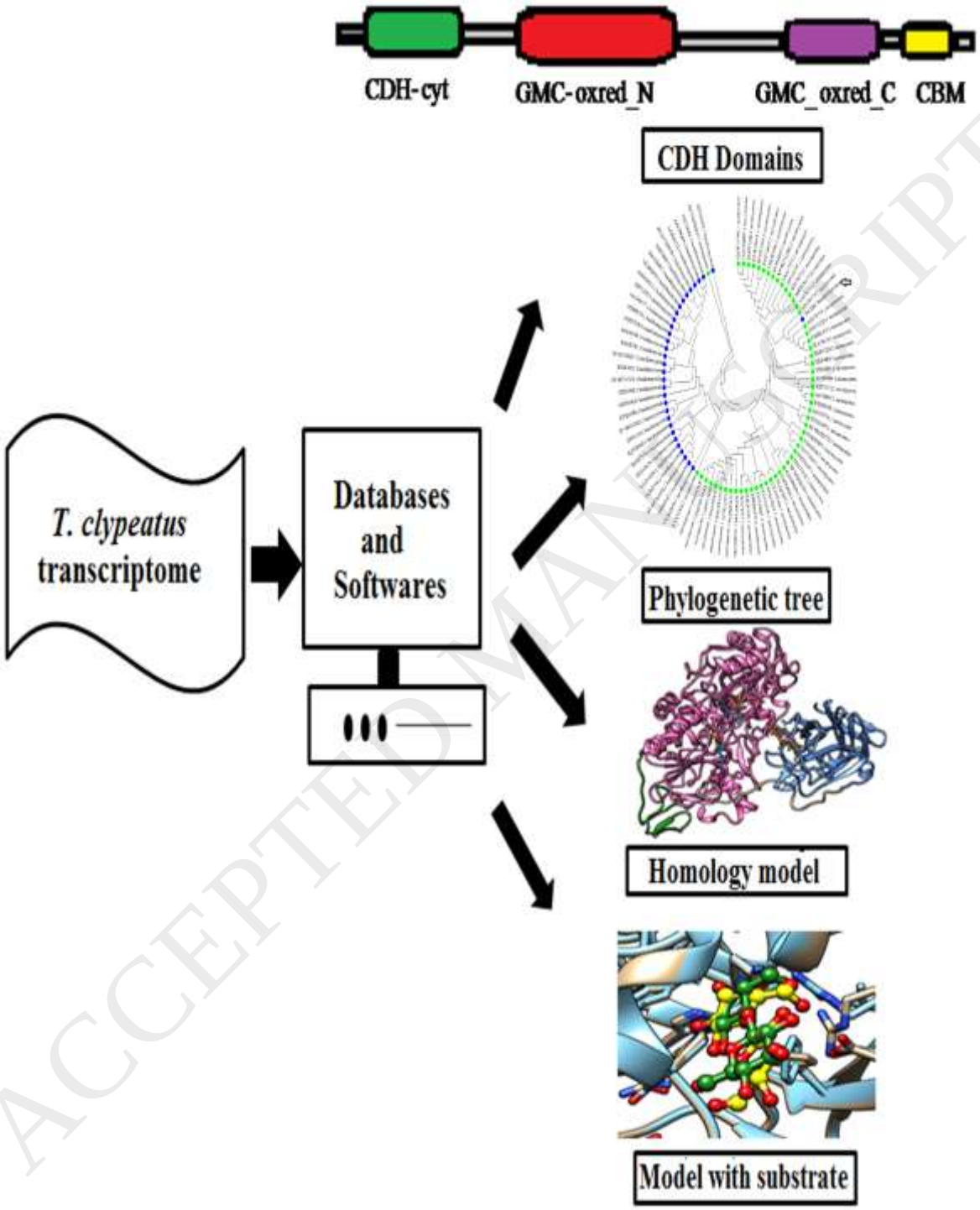
[1]Department of Biotechnology, Visva-Bharati University,
Santiniketan- 731235, West Bengal, India.

[2]Biology Wing, Indian Institute of Science Education and Research, Pune-411008, Maharashtra, India.

#Present Address: Department of Biophysics, All India Institute of Medical Science (AIIMS), New Delhi- 110029, India

***Corresponding author**: (*E-mail: amit.roy@visva-bharati.ac.in)*

Graphical abstract



Highlights:

- CDH of *Termitomyces clypeatus* possesses distinct carbohydrate binding module.
- First report of three domains- cytochrome, dehydrogenase, CBM in basidiomycete CDH.
- Dehydrogenase domain is devoid of cellulose binding residues.
- Phylogenetic analysis clusters it with the CDH of ascomycetes group of fungi.
- Possible phylogenetic status of this fungus have been discussed.

Abstract

Filamentous fungi secrete various oxidative enzymes to degrade the glycosidic bonds of polysaccharides. Cellobiose dehydrogenase (CDH) (E.C.1.1.99.18) is one of the important lignocellulose degrading enzymes produced by various filamentous fungi. It contains two stereo specific ligand binding domains, cytochrome and dehydrogenase - one for heme and the other for flavin adenine dinucleotide (FAD) respectively. The enzyme is of commercial importance for its use in amperometric biosensor, biofuel production, lactose determination in food, bioremediation etc. *Termitomyces clypeatus*, an edible fungus belonging to the basidiomycetes group, is a good producer of CDH. In this paper we have analyzed the structural properties of this enzyme from *T. clypeatus* and identified a distinct carbohydrate binding module (CBM) which is not present in most fungi belonging to the basidiomycetes group. In addition, the dehydrogenase domain of *T. clypeatus* CDH exhibited the absence of cellulose binding residues which is in contrast to the dehydrogenase domains of CDH of other basidiomycetes. Sequence analysis of cytochrome domain showed that the important

residues of this domain were conserved like in other fungal CDHs. Phylogenetic tree, constructed using basidiomycetes and ascomycetes CDH sequences, has shown that very surprisingly the CDH from *T. clypeatus*, which is classified as a basidiomycetes fungus, is clustered with the ascomycetes group. A homology model of this protein has been constructed using the CDH enzyme of ascomycetes fungus *Myricoccum thermophilum* as a template since it has been found to be the best match sequence with *T. clypeatus* CDH. We also have modelled the protein with its substrate, cellobiose, which has helped us to identify the substrate interacting residues (L354, P606, T629, R631, Y649, N732, H733 and N781) localized within its dehydrogenase domain. Our computational investigation revealed for the first time the presence of all three domains - cytochrome, dehydrogenase and CBM - in the CDH of *T. clypeatus*, a basidiomycetes fungus. In addition to discovering the unique structural attributes of this enzyme from *T. clypeatus*, our study also discusses the possible phylogenetic status of this fungus.

Keywords: Cellobiose dehydrogenase; *Termitomyces clypeatus*; carbohydrate binding module; cellulose binding residue; homology model; phylogenetic relationship.

1. Introduction

Cellobiose dehydrogenase (CDH) (E.C.1.1.99.18) is an extracellular enzyme secreted by various wood degrading fungi. It was first discovered in basidiomycetes fungi, *Trametes versicolor* [1] and *Phanerochaete chrysosporium* [2]. CDH belongs to glucose-methanol-choline (GMC) family of enzymes and is a monomeric protein [3]. The enzyme is of commercial importance for its use in biosensors and biofuel cells since it can transfer electrons from the reduced isoalloxazine ring of FAD to electrochemical cell via heme *b* [3,4]. It also has the potential for use in bioremediation by demethoxylations and hydroxylations of non phenolic and poisonous aromatic waste products [3, 4]. It is also used

in bleaching of pulp [3], production of lactobionic acids [4] etc. It consists of two domains, cytochrome (CYT) and dehydrogenase (DH), each containing a prosthetic group, heme type *b* and flavin adenine dinucleotide (FAD) respectively [5] and is connected by a flexible linker [3]. Its catalytic sequence consists of both oxidation and reduction [4]. During the oxidative half of reaction, the C1 position of saccharide is oxidized to δ-lactone and further hydrolyzed to corresponding acid in aqueous medium [6,7]. Electrons produced by carbohydrate oxidation, are transferred by $FADH_2$ of DH domain to heme in CYT domain by inter domain electron transfer [3]. CYT domain acts as mediator and electrons are re-transferred to an electron acceptor [3]. CDHs found in fungi belonging to a major group, known as ascomycetes, possess an additional third domain, a carbohydrate binding module (CBM), which is absent in most of the CDH of fungi belonging to the basidiomycetes group [4,8]. However, a CDH protein from the basidiomycete fungus, *P. chrysosporium* has been known to own a CBM connected to cytochrome domain and the DH domain is not present [9] in it. The CBM, in general, plays an important role by bringing the enzyme to close association with the substrate for increasing its rate of catalysis [10]. In spite of absence of CBM, most basidiomycetes CDHs binds cellulose with specific cellulose binding residues located within dehydrogenase domain which is not observed in any of the ascomycetes CDHs [8]. Phylogenetic analysis of CDH showed its separation into two classes - Class I and Class II produced by basidiomycetes and ascomycetes fungi respectively with the Class II type sometimes containing the CBM [8]. If the ascomycetes CDH contain the CBM, it is classified as Class IIA type while the absence of CBM classifies it as a Class IIB enzyme [7]. A third class of CDHs has been found in ascomycetes but no CDH from class III has been characterized [8].

Termitomyces clypeatus is an edible basidiomycetes fungus [11] and produces wide variety of polysaccharide hydrolysing enzymes like cellulase, sucrase, cellobiase, endo-1,4-β-D-

xylanase, 1,4-β-D-xylosidase, α-L-arabinofuranosidase, α-amylase, amyloglucosidase, including CDH [12]. It produces highest amount of CDH in cellulose medium among all the fungal species [13]. Therefore, considering its industrial importance as stated [3,4], we have studied this promising enzyme from *T. clypeatus* structurally using computational techniques for discovering additional unique properties that may increase its potential further.

We have analyzed the sequence of *T. clypeatus* CDH (*Tc*CDH) in detail and found a distinct CBM along with dehydrogenase and cytochrome domains. Thorough study of the domains has exhibited number of features similar to that of ascomycetes class II CDH. The second part of our study deals with the examination of its 3D structure by homology modelling [14] followed by validation of the modelled structure. Modelling of *Tc*CDH with its specific ligands, i.e. heme of CYT domain, flavin adenine dinucleotide (FAD) of DH domain and substrate has also been carried out which helped us to identify the residues responsible for ligand interacting. This structural characterization showed the similarities between the *Tc*CDH and the CDHs of other basiomycetes and ascomycetes fungi signifying their evolutionary relationship.

2. Methods and materials

2.1 Analysis of *Tc*CDH sequence and phylogenetic tree

CDH amino acid sequence of *T. clypeatus* with accession no: KX576458 and GenBank id: AOW70087.1 [15] was obtained from NCBI database. The protein sequence was analyzed by Basic local alignment search tool for protein (BLASTP) [16] to look for different CDH domains. Pfam (a database of protein families) [17] analysis and multiple sequence alignment using *Tc*CDH sequence was performed to study the domain organization and conservation of residues respectively.

Phylogenetic analysis was performed using 87 CDH protein sequences (32 from basidiomycetes, 55 from ascomycetes) which were collected from NCBI database. The phylogenetic tree was constructed using the multiple sequence alignment program MUSCLE [18] and then maximum likelihood [19] method included in the computer program MEGA7.0.21 (Molecular Evolutionary Genetics Analysis) [20].

2.2 Homology modelling of *Tc*CDH

Protein BLAST [16] was done with *Tc*CDH amino acid sequence [15] to identify homologues of the enzyme along with their known conserved domains. The CDH sequences from ascomycetes and basidiomycetes groups of fungi that were homologous to the *Tc*CDH sequence were considered and aligned to identify the conserved domains from these organisms.

Homology modelling of *Tc*CDH was performed based on the crystal structure of homologous CDH from *Myriococcum thermophilum* [7] as template (PDB id 4QI6), using the modelling tool, MODELLER9.17 [14]. The ligands, heme and flavin adenine dinucleotide (FAD), were modelled and placed in specific geometrical orientation at cytochrome and dehydrogenase domains respectively using the same modelling program.

2.3 Modelling the substrate at active site pocket

The *Tc*CDH substrate, cellobiose was modelled in two steps; a) transforming the coordinates of cellobiose analog (ABL) from *Tc*CDH homologue (PDB id: 1NAA [21]) by structural superimposition and b) replacing the cellobiose analog with cellobiose (CBI) by structural superimposition of ligands. Structural alignment of 1NAA and *Tc*CDH was performed using CLICK [22] with CA as representative atoms. Transformed coordinates of the cellobiose analog, ABL (5-amino-5-deoxy-cellobiono-1,5-lactam), was added to *Tc*CDH. CBI coordinates were obtained from an unrelated protein structure (PDB id: 5CVY [23]) and

superimposed to ABL using CLICK [22] with all carbon and oxygen atoms as representatives. Transformed CBI coordinates were used to replace ABL.

The modelled protein (along with ligands) was validated next by model assessment tools such as, MODELLER's Discrete Optimized Potential Energy (DOPE) method [24], ProSA [25] VERIFY3D [26,27] and RAMPAGE [28]. DOPE is an atomic distance dependent statistical potential calculated from a sample of native protein structures [24]. ProSA is a statistical potential that assesses the modelled protein using energy profile and compares it with the potential mean force from a set of known protein structures [25]. Verify3D program from Protein Structure Validation Software suit (PSVS) 1.5 [29] was used for model validation. This program analyzes the reliability of the modelled protein and characterizes the position of each amino acid with an environmental score by constructing a 3D profile [26,27]. Overall, stereochemistry of the model was analysed by RAMPAGE [28]. Further, CLICK [22] was used to obtain root mean square deviation (RMSD) for CA atoms by superimposing modelled *Tc*CDH and other PDB structures. The PDB structures used for validation study of modelled protein in terms of Root Mean Square Deviation (RMSD) and structure overlap percentage are listed in Table S3.

3. Results

3.1 Study of *Tc*CDH domains and phylogenetic analysis

BLASTP [16] study using *Tc*CDH sequence (data not shown) reveals the presence of a carbohydrate binding module (CBM) in its structure, in addition to the cytochrome and dehydrogenase domains; the latter two domains are known to be present customarily in most CDH enzymes. The CBM domain, when present in the CDH enzyme, enhances the rate of cellulose hydrolysis by increasing the enzyme concentration on the substrate surface [10]. It is to be clearly noted here that the CBM (Carbohydrate Binding Module) is known to be

present in enzymes other than CDH also, that are involved in various types of carbohydrate degradation. Where present within a carbohydrate degrading enzyme, a CBM involves amino acid sequence with discrete fold having carbohydrate binding activity [10]. It consists of 30-200 amino acids, may be located either in C or N terminal ends or occasionally centrally positioned within the polypeptide chain [10]. Based on amino acid sequence similarity, the CBMs from various carbohydrate degrading enzymes have been classified into at least 71 different families of which CBMs that are found in the CDHs are positioned within Family 1 [7, 30]. The 3D structures of 71 representative families of CBM have been interpreted so far (http://www.cazy.org) [30]. Our Pfam analysis reveals that residue numbers 808-836 (amino acid number with signal sequence not removed from N-terminal end of pre-protein) consists of CBM domain of *Tc*CDH (Figure 1(a), upper panel). Alignment of 758 CBM motifs from various carbohydrate active enzymes archived in Pfam database (Table S1) with *Tc*CDH CBM has been done using ClustalX2.1 [31] program and subsequently analysed in Jalview2.10.3b1 for multiple sequence alignment editing and visualization [32]. The results indicate that the key residues responsible for cellulose binding i.e. three aromatic amino acids (Y809, Y835 and Y836, brown colour in figure 1(a) lower panel) and three cysteine residues (C812, C823, C829, blue colour in figure 1(a) lower panel) are in the expected conserved positions (partial alignment data shown in Figure 1(a), lower panel) in the CBM of *Tc*CDH.

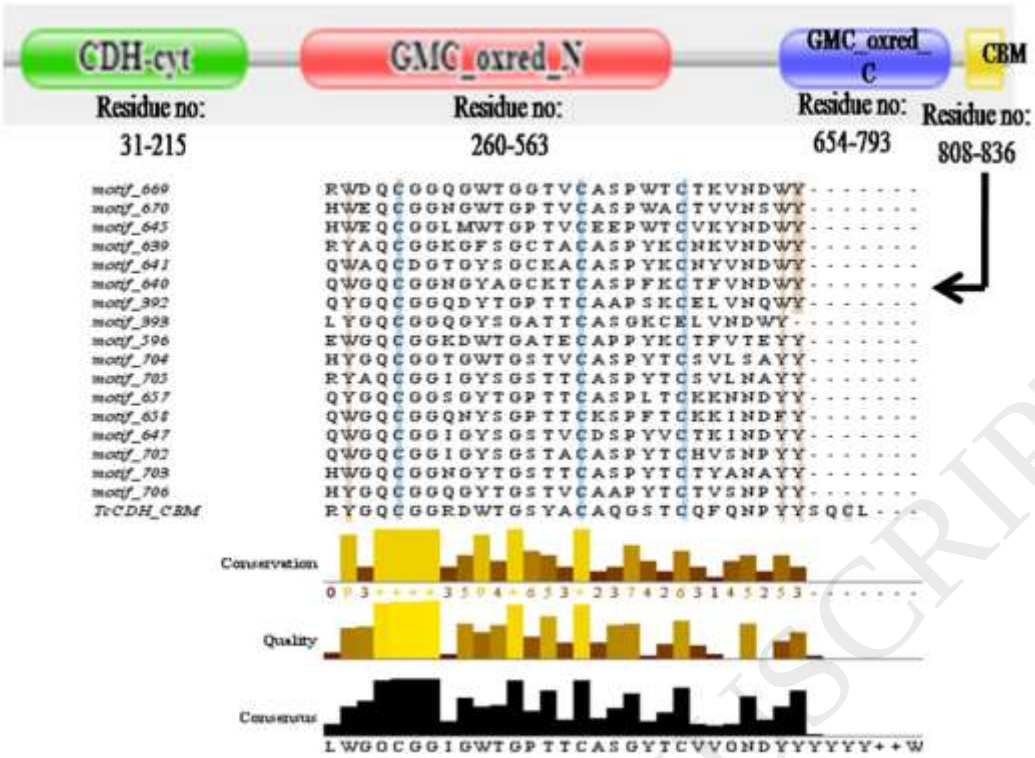


Figure 1(a). Domain organization of *Tc*CDH (upper panel) and CBM residues alignment (lower panel). Upper panel shows *Tc*CDH domain organization as exhibited by Pfam. Lower panel shows the Jalview2.10.3b1 image of alignment of the CBM residues of *Tc*CDH (TcCDH_CBM) with some other representative CBM motifs (of the 758 sequences analysed; data no shown) archived in Pfam. Three cysteine (C812, C823, C829 in blue colour) and aromatic amino acid residues (Y809, Y835 and Y836, brown colour) are well conserved.

These central residues of CBM which contribute to the binding on surface of cellulose are already reported for the Class IIA CBMs of ascomycetes CDH enzymes [8] and a class I CDH variant from *P. chrysosporium* [9]. Additionally, the Pfam study also indicates that the CBM of *Tc*CDH belongs to CBM 1 family by virtue of its possessing four cysteine residues (C812, C823, C829, C839 - blue colour in figure 1(a) lower panel) which are involved in disulphide bond formation. To confirm the position of CBM within *Tc*CDH, BLASTP was done again using CDH sequence of *T. clypeatus* with other CBM motifs (data not shown).

These results showed that the C-terminal end of the *Tc*CDH sequence contain the CBM, like other CBM containing CDHs.

BLASTP analysis reveals that the dehydrogenase domain belongs to Glucose-Methanol-Choline (GMC) family of enzymes similar to other fungal CDHs [3]. Pfam analysis shows that residue number 260-563 consists of N-terminal part of this domain and its C-terminal part consists of residue number 654-793 (amino acid number with signal sequence not removed from N-terminal end of pre-protein) (Figure 1(a), upper panel). Most of the class I CDHs, which do not possesses a CBM, bind to cellulose by specific cellulose-binding residues present within dehydrogenase domain [8]. This cellulose binding residues are attributed to nine aromatic amino acids in class I CDHs which is not present in any of the class II CDHs [8]. When compared by aligning dehydrogenase domain sequence of *T. clypeatus* CDH with class I and class II CDHs (13 sequences from class I and 15 sequences from class II) using Clustal Omega1.2.4 [33] program, we observe the absence of those aromatic residues in *Tc*CDH like class II CDHs (Figure 1(b)).

The cytochrome domain of *Tc*CDH consists of residue number 31-215 (Pfam analysis of *Tc*CDH with signal sequence not removed from N-terminal end of pre-protein, Figure 1(a) upper panel). The cytochrome domain sequence of *Tc*CDH was aligned with other 158 cytochrome CDH motifs documented in Pfam database (Table S2) using ClustalX2.1 [31] program followed by analysis in Jalview2.10.3b1 [32] program. From the alignment result (partial alignment shown in figure S2) the significant residues of this domain could be identified. Methionine and histidine residues (M92, Figure S2 (a); H194, Figure S2 (b) of *Tc*CDH) are conserved in all cytochrome domain motifs archived in Pfam and are predicted to be the axial ligands of heme iron as discussed for the basidiomycete, *P. chrysosporium* CDH (M65, H163 [5,34]). A pair of conserved cysteine residues is present in the cytochrome

domain (C147 and C150 of *Tc*CDH, Figure S2(b)) that form a disulphide bond and are conserved in other cytochrome CDH motifs as well.

Figure 1(b). Alignment of cellulose binding sequences of dehydrogenase domain in basidiomycetes (class I) and ascomycetes (class II) CDHs and *Tc*CDH using Clustal Omega1.2.4. Cellulose binding aromatic residues of class I basidiomycetes CDHs are marked in sky blue and the corresponding residues of class II ascomycetes CDHs in olive green. *Tc*CDH (AOW70087.1, marked in black box) is devoid of cellulose binding aromatic residues.

In order to investigate the similarity of CDH of *T. clypeatus* with those of other fungal CDHs, phylogenetic tree was constructed with 32 and 55 amino acid sequences from basidiomycetes and ascomycetes CDHs respectively using MEGA7.0.21 [20].

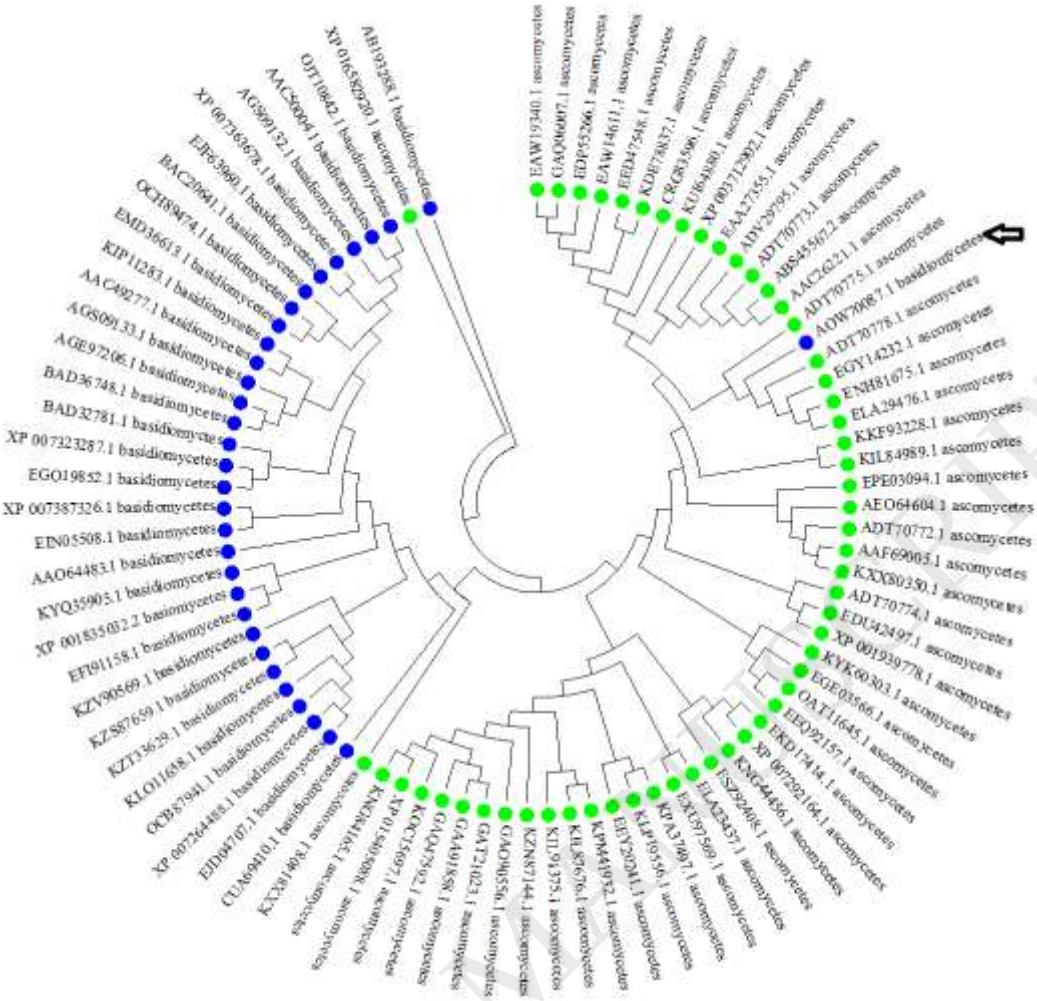


Figure 2. Phylogenetic tree was created using MEGA7.0.21 program with 55 ascomycetes and 32 basidiomycetes CDH sequences along with *Tc*CDH (AOW70087.1) sequence. Green and blue colour codes indicate ascomycetes and basidiomycetes group members respectively. *Tc*CDH (indicated by arrow) is the only basidiomycete to possess the all the three domains - cytochrome, dehydrogenase, carbohydrate-binding module and its dehydrogenase domain is devoid of the cellulose binding residues like ascomycetes CDHs; hence, it is clustered with ascomycetes group members.

Phylogenetic tree revealed that the enzyme protein is divided into two major groups for ascomycetes and basidiomycetes (Figure 2). *Tc*CDH is the only basidiomycete to possess all the three domains – it possesses both the cytochrome and dehydrogenase domains found in the ascomycetes and basidiomycetes in addition to the CBM domain found in the class IIA

CDHs of ascomycetes. In addition, its dehydrogenase domain has been found to be devoid of the cellulose binding residues as found in case of the ascomycetes; this is in contrast, on the other hand, with the dehydrogenase domain found in basidiomycetes where this domain contains specific cellulose binding residues discussed earlier. From these findings it is clear that the *Tc*CDH is more similar to the ascomycetes CDH than that of the basidiomycetes CDH (Figure 2) in spite of itself being classified as belonging to the basidiomycetes group.

3.2 Modelling the 3D structure of *Tc*CDH

In spite of its high industrial relevance, three dimensional structure of the CDH enzyme from *T. clypeatus* is not available in the Protein Data Bank. Our findings of three distinct domains (cytochrome, dehydrogenase and CBM) in the CDH of this fungus, like class IIA CDHs, merits further investigation. Homology modelling of CDH from *T. clypeatus* has been carried out to partially fill this lacuna. The CDH from *M. thermophilum*, an ascomycete, has been selected as the template for this homology modelling study because it has been found to be the best hit sequence by the BLASTP program using the *Tc*CDH as query sequence and has 63% sequence identity to the query sequence, 96% query coverage and E-value 0.0. The first 24 residues in the N-terminal part of the protein sequence did not align well to the template (Figure S1) and were predicted to be part of the signal sequence. Hence, this part of the sequence was not considered for this computational study.

Homology model using MODELLER9.17 produces the 3D structure of the protein (Figure 3). Analysis and visualization of protein model has been done in UCSF Chimera1.11.2 [35]. It displays the three distinct domains- cytochrome, dehydrogenase and carbohydrate binding module of *Tc*CDH.

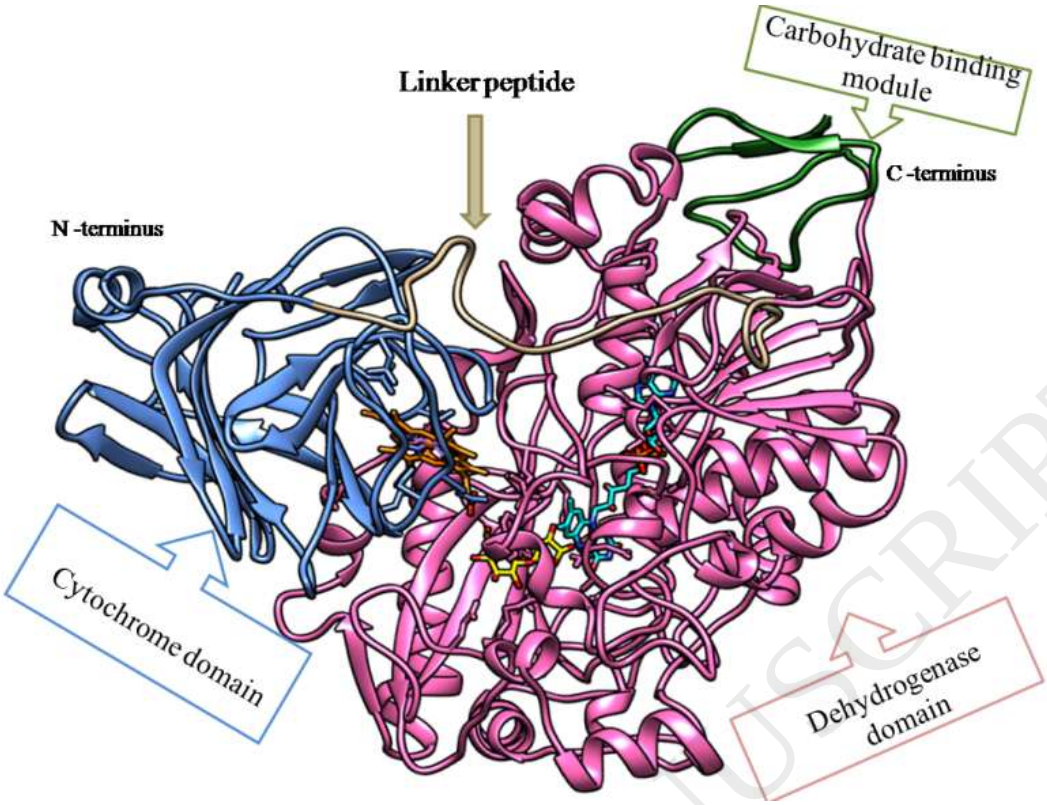


Figure 3. Model structure of *Tc*CDH. Cytochrome domain in blue colour, dehydrogenase domain in pink, carbohydrate binding module in green, linker in tan, heme in orange, FAD in cyan and cellobiose in yellow colour. The model was visualized and analyzed in Chimera1.11.2.

The cytochrome domain of *Tc*CDH is located in the N-terminus region and folds like a β-sandwich similar to Fab VH domain of antibody [4]. The dehydrogenase domain is larger, and belongs to Glucose-Methanol-Choline (GMC) family of enzymes similar to other fungal CDH [3]. These two domains are connected with linker peptide. The CBM domain is located at the C-terminal end of the protein. Pfam displayed four cysteine residues which are involved in disulfide bond formation and three of them are well conserved among other CBM motifs (Figure 1(a) lower panel). In addition, three aromatic amino acid residues attributed for cellulose binding are well conserved in the same CBM motifs (Figure 1(a) lower panel).

Modelling with ligands has been done using the same modelling tool. The cytochrome domain binds heme iron at the exterior of the protein. The dehydrogenase domain, like other

fungal CDH [4] binds FAD. The modelled protein is presented in Figure 3 and a PDB file (TcCDH_Homology model.pdb) for this model has been included in the supplementary section.

Residues that interact with the heme, FAD and cellobiose, through H-bonding (Figure S3(a), S3(b) and 4(a) respectively) and those that are within 5Å of the respective ligands are analysed from Chimera [35] and listed in Table 1.

Ligand	**Interacting residues**	
	<5Å	**H-bonding**
Heme	W83, G85, V86, S87, G90, P91, M92, T93, L97, L98, M99, L112, Y117, D118, Q119, P120, G172, W173, A174, M191, K192, Q193, H194, Q197, G198, I199, W325, S328, M339, S727, R730	M92, H194
Flavin adenine dinucleotide (FAD)	V265, G266, A267, G268, A269, G270, G271, I288, E289, K290, G291, I324, W325, C333, M339, A340, G341, C342, V343, L344, G345, G346, G347, T348, I350, N351, A352, G353, L354, T468, S469, V470, S509, A510, G511, T512, G514, L518, N732, H733, W734, D770, A771, G772, N781, P782, S783, A784, I786	G270, C342, V343, G347, L354, V470, A771, S783

Cellobiose (CBI)	N322, W325, A352, L354, A605, P606, T629, R631, E633, Y649, R730, A731, N732, H733, N781	E633, Y649, A731, N732, H733

Table 1. Interaction of *Tc*CDH residues with ligands (heme, FAD and CBI).

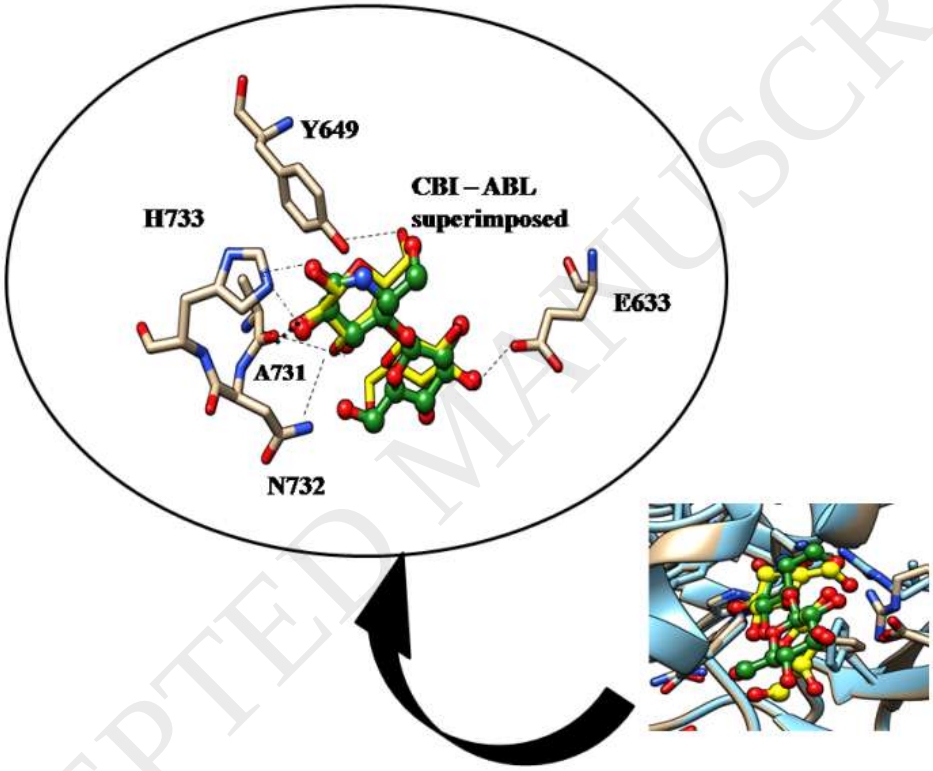


Figure 4(a.) H-bonding interaction (black dashed lines) of cellobiose in *Tc*CDH. Cellobiose (CBI) superimposed on analog (ABL) interacts with E633, Y649, A731, N732, H733. Colour codes: red - oxygen, blue - nitrogen, green - ABL, yellow - CBI, tan and sky blue - residues. The results were rendered in Chimera1.11.2.

Conservation of residues interacting with these ligands (heme and FAD) using respective crystal structures are presented in S4(a) and S4(b) respectively. M92 and H194 (pink colour in Figure S4(a)) of *Tc*CDH cytochrome domain are determined to be the axial ligands of

heme iron [4] on the basis of H-bonding interaction study and are within 5Å distance of heme. These residues are also conserved among the mentioned crystal structures of heme (1D7B, 4QI3, 4QI7). Interaction study of the *Tc*CDH residues with substrate, cellobiose (CBI) have been done by comparing the modelled structure with three relevant crystal structures (at resolutions 3.2Å or better) available in the PDB database – these are 1NAA from *Phanerochaete chyrosporium* (basidiomycetes), 4QI6 from *Myriococcum thermophilum* (ascomycetes) and 4QI7 from *Neurospora crassa* (ascomycetes).

Figure 4. (b) Clustal Omega1.2.4 program showing the conservation of amino acid residues which interacts with the cellobiose. L354, P606, T629, R631, Y649, N732, H733 and N781 are within 5Å of the substrate binding site (blue, pink). Y649, N732 and H733 interact with substrate by H-bonding (pink).

The substrate interacting residues reside within the dehydrogenase domain like other CDHs. It is found that Y649, N732 and H733 are well conserved in the PDB structures 1NAA, 4QI6 and 4QI7 (pink colour in Figure 4(b)) that interacts with CBI through H-bonding.

Furthermore, L354, P606, T629, R631, Y649, N732, H733 and N781 are well conserved among these same PDB structures (blue and pink colour in Figure 4(b)) that are within 5Å from substrate. Therefore, L354, P606, T629, R631, Y649, N732, H733 and N781 are predicted to be the residues which interact with the substrate. Multiple sequence alignment of the protein sequences has been done using Clustal Omega1.2.4 [33].

Different servers of validation mentioned in materials and methods section have been used to check the overall model quality for further application. MODELLER's DOPE calculates the DOPE energy of the model [24] which is displayed in Figure S5. Despite a region with a high energy (around residue 200) all other regions show good DOPE energy when compared to the template. ProSAII energy plot was also used to evaluate the overall accuracy of the protein structure. The ProSAII profile confirms that the residues are all in energetically well-optimised environments (negative energy values; Figure S6), with very few exceptions (positive energy values) [25]. Verify3D program provides a score to each residue for determining their packing quality and majority of residues should have 3D-1D score above 0.2 [27]. The overall score of the model is 0.48 which is in the acceptable range according to Verify3D rules as shown in Figure S7. Ramachandran plot using RAMPAGE indicate that 96.2% residues are in the favoured region, 3.2% residues are in allowed regions, 0.6% residues are in the outlier region (Figure S8) [28]. Furthermore, the cytochrome and dehydrogenase domains of *Tc*CDH are superimposed on respective domains of *M. thermophilum* (4QI3 and 4QI4). The RMSD value (CA atoms) between those superimposed complex structures are 0.65Å and 0.62Å for cytochrome and dehydrogenase domain respectively (Figure S9). The value of RMSD less than 1 indicates that the model protein is reasonably similar to the template and of good quality.

4. Discussion

Traditionally, organisms were classified in to different groups based on their morphological properties. But with the advent of Molecular Biology, plentiful molecular features of various biomolecules became available in the form of DNA and protein sequences. This helped researchers to categorize organisms more precisely on the basis of their molecular properties. The fungus *T. clypeatus* was classified into basidiomycetes group according to its morphological characteristics [11] way back in 1951 at which time all these molecular data about various organisms were virtually non-existant. Our own analysis of *T. clypeatus* mycellial cells under microscope reveals the presence of basidium (data not shown), which is the main feature of all basidiomycetes fungi and supports this classification from the morphological viewpoints. But investigations of CDH enzyme sequence and structure from *T. clypeatus* using computational tools reveal a number of similarities with ascomycetes CDHs, such as, i) CDH from this source *T. clypeatus* possesses a discrete carbohydrate binding module (CBM) (Figure 1) which is mostly found in ascomycetes class IIA CDHs and is generally absent in basidiomycetes CDHs. This inference has been drawn from a characterization of CBM domain of various groups described in the Pfam database which shows that the important residues for carbohydrate binding - three aromatic amino acid residues and three cysteine residues - are well conserved in the carbohydrate binding motifs (CBM) of 758 carbohydrate active proteins including the ascomycetes IIA CDHs (Figure 1(a), lower panel and Table S1). *Tc*CDH also contains these residues and this confirms the CBM nature of this particular domain in *Tc*CDH. CDH of all other basidiomycetes (except *P. chrysosporium* [9]) and Class IIB CDH from ascomycetes, are devoid of this particular CBM domain. ii) dehydrogenase domain of *Tc*CDH do not contain cellulose binding residues; similarly class IIA and IIB ascomycetes CDHs (Figure 1(b)) are also devoid of these residues within their dehydrogenase domains. Basidiomycetes CDHs (except *P. chrysosporium* [9]) are also known to have the dehydrogenase domain which, on the other

hand, contains these conserved cellulose binding residues [8]. This is yet another major similarity of the *Tc*CDH with the CDHs belonging to the ascomycetes group. The cytochrome domain of *Tc*CDH, which is present in both ascomycetes and basidiomycetes groups, possesses the specific conserved residues which interact with heme and a pair of cysteines (Figure S2) and this finding is in agreement with other CDH cytochrome motifs archived in Pfam indicating the importance of this domain in catalytic process of CDH activity.

The phylogenetic analysis presented here provides a new perception of classification of *T. clypeatus* CDH and categorizes it with ascomycetes group since it possesses all the three domains (cytochrome, dehydrogenase and CBM) as found in the ascomycetes IIA CDHs and an ascomycetes-like dehydrogenase domain devoid of the cellulose binding residues (Figure 2). Thus, the phylogenetic tree signifies a close relation of *Tc*CDH with class II CDH of ascomycetes. This is inspite of *T. clypeatus* sharing many important morphological features because of which it was classified as basidiomycetes in the first place by Heim [11] in 1951. It is to be noted here that this is not the only enzyme of *T. clypeatus* which showed ascomycetes like features. Another carbohydrate metabolizing enzyme, phosphoketolase (Pk), which was cloned from *T. clypeatus* (accession no: KF690709) and expressed in *Escherichia coli*, showed noticeable similarity with ascomycetes phosphoketolases as well [36]. Phylogenetic analysis of Pk sequence from *T. clypeatus* with other bacterial and fungal Pk sequences revealed that *Metarhizium anisopliae*, an ascomycete fungus, is its closest relative [36]. Again, phylogenetic analysis of yet another gene encoding endoxylanase, a pentose metabolising enzyme, from the same source (accession no: KX442586 [37]), revealed that *Bispora antennata*, an asomycete fungus, is its nearest relative [Dey and Roy, Unpublished observation].

Homology model of CDH from *T. clypeatus* has been constructed in order to evaluate its detail structure (Figure 3). The model has displayed the two distinct domains, cytochrome and dehydrogenase along with the ligands, heme and FAD respectively (Figure 3). In addition, the model also has displayed the CBM domain very clearly (Figure 3). Hence, the presence of all three domains in *Tc*CDH as determined by the Pfam analysis earlier is fully corroborated by the homology model of the enzyme. We have also rebuilt the protein along with its substrate, cellobiose, which helped us to investigate the substrate interacting residues. These residues are localized in the dehydrogenase domain and compared by aligning with other PDB structures from both basidiomycete and ascomycetes fungi. Eight residues (L354, P606, T629, R631, Y649, N732, H733, N781) and three residues (Y649, N732, H733) are found to be conserved in all of them on the basis of 5Å resolution and H-bonding interactions respectively (Figure 4(b)).

In conclusion, the results of our computational investigations reveal true identity of a gene (accession no: KX576458 and GenBank id: AOW70087.1 [15]) from *T. clypeatus* as a CDH enzyme. Four reasons in support of this conclusion and as discussed above are: i) presence of all three domains that are also found in the CDH of ascomycetes, i.e. cytochrome domain, dehydrogenase domain and the CBM; ii) the CBM of the *Tc*CDH, with its three cysteine and three aromatic amino acid residues, has been found to be highly conserved structurally when compared with the CBM domains of CDH from the ascomycetes group of fungus; iii) its cytochrome domain contains the conserved histidine and methionine (that bind with heme) as well as two cysteine residues (that help in the formation of disulphide linkages) that are universally present within the cytochrome domains of CDH of both the ascomycetes and basidiomycetes fungi and iv) presence of cellobiose specific eight substrate binding residues that are found within the dehydrogenase domain of the CDH of both the groups.

From an examination of its CDH sequence, phylogenetic placement and homology model, it is evident that *T. clypeatus* is close to ascomycetes group in spite of being morphologically similar to basidiomycetes group. The discordance in morphological [11] and molecular properties may be perhaps due to *T. clypeatus* being an intermediate fungus between ascomycetes and basidiomycetes since it bears mixed properties of both the groups. The work accomplished here reports the lack of cellulose binding residues within the dehydrogenase domain of *Tc*CDH similar to the ascomycetes CDHs but in deference to that of the basidiomycetes CDH. It seems that lack of those residues in *Tc*CDH and other ascomycetes made them to retain CBM to perform cellulose binding function. By combining the morphological features [11] and molecular properties of *T. clypeatus*, we can presume that it is a new group of fungus having properties of both ascomycetes and basidiomycetes and it is on its way to be fully differentiated as a basidiomycete in course of evolution in distant future. The CBM and dehydrogenase domains of *Tc*CDH may fuse during evolutionary process in coming days to perform both cellulose binding and catalytic function in a single domain like in the dehydrogenase domain of basidiomycetes CDHs. This scheme of thinking is in line with the general idea that basidiomycetes evolved from the ascomycetes during evolution and are more advanced in nature [38, 39]. Our study illustrates the unique features of CDH in the basidiomycetes fungus, *T. clypeatus* and the possible phylogenetic status of *T. clypeatus* discussed will help to gain further insight into the evolutionary position of this fungus in future.

Conflict of interest

Authors declare no conflict of interest.

Acknowledgement

This work has been supported in part by a grant from UGC, India to AR[1*] and other departmental funds by DBT, GoI. SB[1] is thankful to UGC and Visva-Bharati University for

fellowship support. MSM^2 and AR^2 would like to acknowledge a Wellcome trust-DBT, India alliance for a senior fellowship to MSM^2. We also thank members of the MSM^2 lab for useful discussions and criticisms.

References

[1] U. Westermark, K.E. Eriksson, Cellobiose: quinone oxidoreductase, a new wood-degrading enzyme from white-rot fungi, Acta.Chem.Scand, Series B: Organic Chemistry and Biochemistry. 28 (1974) 209-214.

[2] U. Westermark, K.E. Eriksson, Purification and properties of cellobiose: quinone oxidoreductase from *Sporotrichum pulverulentum*, Acta.Chem.Scand , Series B: Organic Chemistry and Biochemistry. 29 (1975) 419-424.

[3] M. Zamocky, R. Ludwig, C. Peterbauer, B.M. Hallberg, C. Divne, P. Nicholls, D. Haltrich, Cellobiose dehydrogenase - a flavocytochrome from wood-degrading, phytopathogenic and saprotropic fungi, Curr Protein Pept Sci. 7 (2006) 255-80.

[4] G. Henriksson, G. Johansson, G. Pettersson, 2000. A critical review of cellobiose dehydrogenases, J Biotechnol. 78 (2000) 93-113.

[5] B.M. Hallberg, T. Bergfors, K. Backbro, G. Pettersson, G. Henriksson, C. Divne, A new scaffold for binding heme in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase, Structure Fold Des. 8 (2000) 79-88.

[6] B.M. Hallberg, G. Henriksson, G. Pettersson, C. Divne, Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase, J. Mol. Biol. 315 (2002) 421-34.

[7] T.C. Tan, D. Kracher, R. Gandini, C. Sygmund, R. Kittl, D. Haltrich, B.M. Hallberg, R. Ludwig, C. Divne, Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation, Nat Commun. 6 (2015) 1-11.

[8] W. Harreither, C. Sygmund, M. Augustin, M. Narciso, M.L. Rabinovich, L. Gorton, D. Haltrich, R. Ludwig, Catalytic properties and classification of cellobiose dehydrogenases from ascomycetes, Appl. Environ. Microbiol. 77 (2011) 1804-1815.

[9] M. Yoshida, K. Igarashi, M. Wada, S. Kaneko, N. Suzuki, H. Matsumura, N. Nakamura, H. Ohno, M. Samejima, Characterization of carbohydrate-binding cytochrome b562 from the white-rot fungus *Phanerochaete chrysosporium*, Appl. Environ. Microbiol. 71 (2005) 4548−4555.

[10] O. Shoseyov, Z. Shani, I. Levy, Carbohydrate binding modules: biochemical properties and novel applications, Microbiol Mol Biol Rev. 70 (2006) 283-295.

[11] R. Heim, Les Termitomyces du Congo Belge recueillis par Medame M. Goossens-Fontana, Bulletin du Jardin Botanique de l'État à Bruxelles (in French). 21 (1951) 205–222.

[12] S. Ghorai, S.P. Banik, D. Verma, S. Chowdhury, S. Mukherjee, S. Khowala, Fungal biotechnology in food and feed processing, Food Res Int. 42 (2009) 577-587.

[13] T. Saha, D. Ghosh, S. Mukherjee, S. Bose, M. Mukherjee, Cellobiose dehydrogenase production by the mycelial culture of the mushroom *Termitomyces clypeatus*, Process Biochem. 43 (2008) 634-641.

[14] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779-815.

[15] S. Banerjee, A. Roy, Molecular cloning, characterisation and expression of a gene encoding cellobiose dehydrogenase from *Termitomyces clypeatus* (unpublished). https://www.ncbi.nlm.nih.gov/nuccore/KX576458.1?from=1&to=2523

[16] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403-410.

[17] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279-D285.

[18] R.C. Edgar, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792-7.

[19] J. Felsenstein, Evolutionary Trees from DNA Sequences: A maximum likelihood approach, J. of Mol. Evol. 17 (1981) 368-376.

[20] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol Biol. Evol. 33 (2016) 1870-4.

[21] B.M. Hallberg, G. Henriksson, G. Pettersson, A. Vasella, C. Divne, Mechanism of the reductive half-reaction in cellobiose dehydrogenase, J. Biol. Chem. 278 (2003) 7160-6.

[22] M.N. Nguyen, K.P. Tan, M.S. Madhusudhan, CLICK-topology-independent comparison of biomolecular 3D structures, Nucleic Acids Res. 39 (2011) W24-W28.

[23] P.M. Alahuhta, V.V. Lunin, The structure of *Bacillus pumilus* GH48 in complex with cellobiose and cellohexaose (to be published) 10.2210/pdb5CVY/pdb

[24] M. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Sci. 15 (2006) 2507-2524.

[25] M.J. Sippl, Recognition of errors in three- dimensional structures of proteins, Proteins Struct. Funct. Genet. 17 (1993) 355-62.

[26] J.U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science. 253 (1991) 164-70.

[27] R. Lüthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles. Nature. 356 (1992) 83-5.

[28] S.C. Lovell, I.W. Davis, W.B.I. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Cα geometry: φ/ψ and Cβ deviation, Proteins Struct. Funct. Genet. 50 (2002) 437-50.

[29] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia, Proteins. 66 (2007) 778-95.

[30] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho & B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2013) D490–5 (2014).

[31] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, ClustalW and ClustalX version 2, Bioinformatics. 23 (2007) 2947-8.

[32] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview Version 2-a multiple sequence alignment editor and analysis workbench, Bioinformatics. 25 (2009) 1189-91.

[33] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol Syst Biol. 7 (2011) 1-6.

[34] F.A. Rotsaert, B. Li, V. Renganathan, M.H. Gold, Site-directed mutagenesis of the heme axial ligands in the hemoflavoenzyme cellobiose dehydrogenase, Arch Biochem Biophys. 390 (2001) 206-14.

[35] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, 2004. UCSF Chimera--a visualization system for exploratory research and analysis, J Comput Chem. 25 (2004) 1605-12.

[36] P. Sarkar, A. Roy, Molecular cloning, characterization and expression of a gene encoding phosphoketolase from *Termitomyces clypeatus*, Biochem Biophys Res Commun. 447 (2014) 621-5.

[37] P. Dey, A. Roy, Molecular cloning, characterization and expression of a gene encoding endoxylanase from *Termitomyces clypeatus* (unpublished). https://www.ncbi.nlm.nih.gov/nuccore/KX442586.1?from=1&to=678

[38] https://www.nature.com/articles/srep06746 (accessed on 11 July, 2018)

[39] X. Liu, Q. Zhang, X. Xia, Phylogenetic relationships of ascomycetes and basidiomycetes based on comparative genomics analysis, Genes Genom (2017) 39: 1307-1316