Biologia (2023) 78:1875–1887
https://doi.org/10.1007/s11756-022-01309-y

ORIGINAL ARTICLE

Pcal_0976, a pullulanase homologue from *Pyrobaculum calidifontis*, displays a glycoside hydrolase activity but no pullulanase activity

Iqra Aroob[1] · Asifa Maqbool[1] · Ayesha Pervez[1] · Nasir Ahmad[1] · Mehwish Aslam[1] · Abeera Shaeer[1] · Naeem Rashid[1]

Received: 22 August 2022 / Accepted: 29 December 2022 / Published online: 10 January 2023

Abstract

In this study, we have cloned and characterized a novel protein, Pcal_0976, annotated as pullulanase in the genome sequence of hyperthermophilic archaeon *Pyrobaculum calidifontis*. Motif search showed two glucodextran_C like domains and a domain of unknown function (DUF4134) in Pcal_0976. Multiple alignment with close homologues demonstrated six stretches of conserved regions. When the gene encoding Pcal_0976 was expressed in *Escherichia coli*, the recombinant protein was produced in insoluble and inactive form, which was solubilized using guanidine hydrochloride and refolded in an active form in the presence of arginine. Refolded Pcal_0976 displayed hydrolysis of glycogen, dextran, dextrin and starch. No hydrolytic activity was detected against pullulan. These results indicate that Pcal_0976 may not be a pullulanase but a novel glycoside hydrolase. Further studies are needed to establish the role of Pcal_0976 in carbohydrate metabolism in this archaeon.

Keywords *Pyrobaculum calidifontis* · Hyperthermophile · Glycoside hydrolase · Pullulanase · Glycogenase · Glucodextran_C domain

Introduction

Pullulanases are a class of endo-acting glycoside hydrolases (GHs). Pullulanase I primarily acts on α-1,6-linkages in pullulan, starch and other glucans such as amylopectin, glycogen and limit dextrins. However, it is unable to cleave α-1,4-glycosidic bonds. Pullulanase II or amylopullulanase hydrolyses both α-1,4- and α-1,6-linkages in starch and other polysaccharides, while in pullulan it hydrolyses only α-1,6-glycosidic bonds (Ahmad et al. 2014). Both of these enzymes possess retaining mechanism of action, which means that the anomeric configuration of their reaction end products is retained. In the sequence-based classification of Carbohydrate-Active enZymes, the CAZy database, pullulanase I and pullulanase II (amylopullulanase) are classified into glycoside hydrolase families 13 (GH13), and 57 (GH57), respectively (Drula et al. 2022). These enzymes are multi-domain proteins having two to three catalytic residues (MacGregor et al. 2001; Satyanarayana and Nisha 2018; Janeček and Svensson 2022). The GH57 pullulanases are evolutionary important due to presence of DOMON like glucodextranase domains, which are 110 – 125 residue long domains (Aravind 2001) and are known for mediating extracellular interactions with heme and sugars (Iyer et al. 2007). There is a high demand for pullulanases in industrial utilization of biomass, which makes them an attractive choice of study (Wang et al. 2019). In particular, pullulanases with higher thermostability are of significant interest due to their applicability in processes requiring higher temperatures (Lévêque et al. 2000). Keeping this in view, we cloned an open reading frame from *Pyrobaculum calidifontis*, annotated as pullulanase. *P. calidifontis* is a hyperthermophilic archaeon isolated from the terrestrial hot spring in the Philippines (Amo et al. 2002b). Several novel and industrially potential thermostable enzymes from this archaeon have already been characterized (Ali et al. 2011; Amo et al. 2002a; Jamroze et al. 2014; Mehboob et al. 2020; Satomura et al. 2011; un Naeem et al. 2020). Draft genome sequence

Iqra Aroob and Asifa Maqbool contributed equally and, in their opinion, should be considered joint first author.

✉ Naeem Rashid
naeem.ff.sbs@pu.edu.pk; naeemrashid37@hotmail.com

[1] School of Biological Sciences, University of the Punjab, Quaid-e-Azam Campus, 54590 Lahore, Pakistan

of *P. calidifontis* (GenBank; CP000561.1) contains two open reading frames, Pcal_0976 (annotated as pullulanase) and Pcal_1616 (annotated as pullulanase/α-amylase). They exhibit 38% sequence identity with each other. Gene product of Pcal_1616 is the closest characterized counterpart of Pcal_0976. Pcal_1616 belongs to family GH57 and was biochemically characterized as pullulan hydrolase type II (Siddiqui et al. 2014; Rehman et al. 2018). After Pcal_1616, a characterized amylopullulanase, also belonging to family GH57, from *Thermococcus kodakarensis* (Guan et al. 2013) displayed 34% sequence identity with Pcal_0976. Keeping into consideration the importance of thermostable glucan hydrolyzing enzymes and novel sequence features of Pcal_0976, the present study describes *in silico* analysis followed by recombinant production and biochemical characterization of Pcal_0976.

Materials and methods

Reagents and chemicals

The reagents and chemicals used in this study were purchased either from Sigma-Aldrich (St. Louis, MO) or Thermo Fisher Scientific (Maryland, USA), if not mentioned otherwise. The restriction endonucleases, T4 DNA ligase, DNA and protein size markers, *Taq* DNA polymerase, RNase, and deoxynucleotide triphosphates (dNTPs) were from Thermo Fisher Scientific. Starch, pullulan, glycogen, dextran, dextrin and cyclodextrins (α- and γ-) were purchased from Sigma-Aldrich, while β-cyclodextrin was from Acros Organics (Maryland, USA).

Strains, plasmids and media

P. calidifontis strain VA1 was used to obtain Pcal_0976 gene. *Escherichia coli* DH5-α cells and plasmid pTZ57R/T (Novagen Merck, Germany) were used for cloning of the target gene. *E. coli* BL21 CodonPlus (DE3)-RIL cells (Stratagene, La Jolla, CA) and pET-21a(+) expression vector (Thermo Fisher Scientific) were used for heterologous expression of the target gene. *E. coli* strains were routinely grown in Luria-Bertani (LB) medium at 37 °C. Recombinant *E. coli* cells containing pET-21a(+) were selected on LB agar containing ampicillin (100 μg mL^{-1}), whereas X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; 40 μg mL^{-1}) and IPTG (isopropyl-β-D-galactopyranoside; 1 mM) were added when blue/white screening of recombinant *E. coli* cells containing pTZ57R/T was required.

Phylogenetic analysis and multiple sequence alignment

Uniprot-Beta Blast tool, with the target set to UniprotKB reference proteomes plus Swiss-Prot, was used to obtain the top 100 sequences with maximum identity. The sequences were analyzed carefully based on annotation scores and predicted domains. The sequences were aligned using ClustalOmega program (Sievers et al. 2011) freely provided at https://www.ebi.ac.uk/Tools/msa/clustalo/ by the European Bioinformatics Institute. Following initial alignment, 64 sequences were eliminated due to the absence of significant sequence similarity or momentous disruption of the multiple alignment or phylogenetic tree. Final multiple sequence alignment of the selected sequences was performed using ClustalW accessory application in BioEdit Sequence Alignment Editor (Hall et al. 2011). The alignment file obtained was processed in MEGA-X (Kumar et al. 2018) for UPGMA phylogenetic tree construction.

Estimation of the signal sequence

The signal sequence in Pcal_0976 was estimated using SignalP-5.0 Server (Almagro Armenteros et al. 2019) (https://services.healthtech.dtu.dk/service.php?SignalP-5.0).

Molecular modelling

Three-dimensional structure of Pcal_0976 was obtained using Alphafold structure prediction tool (https://alphafold.ebi.ac.uk/search/text/Pcal_0976), which directly predicts the 3D coordinates of all heavy atoms for a given protein using the primary amino acid sequence and aligned sequences of homologues as inputs (Jumper et al. 2021). Visualization of models and drawings were made in PyMOL (http://pymol.org).

Cloning of Pcal_0976 gene

Pcal_0976 gene, without signal sequence, was amplified by polymerase chain reaction (PCR) using genomic DNA of *P. calidifontis* strain VA1 as template and a set of sequence-specific forward (5′-CATATGGCCACAGACCCCACTGGCGACTAC)andreverse(5′-CTACTTCCGCCTGGCTGCTG) primers. These primers were commercially synthesized by Macrogen Inc. (Republic of Korea). An *Nde*I restriction enzyme site (underlined sequence) was incorporated in the forward primer. The PCR-amplified gene was inserted in pTZ57R/T cloning vector and the resulting plasmid was named Pcal_0976-pTZ. Recombinant plasmid, Pcal_0976-pTZ, was digested with *Nde*I and *Eco*RI to liberate the gene which was then cloned in pET-21a(+) expression vector by

utilizing the same restriction sites. The resulting plasmid was named Pcal_0976-pET.

Recombinant production of Pcal_0976 in *E. coli*

Transformation of *E. coli* BL21-CodonPlus (DE3)-RIL cells was carried out using Pcal_0976-pET recombinant plasmid. *E. coli* cells carrying Pcal_0976-pET plasmid were grown in LB medium, containing 100 µg/mL ampicillin, at 37 °C until the optical density at 600 nm reached 0.4. Isopropyl β-D-1-thiogalactopyranoside (IPTG), at a final concentration of 0.2 mM, was used to induce the gene expression. After 4 h of post-induction incubation at 37 °C, the cells were harvested by centrifugation for 10 min at 6000 × *g* and 4 °C. The cells, after resuspension in 50 mM Tris-Cl (pH 8.0), were lysed by sonication using the Bandelin SonoPlus HD 2070 sonication system (Bandelin Electronic, GmbH). After lysis, soluble and insoluble fractions were separated by centrifugation at 11,500 × *g*. Analysis of proteins was done by 12% denaturing polyacrylamide gel electrophoresis (SDS-PAGE).

Solubilization and refolding of recombinant Pcal_0976

Insoluble fraction, containing inclusion bodies of recombinant Pcal_0976, was washed 3 times with washing buffer (50 mM Tris-Cl pH 8.0, 5 mM EDTA, 10 mM NaCl, 1 mM PMSF, 0.5% Triton X-100) to remove the impurities (Singh et al. 2015) and dissolved in solubilization buffer containing 6 M guanidine, 10% glycerol, 20 mM DTT and 50 mM Tris-Cl pH 8.0). For refolding, the solubilized protein sample (0.5 mg/mL) was diluted in 100 mL of dilution buffer (10% glycerol, 20 mM arginine, 50 mM Tris-Cl pH 8.0) by adding 100 µL sample at a time after every 2 h. The refolded protein was concentrated by using 10 kDa MWCO ultrafiltration centrifugal devices (ThermoFisher Scientific). Protein concentration was estimated by the Bradford assay (Harlow and Lane 2006). Known concentrations of bovine serum albumin (BSA) were used to draw a standard curve.

Enzyme activity assay

Enzyme activity of recombinant Pcal_0976 was measured in terms of the amount of reducing sugars liberated upon incubation with the substrate as described previously (Ahmad et al. 2014; Aroob et al. 2019, 2022). The standard assay mixture containing 200 µL of 1% (w/v) substrate in 50 mM Tris-Cl buffer (pH 8.0) and desired amount of Pcal_0976 was incubated at 85 °C. The reaction was stopped by quenching in ice water, and the released reducing ends were determined by the dinitrosalicylic acid (DNS) method (Bernfeld 1955). One unit activity was defined as the amount of enzyme that released 1 nmol of reducing sugars in 1 min under standard assay conditions.

Biochemical characterization

To estimate the optimum temperature, the enzyme activity of Pcal_0976 was measured at different temperatures (60–90 °C) without changing the pH. Similarly, optimal pH was estimated by measuring the activity at various pH (6.0–8.5) and keeping the temperature unchanged.

Substrate specificity

The substrate preference and relative hydrolysis rates of various polysaccharides, including soluble starch, pullulan, glycogen, dextrin, dextran and cyclodextrins (α, β and γ) were determined by incubating each of these substrates at a final concentration of 1% (w/v) with recombinant Pcal_0976. Substrate solutions were prepared in 50 mM Tris-Cl buffer (pH 8.0) and the reaction was carried out at 85 °C. The relative hydrolysis rates were measured by the DNS method.

Results

Sequence analysis

In the draft genome sequence of *P. calidifontis*, Pcal_0976 (GenBank accession # ABO08401) has been annotated as pullulanase. Analysis of the amino acid sequence showed that Pcal_0976 contained a high number of Val (11.9%), Thr (11.1%), Ala (9.8%), Gly (8.2%), Pro (8.4%) and Leu (7.4%) residues. Overall, these six amino acids constituted nearly 59% of the protein. On the other hand, the number of His (0.3%), Cys (0.5%), Met (1.3%), Glu (1.6%), Trp (2.1%) and Lys (2.4%) were quite low. These six amino acids constituted only 8% of the protein. Among all the amino acids, the number of valine residues was the highest (11.9%) in Pcal_0976. Signal peptide prediction using SignalP 5.0 server revealed that Pcal_0976 contained a signal peptide at the N-terminal comprising 28 amino acids, which was predicted to be cleaved between Ala28 and Thr29. Presence of signal peptide indicates that Pcal_0976 may be a protein destined to perform its role extracellularly.

Pcal_0976 showed the highest homology (76% sequence identity) with the uncharacterized pullulanase (PAE3090) from *Pyrobaculum aerophilum*. Among biochemically characterized enzymes, the highest sequence identity of 38% was found with Pcal_1616, a pullulanase from *P. calidifontis* (Rehman et al. 2018), followed by 34% with an amylopullulanase from

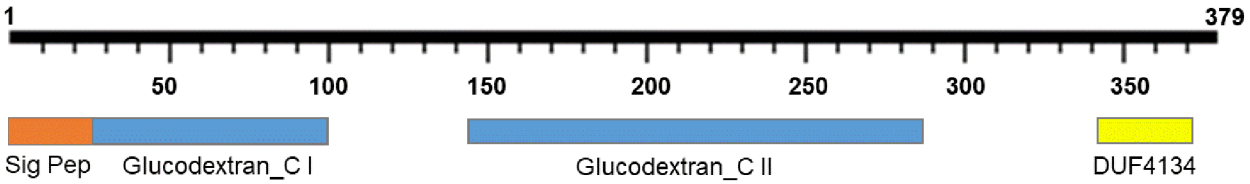

Fig. 1 Schematic diagram showing the distribution of signal peptide and various domains in Pcal_0976

T. kodakarensis (Guan et al. 2013). Both of these enzymes belong to family GH57 but Pcal_0976 (having 379 amino acid residues) has not yet been assigned to any of glycoside hydrolase family by the CAZy curators. Amino acid sequence analysis using SSDB motif search (Sato et al. 2001) revealed the presence of two glucodextran_C domains and a domain of unknown function (DUF4134) present at the C-terminus in Pcal_0976 (Fig. 1). Previously reported pullulanase from *P. calidifontis* Pcal_1616, which also is the closest characterized counterpart, despite having a longer sequence of 1001 amino acid residues, contained only a single glucodextran_C domain (Rehman et al. 2018). Typical catalytic domain (COG1449) of GH57 amylopullulanases was not found in Pcal_0976. The first glucodextran_C domain in Pcal_0976 comprised amino acids from position 28 to102 and the second included residues from 146 to 284. Glucodextran_C (Pfam ID: PF09985) is usually found as C-terminal domain of glucodextranase-like proteins in various prokaryotic membrane-anchored proteins (Mizuno et al. 2004).

The evolutionary relationship for phylogenetic tree construction was inferred using the UPGMA method. The optimal tree (Fig. 2) shows closer relatedness of Pcal_0976 (branch with an asterisk) with putative archaeal pullulanases and amylopullulanases belonging to class Thermoprotei or Thermococci. Little farther branches show proteins containing glucodextran_C domain and other extracellular or membrane-anchored proteins from bacterial or eukaryotic origin. Pcal_0976 can, therefore, possibly be assumed to form an evolutionary link between the prokaryotic and eukaryotic carbohydrate processing enzymes.

Presence of one or two DOMON like glucodextranase domains in addition to catalytic domain (COG1449) is a known characteristic of GH57 amylopullulanases (Jiao et al. 2013). The evolutionary history of glucodextran_C domains, either singular or dual, found in the members of Thermoprotei and Thermococci, was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones et al. 1992). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. The tree with the highest log likelihood (-3986.23) was selected (Fig. 3). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 13 amino acid sequences of glucodextran_C domains from Thermoprotei and Thermococci. There were a total of 251 positions in the final dataset. Evolutionary analyses, conducted in MEGA X (Kumar et al. 2018), revealed that two glucodextran_C domains (Gluc_C I and II) of amylopullulanase from *T. barophilus* appear to originate from the same root, which is an indicative of their evolution possibly due to a duplication event of glucodextran_C within the same organism. Pullulanase I from *P. aerophilum* (PAE3454, 999 AA) shares 71.8% sequence identity with Pcal_1616. Both these enzymes contain singular glucodextran_C domains which are clustered together (Figs. 3 and 4). Pullulanase II from *P. aerophilum* (PAE3090, 384 AA), similar to its closest homologue Pcal_0976, contains two glucodextran_C domains. Gluc_C I and II of Pcal_0976 and *P. aerophilum* pullulanase II appeared into two separate clades which were located very near to their singular counterparts within the same organisms (Fig. 3). Probably the latter two domains (Gluc_C I and II) may have derived from the singular glucodextran_C through intragenomic duplication or from gene fission of a full length glucodextran_C (Fig. 4). A similar intragenomic duplication event of Gluc_C domains was also proposed for *Thermococcus gammatolerans* amylopullulanase (Jiao et al. 2013).

In order to search for the conserved stretches, sequences used for phylogenetic analysis were further screened and only the pullulanases, amylopullulanases and glucodextran_C domain-containing proteins from archaeal origin were selected. The selected sequences were aligned using Clustal Omega (Sievers et al. 2011) and searched manually for the consensus sequences. Six conserved regions were found among these sequences despite overall significant differences in percent identity. A significant proportion of conserved residues was that of branched amino acids (isoleucine, leucine and valine), highlighted in purple in Fig. 5. Since typical signatures of family GH57 i.e., five conserved sequence regions, (β/α)7-barrel domain and catalytic machinery of GH57 members are missing in Pcal_0976, therefore it is worth mentioning that the conserved regions are present in the corresponding glucodextran domains. Sequence followed by these domains (residues 285 to 379) is unique and may be responsible for the catalytic activity and rendering the enzyme a unique candidate in glucan interacting enzymes.

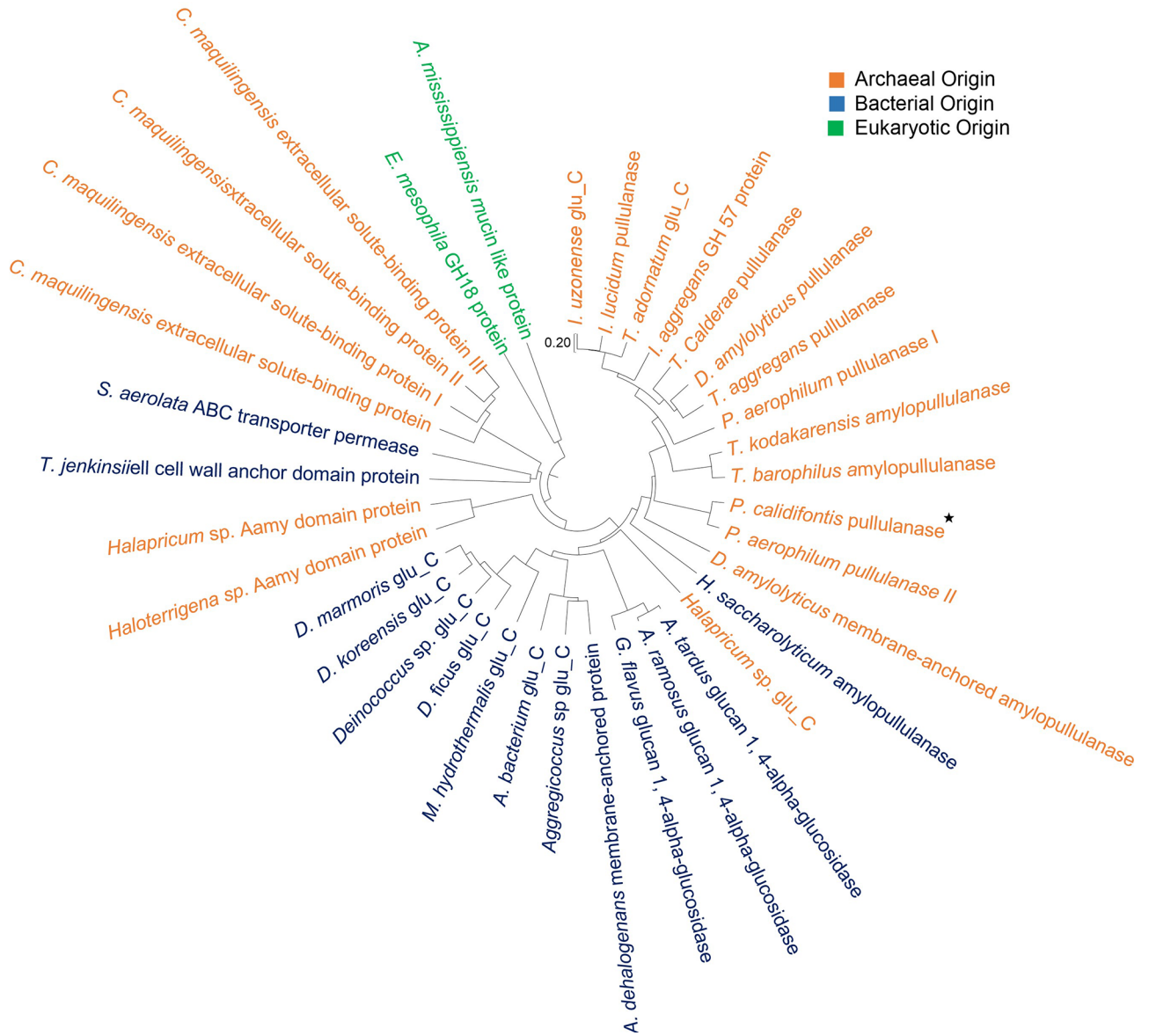

Fig. 2 Phylogenetic tree showing evolutionary relatedness of Pcal_0976 with closely related proteins as searched using BLAST tool in Uniport-Beta. Branches are labelled according to the putative functionality assigned to the protein in that organism. Glu_C refers to glucodextran_C domain containing protein. Following are the Uniprot accession numbers of the sequences used to construct the tree: *Infirmifilum uzonense* glu_C (A0A0F7FKA6), *Infirmifilum lucidum* pullulanase (A0A7L9FJR4), *Thermofilum adornatum* glu_C (S5ZN34), *Ignisphaera aggregans* GH57 protein (E0SSW7), *Thermogladius calderae* pullulanase (I3TCZ1), *Desulfurococcus amylolyticus* pullulanase (B8D5C4), *Thermosphaera aggregans* pullulanase (D5U365), *Pyrobaculum aerophilum* pullulanase I (Q8ZT36), *Thermococcus kodakarensis* amylopullulanase (Q5JJ55), *Thermococcus barophilus* amylopullulanase (F0LJB0), *Pyrobaculum calidifontis* pullulanase (A3MUT4), *Pyrobaculum aerophilum* pullulanase II (Q8ZTU6), *Ignisphaera aggregans* GH 57 protein (E0SSW7), *Halanaerobium saccharolyticum* subsp. *saccharolyticum* amylopullulanase (M5E3T7), *Halapricum* sp. glu_C (A0A6A9SWG2), *Agromyces tardus* glucan 1,4-alpha-glucosidase (A0A3M8AME6), *Agromyces ramosus* glucan 1,4-alpha-glucosidase (A0A4Q7MIT8), *Glaciibacter flavus* glucan 1,4-alpha-glucosidase (A0A4S4FP65), *Anaeromyxobacter dehalogenans* membrane-anchored protein (Q2IH39), *Aggregicoccus* sp. glu_C (A0A6I2GU60), *Alphaproteobacteria bacterium* glu_C (A0A4Q5WW30), *Marinithermus hydrothermalis* glu_C (F2NKS8), *Deinococcus ficus* glu_C (A0A221SZK2), *Deinococcus* sp. glu_C (A0A072N8×6), *Deinococcus koreensis* glu_C (A0A2K3UZS4), *Deinococcus marmoris* glu_C (A0A1U7NZL9), *Haloterrigena* sp. Aamy domain-containing protein (A0A7D5GT65), *Halapricum* sp. Aamy domain-containing protein (A0A6A9SYW8), *Tetrasphaera jenkinsii* cell wall anchor domain protein (A0A077MCP5), *Skermanella aerolata* ABC transporter permease (A0A512DN46), *Caldivirga maquilingensis* extracellular solute-binding protein (A8MB36), *Caldivirga maquilingensis* extracellular solute-binding protein I family 5 (A8M8S2), *Caldivirga maquilingensis* extracellular solute-binding protein II family 5 (A8ME10), *Caldivirga maquilingensis* extracellular solute-binding protein III family 5 (A8MDY2), *Exophiala mesophile* GH18 domain-containing protein (A0A0D1ZED0), *Alligator mississippiensis* mucin like protein (A0A151N7P2)

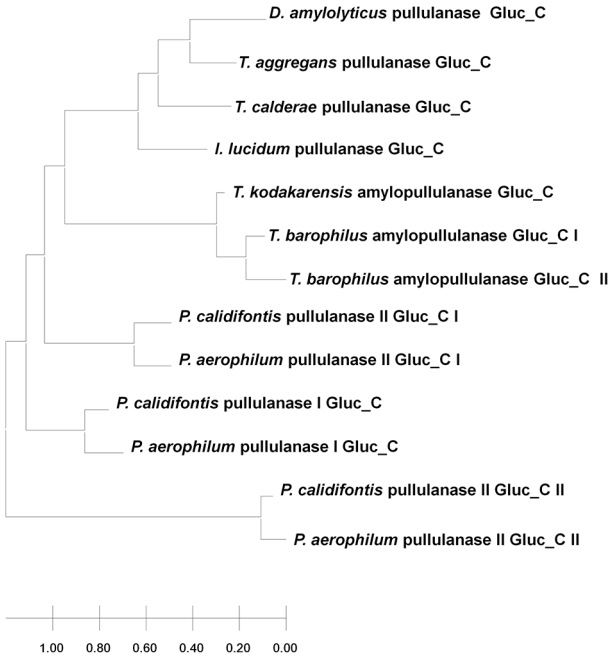

Fig. 3 Evolutionary history of glucodextran_C domains from the members of Thermoprotei and Thermococci

Tertiary structure prediction

Tertiary structure of Pcal_0976 was obtained from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/search/text/Pcal_0976). The closest template for structural modelling was of glucodextranase from *Arthrobacter globiformis* (PDB ID: 1UG9), belonging to the family GH15 (Mizuno et al. 2004). Glucodextran_C domains formed conserved beta sandwich like motifs made of antiparallel beta sheets (indicated in dark and light blue color) (Fig. 6).

Gene cloning and recombinant production of Pcal_0976

PCR using gene-specific primers resulted in amplification of nearly 1.1 kbp DNA fragment (Fig. 7 A), matching the size of Pcal_0976 gene. The amplified DNA fragment was ligated in pTZ57R/T cloning vector and digestion of the resulting plasmid, pTZ-Pcal_0976, with *Nde*I and *Eco*RI released nearly 1.1 kbp DNA fragment indicating the presence of Pcal_0976 gene in the recombinant plasmid (Fig. 7 B). Similarly, digestion of recombinant pET-Pcal_0976 utilizing the same pair of restriction enzymes resulted in the liberation of nearly 1.1 kbp DNA fragment indicating cloning of the gene in pET-21a(+) expression vector (Fig. 7 C). DNA sequencing showed absence of any mutation in the cloned gene.

Heterologous gene expression in *E. coli* BL21-CodonPlus (DE3)-RIL at 37 °C resulted in the production of Pcal_0976 in insoluble and inactive form (Fig. 8 A). Various attempts were made by changing the expression conditions to get the recombinant protein in soluble and active form. However, neither change in the cultivation temperature nor the inducing concentration of IPTG or lactose resulted in soluble production of recombinant Pcal_0976.

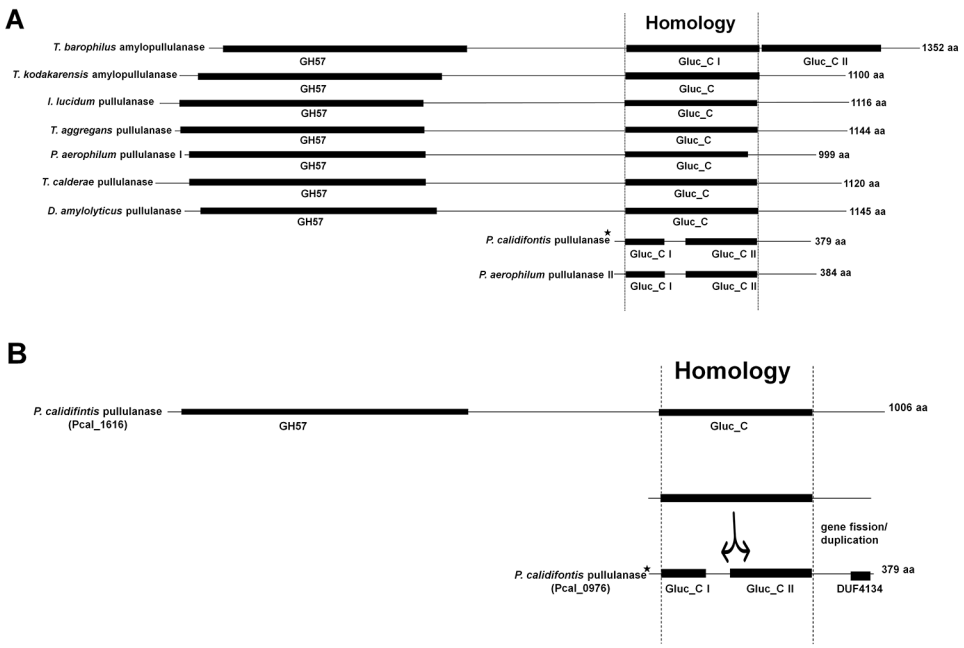

Fig. 4 (A) Conservation of Gluc_C domain in GH57 pullulanases/amylopullulanases from the members of Thermoprotei and Thermococci. Query sequence (*P. calidifontis pullulanase*, Pcal_0976) is indicated by asterik. (B) Proposed hypothetical gene fission/duplication event splitting Gluc_C domain into two separate Gluc_C domains within *P. calidifontis*

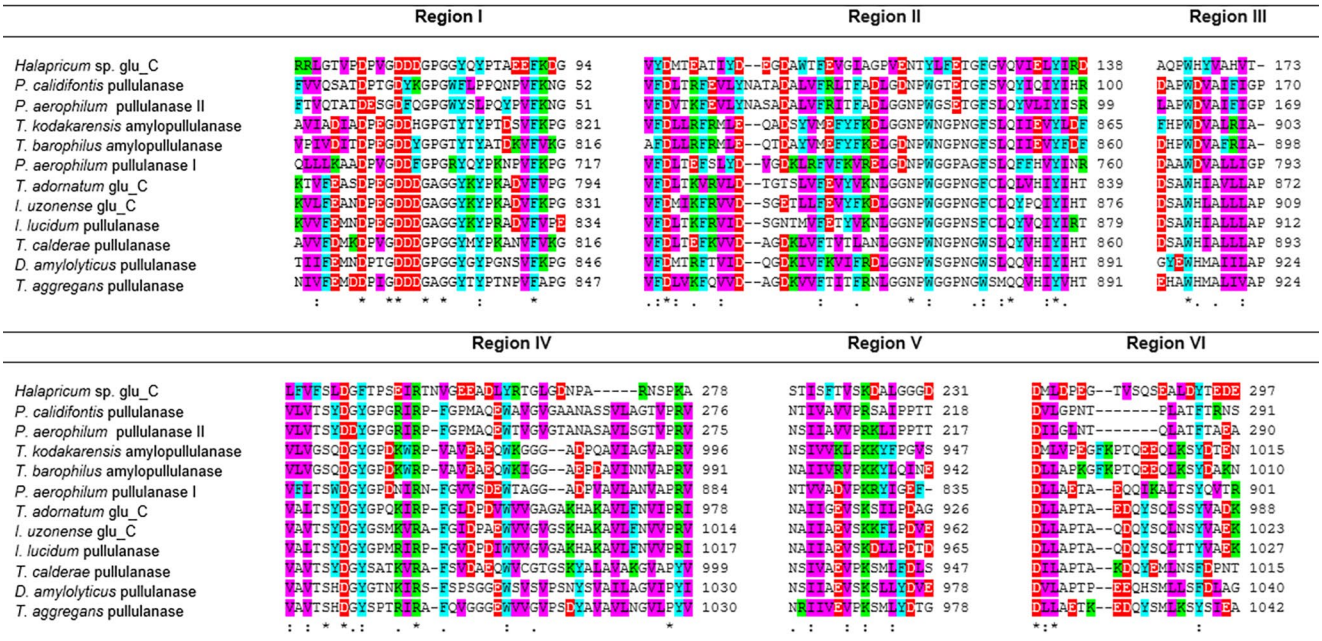

Fig. 5 Conserved sequence stretches identified in Pcal_0976 and archaeal homologs. Archaeal pullulanases, amylopullulanases and closely related glucodextran_C domain containing proteins were selected. Uniprot accession numbers used are the same as mentioned for the phylogenetic tree. Color key of amino acids is as: red (acidic), green (basic), cyan (hydrophobic) and purple (branched). Fully conserved residues are indicated by an asterisk (*), amino acids with strongly similar properties are indicated by a colon (:), while amino acids with weakly similar properties are indicated by a dot (.) at the bottom of the sequence

Solubilization and refolding of recombinant Pcal_0976

When the inclusion bodies containing recombinant Pcal_0976 were solubilized in 6 M guanidine hydrochloride and refolded by gradual removal of the denaturant by fractional dialysis, most of the recombinant Pcal_0976 got precipitated below 2 M guanidine hydrochloride. Therefore, after solubilizing in guanidine hydrochloride, protein refolding was attempted using dilution method. This method, supplemented with the use of arginine in the refolding buffer, was found successful in achieving soluble and active Pcal_0976. Homogeneity of the purified Pcal_0976 in the soluble form is demonstrated by SDS-PAGE (Fig. 8 B).

Biochemical characterization

To examine the effect of temperature, Pcal_0976 activity was assayed at various temperatures. The activity increased gradually with an increase in temperature till 85 °C and thereafter it started decreasing (Fig. 9 A). When analyzed for the optimum pH, Pcal_0976 exhibited highest activity at pH 8.0 in Tris-Cl buffer (Fig. 9B).

In order to determine the substrate preference, enzyme activity was examined against various carbohydrates. Glycogen was the most preferred substrate of Pcal_0976 with a specific activity of 595 mU/mg. Hydrolytic activities in case of typical glycogen branching enzymes are determined using amylose as substrate and these activities are very low ranging from 0.5 – 26 mU/mg (Xiang et al. 2022). Other enzymes acting on glycogen include glycogen phosphorylase from *E. coli* (100 mU/mg) (Alonso-Casajús et al. 2006) and α-amylase from *Bacillus* sp. TS-23 (19,700 mU/mg) (Lo et al. 2001). Other substrates hydrolyzed by Pcal_0976 were in the following order of preference: glycogen > dextran > dextrin > potato starch (Fig. 10).

When we examined the effect of glycerol and detergents on glycoside hydrolase activity of Pcal_0976, it was found that glycerol enhanced the enzyme activity moderately whereas a slight inhibition was observed in the presence of Triton-X100 (Fig. 11).

Discussion

This study aimed at finding a novel and thermostable carbohydrate processing enzyme. Primary structure comparison led us to assume that Pcal_0976 would act as a thermostable pullulanase. Thermostability of enzymes of (hyper)thermophilic origin comes out of their special structural features (Farias and Bonato 2003; Gharib et al. 2016) that are normally not present in proteins from mesophilic sources. The primary structure of Pcal_0976 showed high number of Ala, Gly, Leu, Pro, Thr and Val residues, while His, Cys, Met,

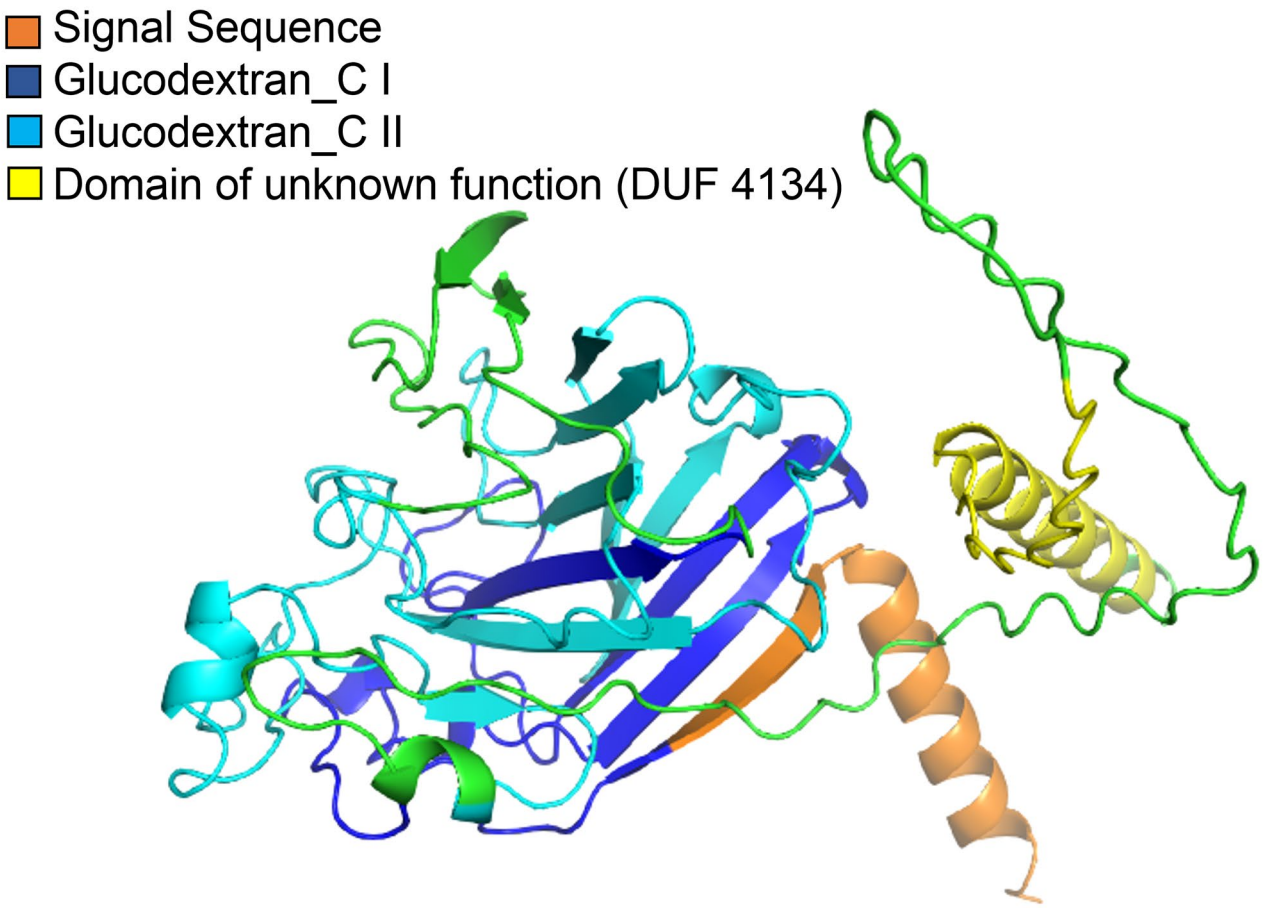

Fig. 6 Alphafold model of Pcal_0976. The orange region shows the signal sequence, while the light blue and dark blue regions correspond to glucodextran_C like domain I and II, respectively. The yellow region is representing the domain of unknown function (DUF4134)

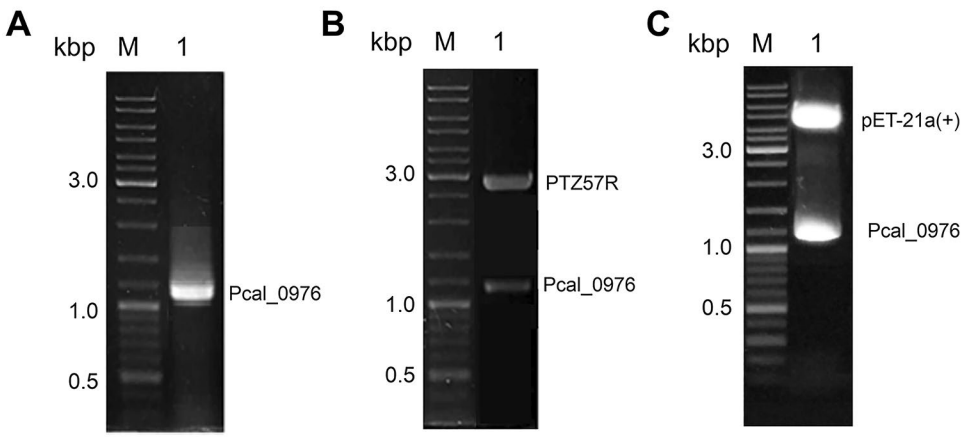

Fig. 7 Ethidium bromide stained agarose gel (1%) demonstrating PCR amplification and cloning of Pcal_0976 gene. (A) PCR amplification of Pcal_0976 gene. Lane M, standard molecular weight marker; lane 1, PCR amplified Pcal_0976 gene. (B) Cloning of Pcal_0976 gene in pTZ57R. Lane M, standard molecular weight marker; lane 1, recombinant Pcal_0976-pTZ digested with *Nde*I and *Eco*RI. (C) Cloning of Pcal_0976 gene in pET-21a(+) expression vector. Lane M, standard molecular weight marker; lane 1, recombinant Pcal_0976-pET digested with *Nde*I and *Eco*RI

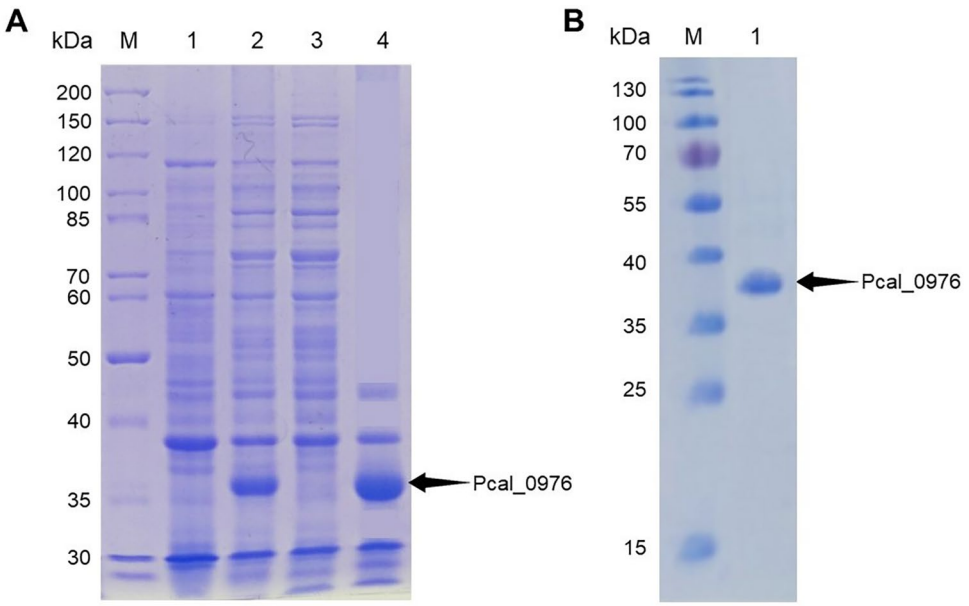

Fig. 8 SDS-PAGE (coomassie brilliant blue-stained) demonstrating production of recombinant Pcal_0976. (A) Lane M, standard marker; lane 1, total cell lysate of cells carrying pET-21a(+); lane 2, total cell lysate of cells carrying Pcal_0976-pET; lane 3, soluble fraction obtained after sonication of cells in lane 2; lane 4, insoluble fraction obtained after sonication of cells in lane 2. (B) Solubilization and refolding of recombinant Pcal_0976. Lane M, standard marker; lane 1: refolded and purified recombinant Pcal_0976

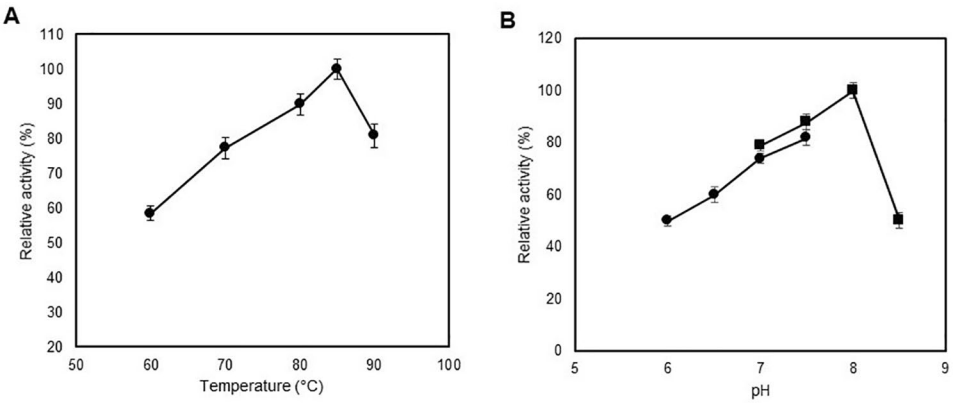

Fig. 9 Effect of temperature (A) and pH (B) on the Pcal_0976 activity. The effect of temperature on the enzyme activity was examined at various temperatures ranging from 60–90 °C using 50 mM Tris-Cl buffer, pH 8.0. The effect of pH was analyzed at 85 °C by determining activity of Pcal_0976 in buffers of different pH. Buffers used were 50 mM sodium phosphate (circles), and Tris–Cl (squares). The error bars represent the standard deviation

Glu, Trp and Lys were present in very low number. Cys, and His are usually avoided in thermostable and hyperthermostable proteins (Chohan et al. 2019; Farias and Bonato 2003) due to their tendency to undergo deamidation or oxidation at high temperatures (Kumar et al. 2000). The very low proportion of such amino acid residues might have contributed to the thermal stability of Pcal_0976. A highly thermostable L-asparaginase from *P. calidifontis* (Pcal_0970) also contained higher number of highly hydrophobic amino acids and very low number of thermolabile residues like Gln, Asp, and Cys (Chohan et al. 2019).

Thermostable pullulanases and amylopullulanases are industrially important enzymes required for liquefaction and saccharification of starch (Ahmad et al. 2014). In sequence based classification system (CAZy classification) pullulanases and amylopullulanases are grouped into family GH13 and GH57, respectively. Some of the amylopullulanases also belong to family GH13 (Jiao et al. 2013), one of the largest families of glycoside hydrolases. The members of family GH13 are characterized by the presence of a typical (β/α)8-barrel like catalytic domain, also known as TIM-barrel like domain. Other distinguishing features of members of family GH13 are activity towards α-glucosidic linkages, presence of four conserved regions in their sequences and a catalytic triad comprising two Asp and one Glu residues (Janeček and Svensson 2022). Members of family GH57 differ from family GH13 enzymes on the basis of sequences. Their catalytic domain adopts a pseudo TIM barrel i.e., (β/α)7-barrel and they possess only two catalytic residues (Glu and Asp) (Janeček and Svensson 2022). Most of the thermostable amylopullulanases belong to family GH57 and display relatively higher thermostability as compared to their counterparts from family GH13 (Janeček 2005; Kang et al. 2005). They possess a highly conserved DOMON_glucodextranase like domain which is usually located at C-terminus, in singular or dual format, preceded by the catalytic COG1449 domain. Thermostable amylopullulanases from family

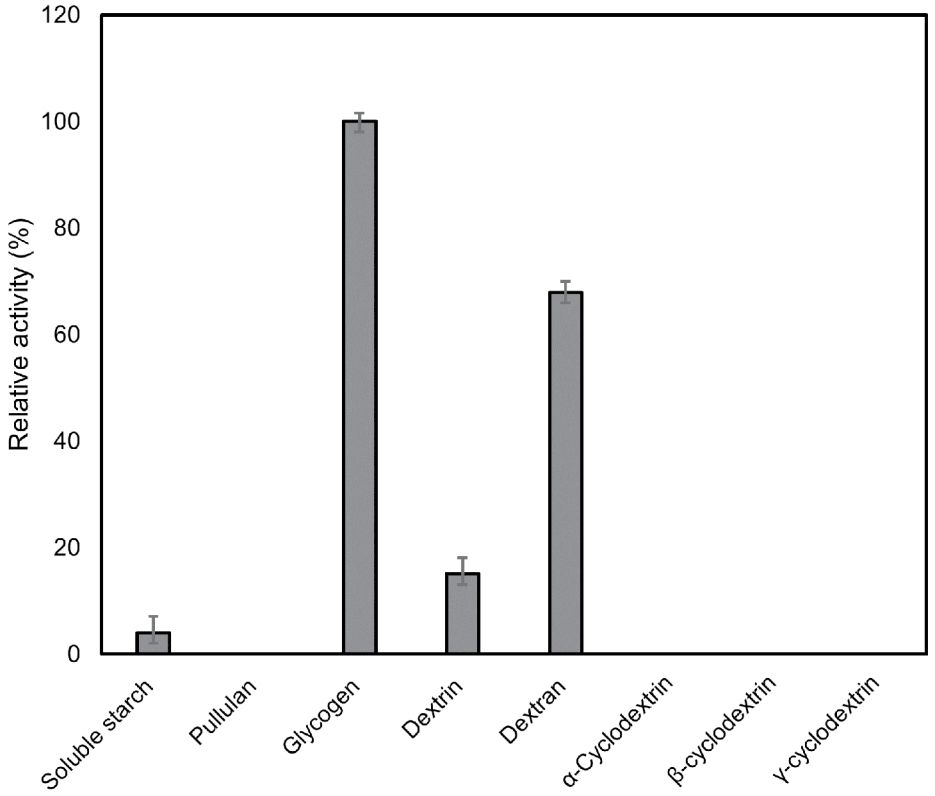

Fig. 10 Relative substrate preference of Pcal_0976 for different substrates. The activity was determined in terms of liberation of reducing ends in Tris-Cl buffer, pH 8.0 at 85 °C

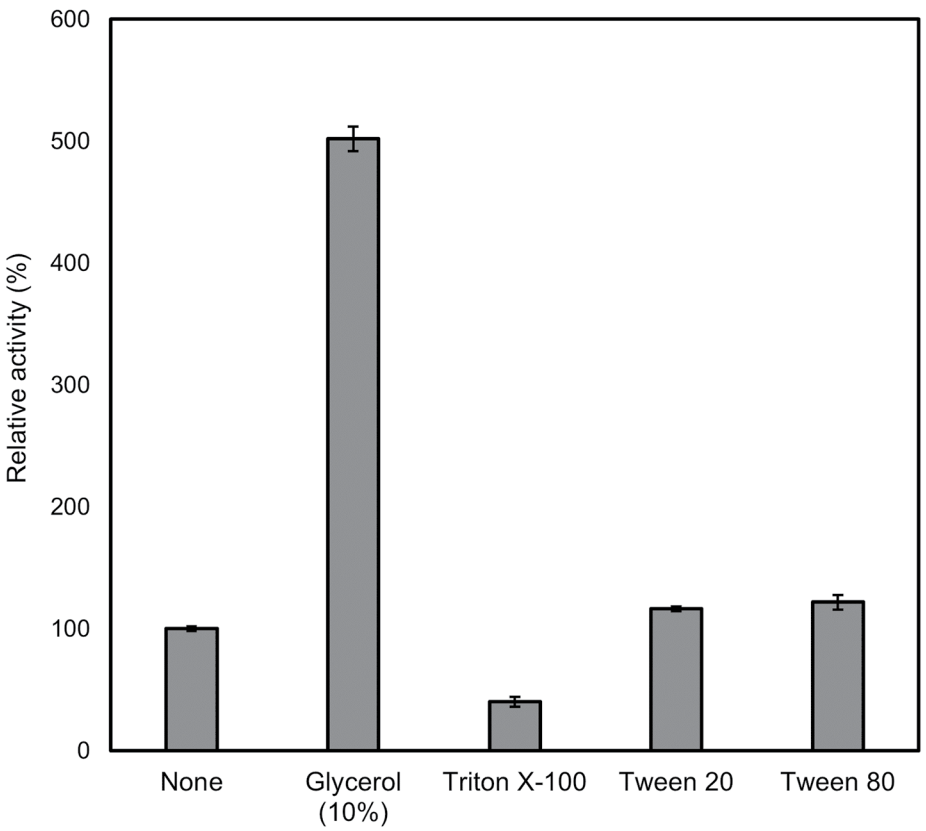

Fig. 11 Effect of additives on the activity of Pcal_0976. Enzyme activity was examined in the presence of various detergents at a final concentration of 1% except for glycerol (10%)

GH13 are reported to lack glucodextranase like domain (Jiao et al. 2013).

Glucodextranases along with glucoamylases are grouped into family GH15 which includes exo-acting amylases possessing inverting mechanism of action (https://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_15). The distinctive structural feature of members of family GH15 is the presence of (α/α)6 barrel like domain and two catalytic Glu residues. Glucodextranases contains four domains (N, A, B and C). Domains N and A are also conserved in glucoamylases and known to play role in catalytic activity. However, domains B and C are not found in these enzymes. Sequence of C domain in glucodextranases (gluc_C) shares homology with the surface layer homology (SLH) domain of GH57 amylopullulanases (Mizuno et al. 2004; Zona and Janeček 2005). It is β-strand-rich domain which adopts a β-sandwich-like fold, also known as DOMON like carbohydrate-binding domain. Gluc_C, SLH and other DOMON like domains (e.g. CBD9) are reported to serve as cell wall anchors and play role in polysaccharide metabolism at cell surface (Mizuno et al. 2004; Iyer et al. 2007).

Domain analysis of Pcal_0976 showed the presence of two DOMON like glucodextran_C domains, one near the N-terminal, the other in the center, and a domain of unknown function (DUF4134) at the C-terminus. On contrary to the previously studied glucodextran domains, which existed at the C-termini, the glucodextran like domains in Pcal_0976 were located at the N-terminus and in the middle region of protein's primary structure. Predicted structure model of Pcal_0976, obtained from the AlphaFold structure prediction tool, revealed that Gluc_C domains retained the conserved beta sandwich like motifs consisting antiparallel beta strands (Fig. 6). The closest structural homolog employed as template was of glucodextranase from *Arthrobacter globiformis* (PDB ID: 1UG9), belonging to the family GH15 (Mizuno et al. 2004). In contrast to members of families GH13, GH15 and GH57, Pcal_0976 lacked any catalytic domain (Figs. 1 and 6). This might be the reason for not being considered by the CAZy curators to be placed in any GH family. The closest characterized counterparts of Pcal_0976, the amylopullulanases from *P. calidifontis* and *T. kodakarensis* possessed singular Gluc_C domains (Siddiqui et al. 2014; Guan et al. 2013). Unlike its closest counterparts, Pcal_0976 contained a domain of unknown function (DUF4134), in addition to glucodextran_C like domains. DUF4134 is known to be found in symbiotic bacteria present in human gut which have role in glycan sensing and carbohydrate metabolism (Sonnenburg et al. 2005, 2006). Presence of this domain in Pcal_0976 indicates an orthology between bacterial and archaeal carbohydrate metabolism. The DUF4134 can be assumed of playing a role in glycan sensing and catalytic activity of the Pcal_0976.

Presence of unique domains and conserved regions make glycoside hydrolases distinct from other hydrolases (Janeček et al. 1997, 2003; Janeček 2002). However, identification of conserved sequence stretches in glucodextran_C containing enzymes is an interesting feature which needs exploration at structural levels.

When the gene encoding Pcal_0976 was expressed in *E. coli*, recombinant protein was produced in insoluble and inactive form. Refolding, after solubilization with denaturants, led us to achieve the protein in soluble and active form. However, it did not display any pullulanase activity. The highest activity was observed against glycogen followed by dextran. These findings were contrary to Pcal_1616, the closest characterized homolog (Rehman et al. 2018), and pullulanase from *Pyrococcus yayanosii* (Pang et al. 2019). Both of these enzymes contained glucodextran_C domain and displayed pullulanase activity. Our results suggest that Pcal_0976 may act as a membrane-anchored protein involved in carbohydrate transport and metabolism (Fujinami et al. 2017) instead of a typical pullulanase or amylopullulanase. Further studies are needed to understand the role of Pcal_0976 in *P. calidifontis*.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Ahmad N, Rashid N, Haider MS, Akram M, Akhtar M (2014) Novel maltotriose-hydrolyzing thermoacidophilic type III pullulan hydrolase from *Thermococcus kodakarensis*. Appl Environ Microbiol 80:1108–1115. https://doi.org/10.1128/AEM.03139-13

Ali SF, Rashid N, Imanaka T, Akhtar M (2011) Family B DNA polymerase from a hyperthermophilic archaeon *Pyrobaculum calidifontis*: cloning, characterization and PCR application. J Biosci Bioeng 112:118–123. https://doi.org/10.1016/j.jbiosc.2011.03.018

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z

Alonso-Casajús N, Dauvillée D, Viale AM, Muñoz FJ, Baroja-Fernández E, Morán-Zorzano MT, Eydallin G, Ball S, Pozueta-Romero J (2006) Glycogen phosphorylase, the product of the glgP gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*. J Bacteriol 188:5266–5272. https://doi.org/10.1128/JB.01566-05

Amo T, Atomi H, Imanaka T (2002a) Unique presence of a manganese catalase in a hyperthermophilic archaeon, *Pyrobaculum calidifontis* VA1. J Bacteriol 184:3305–3312. https://doi.org/10.1128/JB.184.12.3305-3312.2002

Amo T, Paje MLF, Inagaki A, Ezaki S, Atomi H, Imanaka T (2002b) *Pyrobaculum calidifontis* sp. nov., a novel hyperthermophilic archaeon that grows in atmospheric air. Archaea 1:113–121. https://doi.org/10.1155/2002/616075

Aravind L (2001) DOMON: an ancient extracellular domain in dopamine β-monooxygenase and other proteins. Trends Biochem Sci 26:524–526. https://doi.org/10.1016/S0968-0004(01)01924-7

Aroob I, Ahmad N, Aslam M, Shaeer A, Rashid N (2019) A highly active α-cyclodextrin preferring cyclomaltodextrinase from *Geobacillus thermopakistaniensis*. Carbohydr Res 481:1–8. https://doi.org/10.1016/j.carres.2019.06.004

Aroob I, Javed M, Ahmad N, Aslam M, Rashid N (2022) Investigating the role of carbohydrate-binding module 34 in cyclomaltodextrinase from *Geobacillus thermopakistaniensis*: structural and functional analyses. 3 Biotech 12:1–12. https://doi.org/10.1007/s13205-021-03089-9

Bernfeld P (1955) Amylases, alpha and beta. Methods Enzymol I 149–158. https://doi.org/10.1016/0076-6879(55)01021-5

Chohan SM, Rashid N, Sajed M, Imanaka T (2019) Pcal_0970: an extremely thermostable L-asparaginase from *Pyrobaculum calidifontis* with no detectable glutaminase activity. Folia Microbiol 64:313–320. https://doi.org/10.1007/s12223-018-0656-6

Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N (2022) The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 50:D571–D577. https://doi.org/10.1093/nar/gkab1045

Farias ST, Bonato M (2003) Preferred amino acids and thermostability. Genet Mol Res 2:383–393

Fujinami D, Taguchi Y, Kohda D (2017) Asn-linked oligosaccharide chain of a crenarchaeon, *Pyrobaculum calidifontis*, is reminiscent of the eukaryotic high-mannose-type glycan. Glycobiology 27:701–712. https://doi.org/10.1093/glycob/cwx044

Gharib G, Rashid N, Bashir Q, Gardner QA, Akhtar M, Imanaka T (2016) Pcal_1699, an extremely thermostable malate dehydrogenase from hyperthermophilic archaeon *Pyrobaculum calidifontis*. Extremophiles 20:57–67. https://doi.org/10.1007/s00792-015-0797-3

Guan Q, Guo X, Han T, Wei M, Jin M, Zeng F, Liu L, Li Z, Wang Y, Cheong G-W, Zhang S, Jia B (2013) Cloning, purification and biochemical characterisation of an organic solvent-, detergent-, and thermo-stable amylopullulanase from *Thermococcus kodakarensis* KOD1. Process Biochem 48:878–884. https://doi.org/10.1016/j.procbio.2013.04.007

Hall T, Biosciences I, Carlsbad C (2011) BioEdit: an important software for molecular biology. GERF Bull Biosci 2:60–61

Harlow E, Lane D (2006) Bradford assay. CSH protocols 2006: 1121–1132

Iyer LM, Anantharaman V, Aravind L (2007) The DOMON domains are involved in heme and sugar recognition. Bioinformatics 23:2660–2664. https://doi.org/10.1093/bioinformatics/btm411

Jamroze A, Perugino G, Valenti A, Rashid N, Rossi M, Akhtar M, Ciaramella M (2014) The reverse gyrase from *Pyrobaculum calidifontis*, a novel extremely thermophilic DNA topoisomerase endowed with DNA unwinding and annealing activities. J Biol Chem 289:3231–3243. https://doi.org/10.1074/jbc.M113.517649

Janeček S, Svensson B, Henrissat B (1997) Domain evolution in the α-amylase family. J Mol Evol 45:322–331. https://doi.org/10.1007/PL00006236

Janeček S (2002) How many conserved sequence regions are there in the α-amylase family. Biologia 57:29–41

Janeček S, Svensson B, MacGregor EA (2003) Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain. Eur J Biochem 270:635–645. https://doi.org/10.1046/j.1432-1033.2003.03404.x

Janeček S (2005) Amylolytic families of glycoside hydrolases: focus on the family GH57. Biologia 60:177–184

Janeček S, Svensson B (2022) How many α-amylase GH families are there in the CAZy. database? Amylase 6:1–10. https://doi.org/10.1515/amylase-2022-0001

Jiao YL, Wang SJ, Lv MS, Fang YW, Liu S (2013) An evolutionary analysis of the GH57 amylopullulanases based on the DOMON_glucodextranase_like domains. J Basic Microbiol 53:231–239. https://doi.org/10.1002/jobm.201100530

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Kang S, Vieille C, Zeikus JG (2005) Identification of *Pyrococcus furiosus* amylopullulanase catalytic residues. Appl Microbiol Biotechnol 66:408–413. https://doi.org/10.1007/s00253-004-1690-7

Kumar S, Tsai C-J, Nussinov R (2000) Factors enhancing protein thermostability. Protein Eng Des Sel 13:179–191. https://doi.org/10.1093/protein/13.3.179

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Lévêque E, Janeček S, Haye B, Belarbi A (2000) Thermophilic archaeal amylolytic enzymes. Enzyme Microb Technol 26:3–14. https://doi.org/10.1016/S0141-0229(99)00142-8

Lo H-F, Lin L-L, Chen H-L, Hsu W-H, Chang C-T (2001) Enzymic properties of a SDS-resistant *Bacillus* sp. TS-23 α-amylase produced by recombinant *Escherichia coli*. Process Biochem 36:743–750. https://doi.org/10.1016/S0032-9592(00)00273-9

MacGregor EA, Janeček S, Svensson B (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim Biophys Acta 1546:1–20. https://doi.org/10.1016/S0167-4838(00)00302-2

Mehboob S, Ahmad N, Munir S, Ali R, Younas H, Rashid N (2020) Gene cloning, expression enhancement in *Escherichia coli* and biochemical characterization of a highly thermostable amylomaltase from *Pyrobaculum calidifontis*. Int J Biol Macromol 165:645–653. https://doi.org/10.1016/j.ijbiomac.2020.09.071

Mizuno M, Tonozuka T, Suzuki S, Uotsu-Tomita R, Kamitori S, Nishikawa A, Sakano Y (2004) Structural insights into substrate specificity and function of glucodextranase. J Biol Chem 279:10575–10583. https://doi.org/10.1074/jbc.M310771200

Pang B, Zhou L, Cui W, Liu Z, Zhou S, Xu J, Zhou Z (2019) A hyperthermostable type II pullulanase from a deep-sea microorganism *Pyrococcus yayanosii* CH1. J Agri Food Chem 67:9611–9617. https://doi.org/10.1021/acs.jafc.9b03376

Rehman H, Siddiqui MA, Qayyum A, Bano A, Rashid N (2018) Gene expression in *Escherichia coli* and purification of recombinant type II pullulanase from a hyperthermophilic archaeon, *Pyrobaculum calidifontis*. Pak J Zool 50. https://doi.org/10.17582/journal.pjz/2018.50.4.1381.1386

Sato Y, Nakaya A, Shiraishi K, Kawashima S, Goto S, Kanehisa M (2001) Ssdb: sequence similarity database in KEGG. Genome Info 12:230–231. https://doi.org/10.11234/gi1990.12.230

Satomura T, Zhang X-D, Hara Y, Sakuraba H, Ohshima T (2011) Characterization of a novel dye-linked L-proline dehydrogenase from an aerobic hyperthermophilic archaeon, *Pyrobaculum calidifontis*. Appl Microbiol Biotechnol 89:1075–1082. https://doi.org/10.1007/s00253-010-2914-7

Satyanarayana T, Nisha M (2018) Archaeal and bacterial thermostable amylopullulanases: characteristic features and

biotechnological applications. Amylase 2:44–57. https://doi.org/10.1515/amylase-2018-0006

Siddiqui MA, Rehman H, Rashid N (2014) Gene cloning and characterization of a type II pullulanase hydrolase from a hyperthermophilic archaeon, *Pyrobaculum calidifontis*. Pak J Zool 46:1077–1084

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK (2015) Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Factories 14:1–10. https://doi.org/10.1186/s12934-015-0222-8

Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI (2005) Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. Science 307:1955–1959. https://doi.org/10.1126/science.11090

Sonnenburg ED, Sonnenburg JL, Manchester JK, Hansen EE, Chiang HC, Gordon JI (2006) A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. Proc Nat Acad Sci 103:8834–8839. https://doi.org/10.1073/pnas.0603249103

u Naeem S, Ahmad N, Rashid N (2020) Pcal_0842, a highly thermostable glycosidase from *Pyrobaculum calidifontis* displays both α-1, 4-and β-1, 4-glycosidic cleavage activities. Int J Biol Macromol 165:1745–1754. https://doi.org/10.1016/j.ijbiomac.2020.10.012

Wang X, Nie Y, Xu Y (2019) Industrially produced pullulanases with thermostability: discovery, engineering, and heterologous expression. Bioresour Technol 278:360–371. https://doi.org/10.1016/j.biortech.2019.01.098

Xiang G, Leemhuis H, van Der Maarel M, J E C (2022) Structural elements determining the transglycosylating activity of glycoside hydrolase family 57 glycogen branching enzymes. Proteins 90:155–163. https://doi.org/10.1002/prot.26200

Zona R, Janeček S (2005) Relationships between SLH motifs from different glycoside hydrolase families. Biologia 60:115–121