Supporting Information

KnowVolution of a GH5 cellulase from *Penicillium verruculosum* to improve thermal stability for biomass degradation

Francisca Contreras,[†] Martin J. Thiele,[†] Subrata Pramanik,[†] Aleksandra M. Rozhkova,[‡, §] Anna S. Dotsenko,[‡] Ivan N. Zorov,[‡, §] Arkady P. Sinitsyn,[‡, §] Mehdi D. Davari[*,†] and Ulrich Schwaneberg[*,†,‖]

†Affiliation 1: Institute of Biotechnology, RWTH Aachen University, Worringerweg 3, 52074 Aachen, Germany
‡Affiliation 2: Federal Research Centre «Fundamentals of Biotechnology», Russian Academy of Sciences, Leninsky prospect, 33, Bld.2, Moscow 119071, Russia
§Affiliation 3: Department of Chemistry, M.V. Lomonosov Moscow State University, Leninskiye Gory 1-3, Moscow 119991, Russia
‖Affiliation 4: DWI-Leibniz Institute for Interactive Materials, Forckenbeckstraße 50, 52074 Aachen, Germany

*Corresponding author: m.davari@biotec.rwth-aachen.de (MDD), u.schwaneberg@biotec.rwth-aachen.de (US)

Number of pages: 34

Number of figures: 10

Number of tables: 12

Number of methods: 8

Detailed explanation of Material and Method

Method M1: Cloning of synthetic gene eglII from *P. verruculosum* into pBSYA1S1Z

The parent *egl*II was ordered as a synthetic gene from ThermoFischer GeneArt (Regensburg, Germany) with an optimized codon usage for *P. pastoris*, which included *Xho*I and *Not*I restriction sites. After double digestion with *Xho*I and *Not*I, the fragment was subcloned using T4 DNA ligase into pBSYA1S1Z (**Figure S1**). The resultant recombinant plasmid, named pBSYA1S1Z::eglII, was transformed into *E. coli* DH5α and sequenced. The DNA sequencing was conducted at Eurofins MWG Operon (Ebersberg, Germany), and Clone Manager 9 Professional Edition (Sci-Ed software, Cary, USA) was used for all sequence alignments.

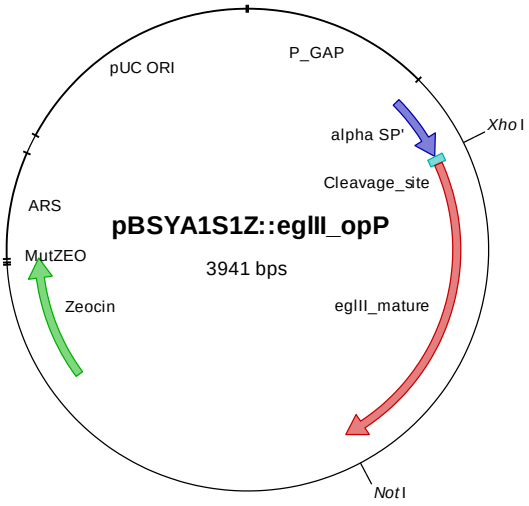


Figure S1: Representation of plasmid pBSYA1S1Z::*eglII*. pBSYA1S1Z is a shuttle vector (pUC Ori and ARS) with zeocin resistance as a selection marker. *egl*II mature represents the codon-optimized mature sequence of the gene. The gene is expressed under the GAP promoter (P_GAP) and secreted to the supernatant by the α-signal peptide (alpha SP).

Table S1: endo-β-Glucanase synthetic gene *egl*II from *P. verruculosum* was purchased as codon-optimized for *P. pastoris*. Underlined and bold nucleotides represent the restriction sites for cloning with *Xho*I and *Not*I into pBSYA1S1Z.

Endo-β-glucanase synthetic gene *egl*II

CTCGAGAAGAGAGAGGCCGAAGCTGCTAACTCTAAAGAGGTTAAGAAGAGAGCTT
CCTCCTTCGAATGGTTCGGTTCTAACGAATCTGGTGCTGAATTCGGTTCCGGTAAC
ATTCCAGGTGTTGAGGGTACTGACTACACTTTCCCAAACACTACTGCTATCCAGATC
TTGATCGACGCTGGTATGAACATCTTCAGAGTTCCATTCTTGATGGAAAGAATGATC
CCAACTGAGATGACTGGTTCCTTGGACACTGCTTACTTCGAAGGTTACTCCGAGGT
TTTGAACTACATCACTGGTAAGGGTGCTCACGCTGTTGTTGATCCACACAACTTCG
GTAGATATTACGGTACTCCAATCTCCTCCACTTCCGACTTCCAAACTTTCTGGTCTA
CTTTGGCTTCCCAGTTCAAGTCCAACGACTTGGTTATCTTCGACACTAACAACGAGT
ACCACGACATGGACGAGTCCGTTGTTGTTGCTTTGAACCAGGCTGCTATCGACGGT
ATTAGAGATGCTGGTGCTACTACTCAGTACATCTTCGTTGAAGGTAACGCTTACTCC
GGTGCTTGGACTTGGACTACTTACAACACTGCTATGGTTAACTTGACTGACCCATC
CGACTTGATCGTTTACGAGATGCACCAATACTTGGACTCCGACGGTTCTGGTACTT
CCGACCAATGTGTTTCTTCCACTGTTGGTCAAGAGAGAGTTGTTGACGCTACTACTT
GGTTGCAGTCCAACGGAAAGTTGGGTATCTTGGGTGAATTTGCTGGTGGTGCTAAC
TCCGTTTGTGAAGAGGCTGTTGAGGGAATGTTGGATTACTTGGCTGAGAACTCCGA
CGTTTGGTTGGGTGCTTCTTGGTGGTCTGCTGGTCCATGGTGGCAAGATTACATCT
ACTCTATGGAACCACCAAACGGTATCGCTTACGAGTCCTACTTGTCCATCTTGGAG
ACTTACTTCTAAGC**GGCCGC**

Method M2: PvCle5A activity assay with Azo-CMC

The hydrolytic activity was screened by using solubilized Azo-CMC as substrate, following the manufacturer instructions with some modifications. After cultivation in 96-well MTP, the PvCel5A containing supernatant was diluted 5-times with sodium acetate buffer (0.1 M, pH 4.5). The diluted supernatant was incubated without the substrate at 78 °C for 60 min. The diluted supernatant with (t60) and without (t0) incubation was transferred (40 µL) into an MTP for activity measurement. The enzyme reaction was initiated by the addition of 40 µL of Azo-CMC 2.0% (Megazyme, Ireland) in sodium acetate buffer (0.1 M, pH 4.5). The reaction mixture was incubated at 50 °C with shaking (ELMI Ltd., SkyLine DTS-4 Digital Thermo Shaker, 900 rpm) and stopped after 10 min with a precipitating solution (40 g/L sodium acetate, 4 g/L $ZnCl_2$, 80% technical grade ethanol). The precipitated reaction mix was centrifuged at 1000 *xg* for 10 min. Afterward, 100 µL of the clear supernatant was transferred into a 96-well MTP, and the absorbance measured at 590 nm (Tecan sunrise, Crailsheim, Germany).

Method M3: Library Generation: epPCR

Random mutagenesis was performed by error-prone PCR (epPCR) on the *egl*II gene in the commercial shuttle vector pBSYA1S1Z (Bisy e.U, Hofstaetten/Raab, Austria). For the epPCR, a thermal cycler (Eppendorf Mastercyler proS, Hamburg, Germany) (First stage: 94 °C for 2 min, 1 cycle; 94 °C 30 s/ 62 °C, 30 s/ 68 °C, 3 min, 25 cycles) Taq Polymerase (1 U), 0.20 mM dNTP mix, and 0.2 µM of each primer (primer sequences are shown in **Table S3**) together with template (~20 ng; pBSYA1S1Z_eglII_wt) were used. After, the PCR products were purified by using a NucleoSpin® Extract II Purification Kit (Macherey-Nagel, Düren, Germany) and *Dpn*I (20 U; New England Biolabs) was supplemented for digestion of maternal template DNA and incubated overnight at 37 °C. The *Dpn*I-digested PCR products were purified by using a NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany). Sub-cloning of the PvCel5A library was performed by MEGAWHOP.[1] Briefly, the generated library was utilized as megaprimer (500 ng) and mixed with the template (50 ng; pBSYA1S1Z_eglII_wt), Phusion polymerase (2 U), and 0.20 mM dNTP mix (72 °C for 3 min, 1 cycle; 95 °C, 30 s, 1 cycle; 95 °C 30 s/ 60 °C, 1 min/ 72 °C, 5 min, 25 cycles; 72 °C, 8 min, 1 cycle). After, the PCR product was supplemented with *Dpn*I (20 U; New England Biolabs) for digestion of maternal template DNA and incubated overnight at 37 °C. The *Dpn*I-digested PCR products were purified by gel excision using a NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany).

The purified PCR product (~ 20 ng) was transformed into *Pichia pastoris* BSYBG11 (*P. pastoris* BG11) and plated on YPD agar plates containing 100 µg/mL zeocin.

Method M4: Library Generation: Site-saturation Mutagenesis (SSM)

Site-saturation mutagenesis of endoglucanase PvCel5A was performed at 10 selected positions on endoglucanase PvCel5A in the commercial shuttle vector pBSYA1S1Z (Bisy e.U, Hofstaetten/Raab, Austria). For the site-saturation mutagenesis (SSM) PCR, a thermal cycler (Eppendorf Mastercyler proS, Hamburg, Germany) (First stage: 98 °C for 30 s, 1 cycle; 98 °C 10 s/ 62 °C, 30 s/ 72 °C, 4 min, 4 cycles. Second stage: 98 °C for 30 s, 1 cycle; 98 °C, 10 s/ 62 °C, 30 s/ 72 °C 4 min, 24 cycles; 72 °C for 10 min, 1 cycle), PhuS DNA Polymerase (2 U), 0.20 mM dNTP mix, and 0.2 µM of each NNK primer (primer sequences are shown in **Table S4**) together with template (~20 ng; pBSYA1S1Z_eglII_wt) were used. After the PCR, *Dpn*I (20 U; New England Biolabs) was supplemented for digestion of maternal template DNA and incubated overnight at 37 °C. The *Dpn*I-digested PCR products were purified by using a NucleoSpin® Extract II Purification Kit (Macherey-Nagel, Düren, Germany), transformed into *P. pastoris* BG11 and plated on YPD agar plates containing 100 µg/mL zeocin. The generated libraries were screened by measuring the initial activity of the variants and after incubation of PvCel5A containing supernatant without the Azo-CMC at 78 °C for 60 min. The activity was measured with the substrate Azo-CMC at 50 °C, 50 mM sodium acetate buffer pH 4.5, and improved PvCel5A variants compared to the WT endoglucanase PvCel5A and sequence analyzed.

Method M5: Recombination: Site-directed mutagenesis (SDM)

The recombination of beneficial positions (see **Table S8**) in the endoglucanase PvCel5A was performed by site-directed mutagenesis (SDM). For the SDM PCR, a thermal cycler (Eppendorf Mastercyler proS, Hamburg, Germany) (First stage: 98 °C for 30 s, 1 cycle; 98 °C 10 s/ 62 °C, 30 s/ 72 °C, 4 min, 4 cycles. Second stage: 98 °C for 30 s, 1 cycle; 98 °C, 10 s/ 62 °C, 30 s/ 72 °C 4 min, 24 cycles; 72 °C for 10 min, 1 cycle), PhuS DNA Polymerase (2 U), 0.20 mM dNTP mix, and 0.2 µM of each primer (primer sequences are shown in **Table S5**) together with template (~20 ng) were used. After the PCR, *Dpn*I (20 U; New England Biolabs) was supplemented for digestion of maternal template DNA and incubated overnight at 37 °C. The *Dpn*I-digested PCR products were purified by using a NucleoSpin® Extract II Purification Kit (Macherey-Nagel, Germany), transformed into *E. coli* DH5α and plated on low salt LB agar plates (tryptone 10 g/L, NaCl 5 g/L, yeast extract 5 g/L, and agar 15 g/L) containing 50 µg/mL zeocin. The generated recombinant variants were confirmed by sequencing in Eurofins MWG (Ebersberg, Germany). The plasmid containing the mutations were transformed into *P. pastoris* BG11 and plated on YPD agar plates containing 100 µg/mL zeocin. The activity of the transformants was measured with the substrate Azo-CMC at 50 °C, 50 mM sodium acetate buffer pH 4.5.

Method M6: Endoglucanase PvCel5A purification via ion-exchange chromatography (IEC)

For large-scale PvCel5A expression, 200 mL of YPD media was transferred to 2 L Erlenmeyer flasks and supplemented with Zeocin 100 µg/mL. The main culture was inoculated with PvCel5A in *P. pastoris* BG11 to an OD_{600nm} of 0.2 and cultured for 96 h (250 rpm, 25 °C, 96 h, and 70% humidity). Clear supernatant was obtained after centrifugation (Sorvall, RC 6 Plus Centrifuge, ThermoScientific, 10000 ×*g*, 10 min, 4 °C) of the cell culture supernatant. The activity of PvCel5A in the supernatant was assessed by the Azo-CMC assay (**Figure S2A**). Afterward, the supernatant was concentrated by cross-flow filtration (VivaFlow 200, 10 kDa MWCO PES; Sartorius) to 90 mL. The concentrated supernatant was loaded into an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO; Merck Millipore), concentrated to 6 mL final volume, and the YPD media exchanged with Bis-Tris buffer pH 6.5, 20 mM (buffer A). The PvCel5A was purified by FPLC (ÄKTAprime plus chromatography system, GE Healthcare, Solingen, Germany) as follows: the concentrated supernatant was loaded into an anion exchange chromatography column (GE Healthcare HiTrap Capto Q ImpRes, 5 mL) equilibrated with buffer A; PvCel5A was eluted in a linear gradient from buffer A 15% to Bis-Tris buffer pH 6.5, 20 mM, NaCl 1M (Buffer B) 50%. The correct peak fractions (~35 kDa) were analyzed by SDS-Page and pooled together (**Figure S2B-D**).[2] Differences in the apparent protein size may be attributed to a heterogeneous glycosylation pattern produced by *P. pastoris*.[3-4] Protein concentration of the purified PvCel5A was normalized by measuring the total protein concentration utilizing A_{280nm} (NanoDrop™ 1000 spectrophotometer by Thermo Scientific™, Bremen, Germany), using the theoretical extinction coefficients determined based on the amino acid composition with ProtParam on the ExPASy server (**Table S2**).[5]

Table S2: Extinction coefficients employed for protein concentration measurements at A_{280nm}

PvCel5A variant	Extinction coefficient [1/M cm]
WT	81 945
V1	80 455
R17	80 455

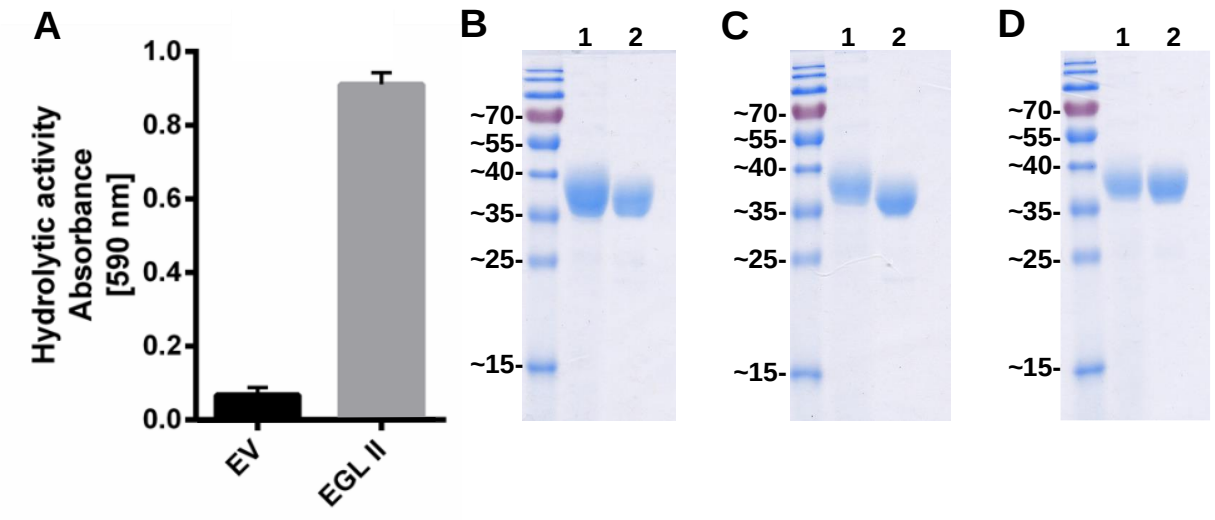


Figure S2: Supernatant containing PvCel5A. **A:** Activity of PvCel5A in the supernatant of *P. pastoris* measured with the Azo-CMC assay at 50 °C pH 4.5. SDS-PAGE of **B:** PvCel5A-WT, **C:** Cel5A variant V1 and **D:** Cel5A variant R17. Concentrated supernatant containing PvCel5A (lane 1, ~35 kDa) and PvCel5A purified via ion-exchange (lane 2, ~35 kDa) obtained with ≥85% purity.

Primers used for cloning *egl*II gene and generation of the epPCR library of PvCel5A

Table S3: Primers used for cloning of the endoglucanase PvCel5A in pBSYA1S1Z. Underlined are the restriction sites utilized for cloning.

Name	Sequence (5‘ to 3‘)	Restriction site
Fw_epPCR_eglII	GTCTCTCTCGAGAAGAGAGAGGCCG	*Xho*I
Rv_epPCR_eglII	CTCTTGAGCGGCCGCTTAGAAGTAAG	*Not*I

Primers used for site-saturation mutagenesis (SSM) of the endoglucanase PvCel5A gene:

Table S4: Primers used for SSM of the endoglucanase PvCel5A gene in pBSYA1S1Z. Underlined are the degenerated codons utilized for saturation.

Name	Sequence (5‘ to 3‘)
Fw_SSM_Y293	GGCAAGATTACATCNNKTCTATGGAACCACCAAAC
Rev_SSM_Y293	CCGTTTGGTGGTTCCATAGAMNNGATGTAATCTTG
Fw_SSM_E263	TGAAGAGGCTGTTNNKGGAATGTTGGATTAC
Rev_SSM_E263	AATCCAACATTCCMNNAACAGCCTCTTCAC
Fw_SSM_T312	TACTTGTCCATCTTGGAGNNKTACTTCTAAGCG
Rev_SSM_T312	CGCTTAGAAGTAMNNCTCCAAGATGGACAAGTAG
Fw_SSM_S305	ATCGCTTACGAGNNKTACTTGTCCATC
Rev_SSM_S305	ATGGACAAGTAMNNCTCGTAAGCGATACC
Fw_SSM_E304	CGGTATCGCTTACNNKTCCTACTTGTCCATC
Rev_SSM_E304	GGACAAGTAGGAMNNGTAAGCGATACC
Fw_SSM_T69	AAAGAATGATCCCANNKGAGATGACTGGTTCC
Rev_SSM_T69	GGAACCAGTCATCTCMNNTGGGATCATTCTTTC
Fw_SSM_S114	ACTCCAATCTCCNNKACTTCCGACTTC
Rev_SSM_S114	GAAGTCGGAAGTMNNGGAGATTGGAGTACC
Fw_SSM_S308	CGAGTCCTACTTGNNKATCTTGGAGACTTAC
Rev_SSM_S308	GTAAGTCTCCAAGATMNNCAAGTAGGACTCG
Fw_SSM_Q289	CTGGTCCATGGTGGNNKGATTACATCTACTC
Rev_SSM_Q289	TAGAGTAGATGTAATCMNNCCACCATGGACCAG
Fw_SSM_H96	GGTAAGGGTGCTNNKGCTGTTGTTG
Rev_SSM_H96	ATCAACAACAGCMNNAGCACCCTTAC

Primers used for site-directed mutagenesis (SDM) of the PvCel5A gene in pBSYA1S1Z

Double variants containing the primers T312R, S305F, E304R, S308P, and Q289G were performed with the PvCel5A-WT gene as a template. Double variants containing the primers S308P_V5 and E304R_V5 were performed with the PvCel5A-S305F gene as a template.

Table S5: Primers used for SDM of the PvCel5A gene in pBSYA1S1Z. Underlined are the codons utilized for SDM

Name	Sequence (5‘ to 3‘)
Fw_T312R	ATCTTGGAGCGTTACTTCTAAGC
Rv_T312R	GCTTAGAAGTAACGCTCCAAG
Fw_S305F	TCGCTTACGAGTTCTACTTGTCCATC
Rv_S305F	GGACAAGTAGAACTCGTAAGCGATAC
Fw_E304R	GGTATCGCTTACAGGTCCTACTTG
Rv_E304R	CAAGTAGGACCTGTAAGCGATACC
Fw_S308P	GTCCTACTTGCCTATCTTGGAGAC
Rv_S308P	CTCCAAGATAGGCAAGTAGGACTC
Fw_Q289G	TCCATGGTGGGGTGATTACATC
Rv_Q289G	GATGTAATCACCCCACCATGGAC
Fw_S308P_V5	TTCTACTTGCCTATCTTGGAGAC
Rv_S308P_V5	CTCCAAGATAGGCAAGTAGAACTC
Fw_E304R_V5	GTATCGCTTACAGGTTCTACTTGTC
Rv_E304R_V5	CAAGTAGAACCTGTAAGCGATACC

Method M7: Degradation of cellulose by a cellulolytic cocktail

Degradation of microcrystalline cellulose (MCC) was assessed with a cellulolytic cocktail composed of an endoglucanase PvCel5A (*Penicillium verruculosum*), cellobiohydrolase HjCBHI (*Hypocrea jecorina*, Sigma-Aldrich, Germany) and β-glucosidase AnBGL (*Aspergillus niger*, Megazyme, Ireland). A reconstituted cocktail was produced with PvCel5A (WT, variant V1 or R17), HjCBHI, and AnBGL in a concentration of 4, 6, and 1 mg of protein per gram of MCC, respectively. First, 500 µL of MCC 1.0% suspension in sodium acetate buffer 50 mM pH 4.5 was pre-incubated with PvCel5A during 16 h at 70 °C, 900 rpm (Thermos scientific Shaking Drybath). Later, the mixture of HjCBHI and AnBGL was added to the pre-incubated MCC suspension and incubated at 50 °C. Samples of 50 µL were extracted at 24, 48, 72, 96, and 120 h after the addition of the CBHI-BGL mixture, and solubilized sugars were separated from MCC by centrifugation (Eppendorf 5424, 12 000 *xg*, 3 min), the supernatant was used for glucose concentration determination. The released glucose was measured with the Glucose HK assay kit (Megazyme, Ireland), in a 96 well MTP following the manufacturers' instructions. The glucose concentration was determined towards a D-glucose calibration curve ranging from 0.2 to 9 g/L.

Method M8: Molecular Dynamics (MD) Simulations

The starting structure of PvCel5A-WT was taken from the X-ray crystal structure of *Penicillium verruculosum* endoglucanase (PDB ID: 5L9C chain A). The PvCel5A variants V1 (F16L/Y293F) and R17 (F16L/Y293F/Q289G) structures were generated, and stabilization energies (ΔΔG) were calculated based on PvCel5A-WT structures using FoldX plugin implemented in YASARA software version 17.8.19.[6-8] Prior to simulation, water molecules and ligands (cellobiose and N-acetyl-D-glucosamine) were removed from the crystal structure using YASARA software version 17.8.19. Molecular dynamics (MD) simulations of PvCel5A-WT and its variants V1 (F16L/Y293F) and R17 (F16L/Y293F/Q289G) were performed at four temperatures, including 26.85 °C, 70 °C, 75 °C and 80 °C. Simulation at 27 °C was considered to identify baseline dynamics and to compare the structural features of the PvCel5A-WT and its variants V1 and R17 at standard conditions, and later at elevated temperatures. Temperatures 70 °C, 75 °C, and 80 °C were considered to reflect thermal stability characteristics in experimental conditions. All MD simulations were performed using the GROMACS v5.1.2 package[9-11] with the Amber99SB-ILDN force field and SPC/E water model.[12-13] The protonation state of titratable residues was determined based on *p*Ka estimation using PROPKA method implemented in the PDB2PQR server using the AMBER99 force field at pH 4.5.[14-17] The protonation was assigned Nδ1 of catalytic residue H156 based on the proton transfer mechanism between the serine, the histidine, and the aspartate residues of the catalytic triad, which is involved in activation of the hydroxyl group of the catalytic serine. Specifically, protonation was assigned to Nδ1 for H96, H102, and H207 based on possible H-bonds networks, and both Nδ1 and Nε2 protonated for H144 based on PROPKA calculation. Subsequently, enzymes were inserted in a cubic box (placed at least 1.5 nm from the box edge), then explicitly solvated in water, and sodium ions were added to neutralize the charge of the systems. The electrostatic interactions were calculated by applying the particle mesh Ewald (PME) method.[18-19] Short-range electrostatic interactions (rcoulomb) and Van der Waals (rvdw) were calculated using a cut-off value of 1.0 nm. Energy minimization of each system was performed individually using the steepest descent minimization algorithm until the maximum force reached 1000.0 kJ/mol/nm. Final system equilibration was conducted in the NVT and NPT ensembles in

four temperatures 27 °C, 70 °C, 75 °C and 80 °C. The production run was carried out using NPT ensemble for 100 ns with a time step of 2 fs for each temperature, respectively. The trajectories were visualized using VMD 1.9.1[20] and analyzed by GROMACS 5.1.2 tool.[9-11]

Supporting Data and Figures

Summary of KnowVolution campaign for increased thermostability of PvCel5A

A KnowVolution campaign consisting of four Phases (I-IV) was performed to identify potential beneficial amino acid positions that govern improved thermal stability of the endoglucanase PvCel5A (**Figure S3**).

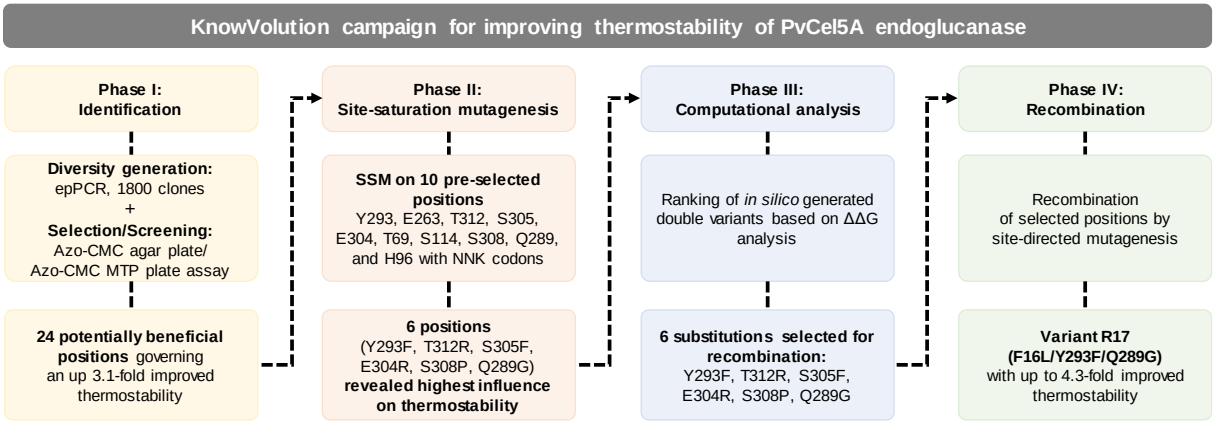


Figure S3. KnowVolution campaign of PvCel5A endoglucanase for improved thermal stability. In the first Phase of KnowVolution,[21] potentially beneficial positions were identified. Ten positions were pre-selected and saturated by SSM in Phase II. Six beneficial positions influencing the thermal stability were identified and subsequently analyzed in-depth by computational analysis. In Phase III, the beneficial substitutions (Y293F, T312, S305, E304, S308P, and Q289G) were computationally selected for recombination following the CompassR rule.[22] In Phase IV, recombination experiments were performed by SDM to generate double and triple PvCel5A variants. A PvCel5A variant was selected as improved if two conditions are satisfied: (1) the residual activity of the improved PvCel5A variant is higher than the residual activity of PvCel5A WT + coefficient of variance of the Azo-CMC assay (~10%) and (2) the initial activity of the PvCel5A variant is higher than 80% of the initial activity of PvCel5A WT.

PvCel5A variants obtained from the KnowVolution phase I with improved thermostability

Variants obtained from a random mutagenesis library of PvCel5A generated by epPCR supplemented with 0.3 mM $MnCl_2$. The thermostability was assessed by measuring the activity before and after incubation at 78 °C for 60 min without the substrate. The activity was measured with the Azo-CMC assay at 50 °C in sodium acetate buffer 50 mM pH 4.5. The residual activity of the variants and WT were used to calculate the improvement fold.

Table S6: Identified PvCel5A variants obtained from the KnowVolution phase I with improved thermostability. Positions in bold were selected for site-saturation mutagenesis.

Variant	Substitution	[a]Improved thermostability in fold	[b]Relative Initial activity [%]	Residual activity [%]
V1	F16L; **Y293F**	3.1	95	61
V2	**Y293F**	2.9	97	57
V3	K130R; I246F; **E263K**	2.5	94	48
V4	**T312S**	2.4	102	46
V5	**S305F**	2.2	91	44
V6	**S305F**	2.0	94	39
V7	T69A; **E304K**	1.9	108	44
V8	**S114P**; **T312A**	1.9	99	37
V9	**E304V**	1.9	96	44
V10	T77A; N299D; **S308P**	1.8	82	35
V11	**S308P**	1.7	100	34
V13	**S308P**	1.7	103	38
V14	**S308P**	1.6	95	37
V15	S27P; S240T	1.6	96	31
V16	F16L; **H96R**; **S308P**	1.6	96	36
V17	S131F; **S308P**	1.6	93	35
V18	**S308P**	1.6	104	31
V19	**Q289R**	1.4	102	32
V20	**S308P**	1.4	102	33
V21	**S114P**	1.4	100	28
V22	**S308P**	1.4	100	31

[a] Improvement is determined as the ratio of the residual activity of the improved PvCel5A variant and the residual activity of PvCel5A-WT.

[b] Initial activity of variants normalized against the initial activity of PvCel5A-WT

Selection criteria for saturation mutagenesis on positions obtained from Phase I of the KnowVolution campaign

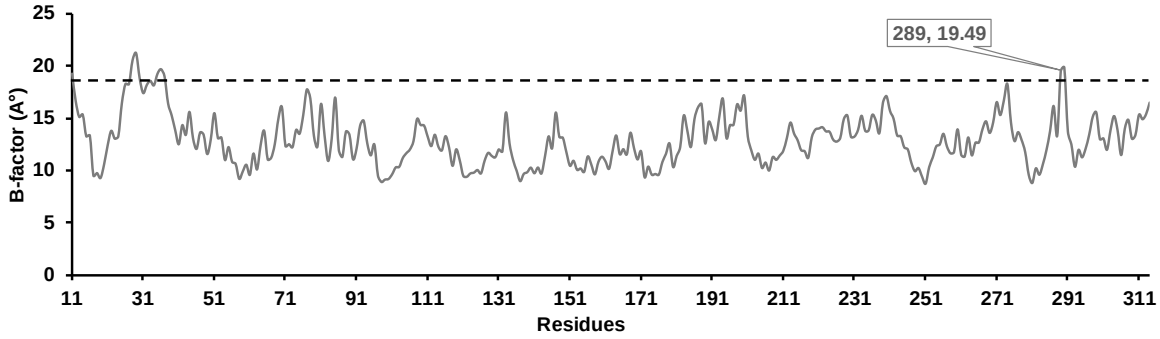
The selection of positions for SSM was based on three criteria:

A: Experimentally proved positions from PvCel5A variants with the highest increase in thermostability detected in Phase I (**Table S5**, Y293, E263, T312, S305, E304, S114, and S308)

B: Computationally identified flexible positions based on B-factor analysis (**Figure S4A**). After B-factor analysis, it was observed that variant Q289G is highly stable (ΔΔG = - 2.328 **Figure S3B**)

C: Positions that offer thermostable PvCel5A variants based on ΔΔG analysis (**Table S7**). After ΔΔG analysis, position H96 presented the highest number of highly stabilizing substitutions (ΔΔG < −1.84), with 2 substitutions. Position T69 presented the highest number of variants in the stabilizing (ΔΔG −1.84 to −0.92) and slightly stabilizing (ΔΔG −0.92 to −0.46) range with 8 substitutions. Therefore, H96 and T69 were selected for further analysis.

A



B

Position	V	Y	W	T	S	P	F	M	K	L	I	H	G	E	Q	C	D	N	R	A
11	0.686	0.126	0.097	0.268	0.000	2.886	0.138	0.135	0.034	0.248	0.680	0.180	-0.785	0.287	0.133	-0.025	0.280	0.105	-0.019	-0.329
28	0.101	-0.404	-0.124	-0.258	-0.319	-0.018	-0.630	-0.445	-0.931	-0.351	-0.106	-0.045	0.000	-0.341	-0.710	-0.147	-0.151	-0.298	-0.622	-0.288
29	0.762	0.328	0.949	1.060	1.246	3.516	0.609	0.897	0.217	0.696	0.164	0.267	1.291	1.078	0.375	0.686	1.470	0.000	0.486	1.011
30	0.966	1.805	1.192	1.852	2.539	3.796	1.745	-0.185	1.028	-0.132	0.000	2.183	2.386	1.420	1.568	1.509	2.100	1.947	1.476	2.008
33	0.000	-0.087	0.279	-0.035	0.388	2.898	-0.124	0.232	-0.608	-0.037	-0.039	-0.362	1.537	0.438	-0.036	0.266	0.267	0.185	-0.547	0.551
35	4.246	3.116	3.238	3.788	3.372	6.002	3.090	2.946	2.740	3.249	4.041	3.362	0.000	3.332	2.856	3.014	3.142	2.880	2.072	3.017
36	-0.372	0.057	0.032	0.000	0.098	3.775	0.008	-0.060	-0.539	-0.299	-0.174	-0.306	0.214	-0.027	-0.748	0.547	0.771	0.075	-1.170	0.436
37	1.875	4.072	3.618	1.665	1.660	1.721	3.707	1.243	1.654	0.879	1.939	1.098	2.333	1.319	1.537	1.667	0.000	1.323	0.881	1.523
289	0.746	0.019	-0.024	0.502	-1.275	3.602	-0.032	0.025	-1.444	0.277	0.668	0.085	-2.328	0.194	0.003	0.050	0.439	0.172	-0.290	0.204
290	0.720	-0.368	-0.019	0.676	0.593	0.153	-0.468	-0.365	0.095	-0.094	0.301	0.158	0.117	0.365	0.102	1.806	-0.020	0.060	-0.362	0.318

Figure S4: Flexibility analysis of PvCel5A. **A:** B-factor for each residue of PvCel5A-WT. **B:** Hit map of *in silico* SSM by ΔΔG prediction. Colour code used, green: highly stabilizing

(< −1.84), light green: stabilizing (−1.84 to −0.92), faint green: slightly stabilizing (−0.92 to −0.46), white: neutral (−0.46 to +0.46), yellow: slightly destabilizing (+0.46 to +0.92), orange: destabilizing (+0.92 to +1.84), and red: highly destabilizing (> +1.84).

The ΔΔG can provide various phenomena in terms of stabilizing or destabilizing property of the *in silico* substitutions. For example, substitutions with proline in various positions contributed to stabilizing (e.g., S27, T77, and S308) or destabilizing (e.g., E263, S305, T69). This indicates that substitutions with proline could provide stabilizing or destabilizing effect, depending on the environment, and interactions with the surrounding the amino acids as observed in previous studies.[23-25] Similar phenomena can be observed for other substitutions (e.g., arginine, tyrosine, and methionine) in different positions (e.g., S305, T69).

Table S7: *In silico* SSM of experimentally identified positions based on thermodynamics stability (ΔΔG*).

Position	V	Y	W	T	S	P	F	M	K	L	I	H	G	E	Q	C	D	N	R	A
F16	4.599	1.274	5.777	4.600	4.593	5.383	-0.011	1.798	5.132	2.112	4.524	4.773	6.699	6.270	5.361	4.077	5.593	4.832	6.767	5.312
Y293	3.543	0.050	6.169	3.765	4.607	5.719	0.397	1.528	3.336	2.363	7.010	3.334	3.932	5.504	3.547	3.323	5.354	4.008	4.052	3.140
I246	0.692	9.535	11.720	2.393	3.672	7.793	6.142	0.193	2.148	0.658	0.000	2.997	5.241	4.404	3.831	3.075	6.146	4.275	0.018	3.427
K130	0.976	1.557	1.978	1.180	1.097	0.744	1.544	0.763	-0.074	1.259	1.262	0.625	1.842	1.893	0.844	1.197	2.942	1.243	-0.205	0.800
E263	-0.311	-1.194	-0.723	0.496	-0.137	3.352	-1.082	-1.662	-1.249	-1.560	-1.102	-0.310	0.575	-0.036	-0.459	0.010	1.724	-0.292	-0.785	-0.224
T312	1.104	0.568	0.571	-0.028	-0.099	1.779	0.494	0.150	-0.095	0.328	0.780	0.218	0.826	0.841	0.062	0.178	1.023	0.304	-0.266	0.237
S305	0.232	-0.894	-0.623	0.769	-0.011	2.510	-1.209	-0.952	-0.430	-0.957	-0.504	-1.229	0.716	0.562	0.056	0.563	1.254	-0.465	-1.082	-0.080
E304	0.233	0.323	-0.257	-0.235	-0.031	3.571	0.299	-0.526	-0.516	-0.401	-0.252	0.258	0.611	0.000	-0.238	0.517	0.578	0.514	-0.480	-0.050
T69	0.053	-0.691	-0.715	-0.004	0.014	3.126	-0.687	-0.653	-0.526	-0.305	-0.190	-0.269	-0.676	-0.082	-0.727	-0.099	-0.340	-0.121	-0.982	-0.084
S114	1.507	1.340	1.365	0.278	0.000	1.503	0.935	1.596	1.484	1.716	1.005	1.588	1.362	-0.605	1.063	1.498	-3.358	-0.209	1.705	1.262
S308	-0.295	-0.188	-0.319	-0.053	0.000	-1.663	-0.324	-0.422	-0.908	-0.486	-0.630	0.339	0.230	-0.023	-0.518	0.259	0.220	0.180	-0.781	0.217
N299	1.701	0.149	0.194	1.613	-0.169	1.666	0.139	1.055	-0.165	0.084	2.465	0.139	-0.070	0.107	0.058	1.287	0.560	-0.001	1.029	0.143
Q289	0.746	0.020	-0.024	0.503	-1.276	3.603	-0.032	0.025	-1.444	0.277	0.668	0.085	-2.328	0.194	0.003	0.050	0.439	0.172	-0.290	0.204
S240	0.592	-0.221	-0.352	-0.153	0.003	2.770	-0.265	-0.445	-0.861	-0.295	0.021	-0.662	0.717	0.295	-0.319	0.618	0.808	-0.203	-0.739	-0.130
S27	0.325	-0.375	0.255	0.299	0.000	-1.110	-0.449	-0.851	-0.331	-0.774	-0.055	0.732	0.631	0.059	-0.563	-0.116	0.355	0.498	-1.154	0.084
H96	0.607	-2.327	-0.577	1.344	0.890	0.813	-2.701	-1.188	0.138	-0.295	0.927	-0.059	2.157	0.787	0.482	0.279	2.436	0.377	0.213	0.300
S131	0.378	-0.071	0.205	0.297	0.000	0.901	-0.219	-0.437	-0.099	0.007	0.163	0.151	0.229	0.226	0.114	0.294	0.046	0.094	-0.290	0.294
T77	-0.036	-0.061	-0.710	0.000	0.176	-1.688	-0.111	-0.984	-0.337	-0.552	-0.521	0.232	0.124	-0.759	-0.453	0.294	-0.716	-0.038	-0.335	-0.126

*Color code used green: highly stabilizing (< −1.84), light green: stabilizing (−1.84 to −0.92), faint green: slightly stabilizing (−0.92 to −0.46), white: neutral (−0.46 to +0.46), yellow: slightly destabilizing (+0.46 to +0.92), orange: destabilizing (+0.92 to +1.84), and red: highly destabilizing (> +1.84).[5]

Substitutions identified from SSM libraries from phase II of the KnowVolution campaign

The screening of the 10 SSM libraries, revealed six positions (Q289, Y293, S305, T312, E304, and S308) that showed an improvement of the thermostability compared to PvCel5A-WT (**Table S8**)

Table S8: Substitutions that increase the thermostability of PvCel5A. The substitutions were identified from SSM libraries from phase II of the KnowVolution campaign.

Variant	Substitution	Times sequenced	Improvement
S3	Q289G	5	2.4
V2	Y293F	5	2.4
V5	S305F	2	2.2
S4	T312R	2	2.1
S5	E304R	2	1.7
S6	E304K	1	1.4
V11	S308P	2	1.2

Recombination variants (R1 to R20) of PvCel5A in phase IV of KnowVolution campaign

After the screening of SSM libraries in Phase II of the KnowVolution campaign, it was shown that 6 substitutions (Y293F, T312R, S305F, E304R, S308P, and Q289G) influence the thermostability of PvCel5A. Recombination was performed *in silico* following two strategies:

A: The generation of double PvCel5A variants by recombining all six positions, variants R1 to R15 in **Table S9**.

B: The generation of triple PvCel5A variants by recombining variant V1 (F16L, and Y293F) with the five remaining positions (variants R16 to R20). Variants are considered energetically stable if their ΔΔG < 0.4 and unpredictable if the ΔΔG < 7.4.

Table S9: Recombination of PvCel5A variants from phase IV of the KnowVolution campaign.

Variant	Substitutions			Stability $\Delta\Delta G_{fold}$	Stability prediction	Improvement fold
R1		Y293F	S305F	-0.658	Stable	3.7
R2		Y293F	Q289G	-2.049	Stable	4.3
R3		Y293F	T312R	-0.048	Stable	3.1
R4		Y293F	S308P	-1.434	Stable	3.0
R6		S305F	Q289G	-3.208	Stable	3.2
R7		T312R	S305F	-1.091	Stable	2.2
R8		S305F	S308P	-2.559	Stable	2.6
R9		S305F	E304R	-1.169	Stable	3.1
R10		T312R	Q289G	-2.698	Stable	3.5
R11		S308P	Q289G	-3.965	Stable	3.8
R12		E304R	Q289G	-2.717	Stable	4.5
R13		T312R	S308P	-2.311	Stable	2.4
R14		T312R	E304R	-0.763	Stable	3.3
R15		E304R	S308P	-1.875	Stable	2.6
R16	F16L	Y293F	S305F	1.349	Unpredictable	3.4
R17	F16L	Y293F	Q289G	0.055	Stable	4.3
R18	F16L	Y293F	T312R	1.913	Unpredictable	3.6
R19	F16L	Y293F	S308P	0.636	Unpredictable	3.9
R20	F16L	Y293F	E304R	1.781	Unpredictable	3.3

Residual activity of the identified PvCel5A variants in the KnowVolution campaign

The relative activity of PvCel5A-WT, variants V1 (F16L/Y293F), R2 (Q289G/Y293F), R12 (Q289G/E304R), and R17 (F16L/Y293F/Q289G) were assessed in *P. pastoris* supernatant (**Table S10**). After a flask expression (see Experimental Section), a standardized amount of protein (70 ng/reaction) was utilized to discard improvement due to increased expression.

Table S10. Activity and thermal stability of PvCel5A-WT and variants V1, R2, R12, and R17. The relative initial activity of PvCel5A variants (V1, R2, R12, and R17) is compared to PvCel5A-WT. Residual activity of PvCel5A-WT and variants was determined after incubation at 80 °C for 60 min and the calculated improvement of the thermal stability of each variant when compared with PvCel5A-WT.

Variant	Substitution	Relative initial activity [%]	Residual activity [%]	Thermal stability improvement fold
WT	-	100	26	1.0
V1	F16L/Y293F	100	65	2.5
R12	Q289G/E304R	92	77	3.0
R2	Q289G/Y293F	87	80	3.1
R17	F16L/Y293F/Q289G	84	94	3.6

Specific activity of purified PvCel5A WT, variants V1 and R17

The specific activity of PvCel5A WT, variants V1 and R17 was determined with the natural substrate carboxymethyl cellulose (CMC) at different temperatures to determine the optimal temperature of PvCel5A (**Table S11**). PvCel5A-WT retained up to 57% of its activity when assessed at 40 °C, while variants R17 and V1 retained 28 and 37%, respectively. Variant R17 had increased specific activity in temperatures between 50 and 90 °C compared to PvCel5A-WT, while variant V1 had increased specific activity in the temperature range of 65 to 80 °C (**Figure S4**).

Table S11: Specific activity of PvCel5A-WT, variants V1 and R17. Specific activity in the temperature range of 40 to 90 °C. Reported values were averaged from three independent measurements performed in technical triplicates.

Temperature [°C]	Specific activity [U/mg]		
	WT	V1	R17
90	66 ± 8	66 ± 2	127 ± 5
80	199 ± 9	259 ± 11	403 ± 18
75	221 ± 2	288 ± 8	427 ± 9
70	218 ± 4	265 ± 9	397 ± 10
60	184 ± 4	176 ± 4	219 ± 21
50	132 ± 4	136 ± 5	179 ± 10
40	114 ± 4	106 ± 11	120 ± 8

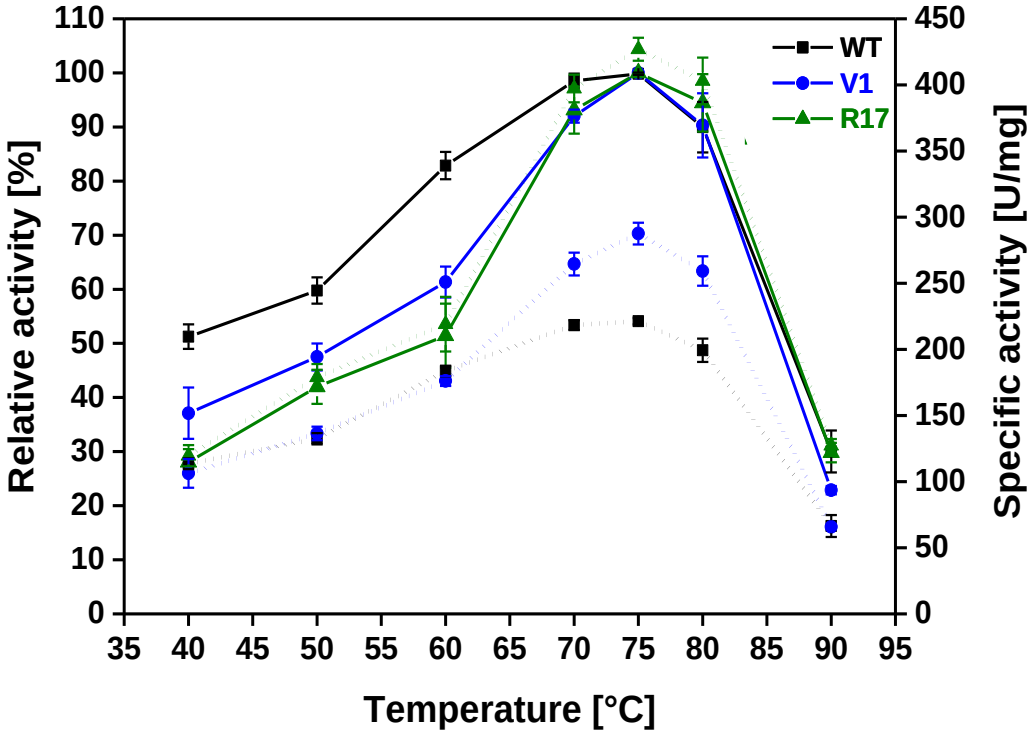


Figure S5: Relative activity (full lines) and specific activity (dotted line) of PvCel5A-WT, variants V1 and R17. In the relative activity, the maximum activity of PvCel5A-WT, variants V1 and R17 was considered as 100%. The activities are measured in the temperature range of 40 to 90 °C. The specific activity is presented in dotted lines.

Evolutionary conservation analysis of PvCel5A

The evolutionary analysis of PvCel5A (PDB ID: 5L9C) was generated using a multiple sequence alignment containing 150 sequences from the protein database UniRef90[26] utilizing the ConSurf server.[27]

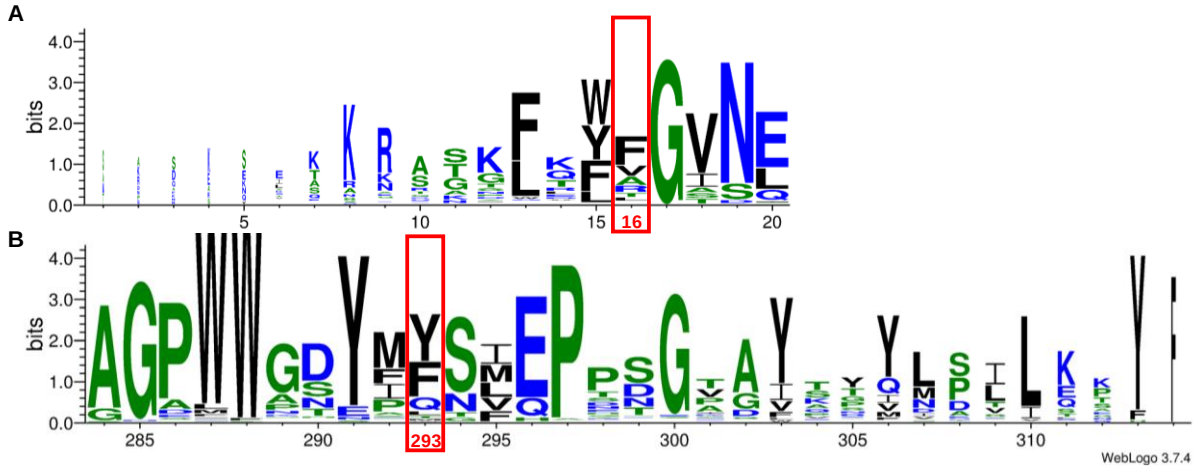


Figure S6. Evolutionary conservation analysis of PvCel5A. Sequence logo generated to visualize conservation analysis of **A:** N-terminal (residues 1-20) and **B:** C-terminal (residues 284-314) region of PvCel5A (PDB ID: 5L9C). The sequence logo was generated using the WebLogo tool.[28] Positions 16 and 293 are highlighted in red.

Fluctuation of the C-terminal region of PvCel5A-WT, variants V1 and R17

Root mean square fluctuation (RMSF) values from MD simulations of each C-terminal region (backbone of amino acids 280-314) at 27 °C, 70 °C, 75 °C, and 80 °C.

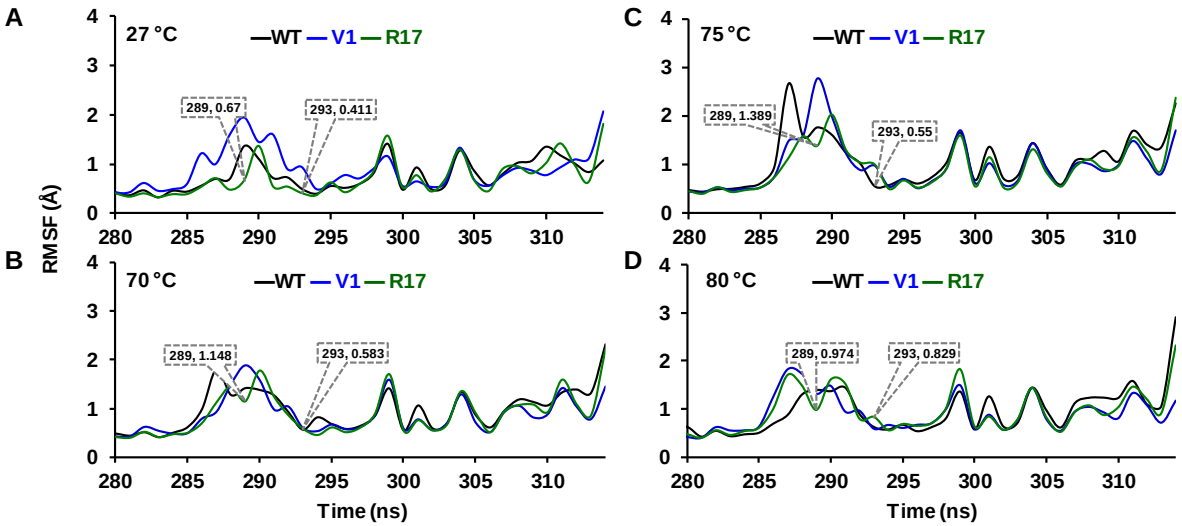


Figure S7. RMSF features of PvCel5A WT, variant V1 (Y293F/Q289G), and R17 (F16L/Y293F/Q289G) of each C-terminal region (backbone of amino acids 280-314) at 27°C (**A**), 70 °C (**B**), 75 °C (**C**), and 80 °C (**D**). Q289G played the main role to reduce RMSF, whereas Y293F substitution slightly distorted the region.

Fluctuation of PvCel5A-WT, variants V1 and R17

Root mean square fluctuation (RMSF) values from MD simulations of PvCel5A, variants V1 and R17

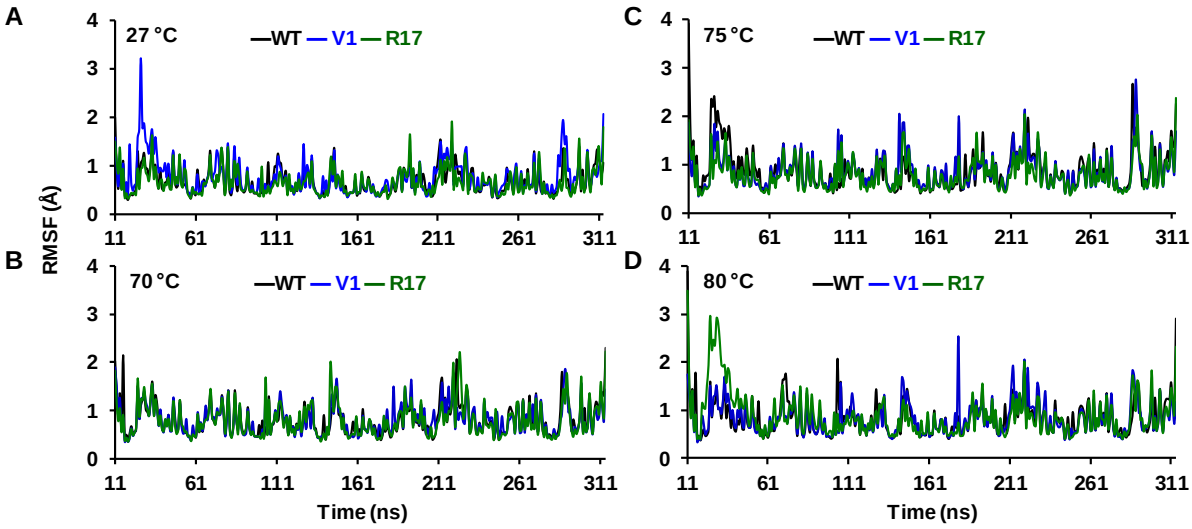


Figure S8. RMSF of PvCel5A WT, variant V1 (Y293F/Q289G), and R17 (F16L/Y293F/Q289G) at 27°C (**A**), 70 °C (**B**), 75 °C (**C**), and 80 °C (**D**). RMSF values remain unchanged in the active site residues E142 and E249, which indicate stability (flexibility and rigidity) was maintained in the active site during MD simulations at 27°C (**A**), 70 °C (**B**), 75 °C (**C**), and 80 °C (**D**).

Cellulolytic cocktail performance on microcrystalline cellulose (MCC) degradation

Degradation of microcrystalline cellulose (MCC) was assayed with a reconstituted cellulase cocktail of PvCel5A, HjCBHI, and AnBGL. MCC was pre-treated at 70 °C for 16 h with the endoglucanase and later treated with the HjCBHI, and AnBGL at 50 °C.

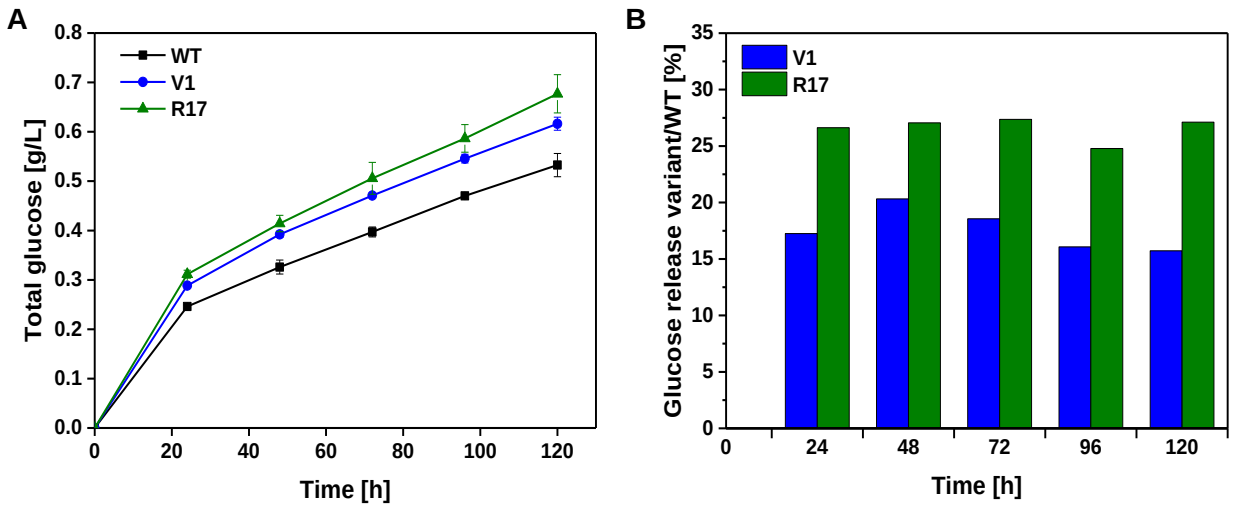


Figure S9. Degradation of microcrystalline cellulose (MCC) with a cellulolytic cocktail. A cellulolytic cocktail containing PvCel5A, HjCBHI, and AnBGL was utilized for the degradation of MCC 1.0% in sodium acetate buffer pH 4.5 50 mM. MCC was pre-incubated with PvCel5A-WT, variant V1, or R17 at 50 °C for 16 h at 70 °C. Later HjCBHI and AnBGL were added to the PvCel5A containing MCC and incubated at 50 °C. Glucose concentration was determined every 24 h after the addition of HjCBHI and AnBGL.

Visualization of identified amino acid positions that influence thermal stability in PvCel5A-WT endoglucanase

Representation of identified positions (Q289, E304, Y293, and F16) after the KnowVolution campaign of PvCel5A-WT towards improved thermostability. PvCel5A is represented in two views for better visualization of the substitutions.

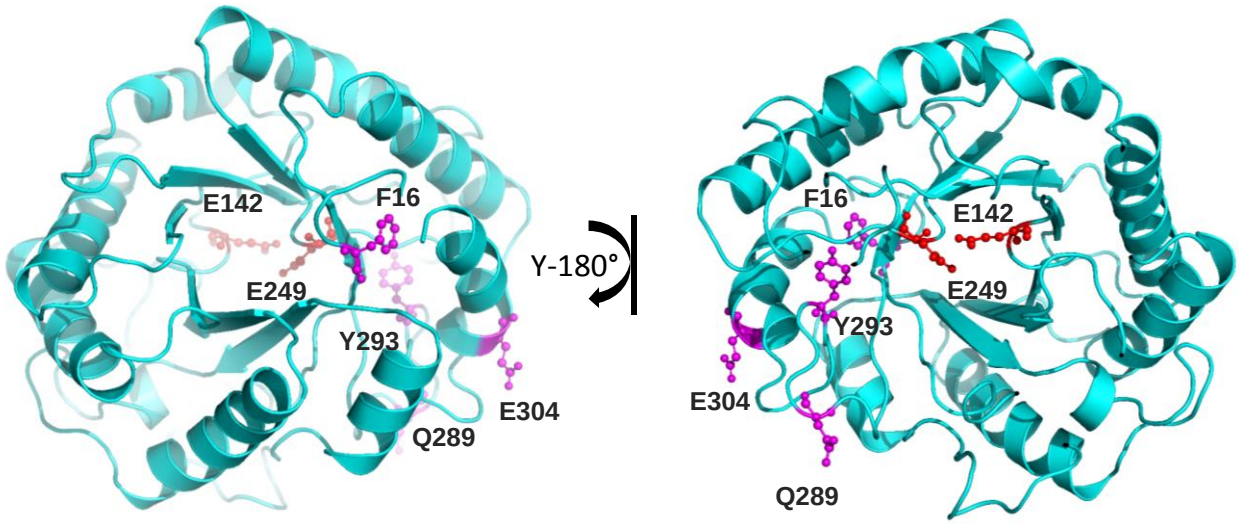


Figure S10. The location of confirmed beneficial positions of the variant R2, R12, and R17 obtained from Phase IV (see **Figure 1**) are shown. PvCel5A structure is shown in cyan, identified positions, and catalytic residues (E142 and E149) are shown as sticks in magenta, and red, respectively.

Contribution of each type of energy to the total energy of substitution Y293F

It shows that contribution of H-bond (backbone and sidechain H-bonds) energy is destabilizing, whereas hydrophobic energy (solvation hydrophobic) plays a significant role towards stabilization. Besides, solvation polar energy played a major role in the stabilization of substitution Y293F. Unfavorable H-bond is compensated by solvation interactions (solvation polar and solvation hydrophobic).

Table S12: Contribution of each energy to the total energy of Y293F (see **Table S7**) using FoldX.

Energy	**Value**
Total energy (ΔΔG)	**0.396882**
Backbone Hbond	0.623299
Sidechain Hbond	0.632599
Van der Waals	0.235589
Electrostatics	-0.007626
Solvation Polar	**-0.777493**
Solvation Hydrophobic	**-0.160492**
Van der Waals clashes	-0.004343
entropy sidechain	-0.090906
entropy mainchain	0.091456
sloop_entropy	0
mloop_entropy	0
cis_bond	0
torsional clash	-0.145201
helix dipole	0
water bridge	0
disulfide	0
electrostatic kon	0
partial covalent bonds	0
energy Ionisation	0
Entropy Complex	0

References

1. Miyazaki, K., MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. In *Methods Enzymol.*, Elsevier: 2011; Vol. 498, pp 399-406.
2. Morozova, V. V.; Gusakov, A. V.; Andrianov, R. M.; Pravilnikov, A. G.; Osipov, D. O.; Sinitsyn, A. P., Cellulases of Penicillium verruculosum. *Biotechnol. J.* **2010,** *5* (8), 871-880.
3. Krainer, F. W.; Gmeiner, C.; Neutsch, L.; Windwarder, M.; Pletzenauer, R.; Herwig, C.; Altmann, F.; Glieder, A.; Spadiut, O., Knockout of an endogenous mannosyltransferase increases the homogeneity of glycoproteins produced in Pichia pastoris. *Sci. Rep.* **2013,** *3*, 3279.
4. Mellitzer, A.; Weis, R.; Glieder, A.; Flicker, K., Expression of lignocellulolytic enzymes in Pichia pastoris. *Microb. Cell Fact.* **2012,** *11* (1), 61.
5. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, Springer: 2005; pp 571-607.
6. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K., Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins: Struct. Funct. Bioinform.* **2009,** *77* (S9), 114-122.
7. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L., The FoldX web server: an online force field. *Nucleic Acids Res.* **2005,** *33* (suppl_2), W382-W388.
8. Van Durme, J.; Delgado, J.; Stricher, F.; Serrano, L.; Schymkowitz, J.; Rousseau, F., A graphical interface for the FoldX forcefield. *Bioinformatics* **2011,** *27* (12), 1711-1712.
9. Lindahl, E.; Hess, B.; Van Der Spoel, D., GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001,** *7* (8), 306-317.
10. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16), 1701-1718.
11. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.
12. Kusalik, P. G.; Svishchev, I. M., The spatial structure in liquid water. *Science* **1994,** *265* (5176), 1219-1221.
13. Fyta, M.; Netz, R. R., Ionic force field optimization based on single-ion and ion-pair solvation properties: Going beyond standard mixing rules. *J Chem. Phys.* **2012,** *136* (12), 124103-124111.
14. Wang, J.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21* (12), 1049-1074.
15. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** *32* (suppl_2), W665-W667.
16. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007,** *35* (suppl_2), W522-W525.
17. Onufriev, A. V.; Alexov, E., Protonation and pK changes in protein-ligand binding. *Q. Rev. Biophys.* **2013,** *46* (2), 181-209.
18. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J Chem. Phys.* **1995,** *103* (19), 8577-8593.
19. Norberg, J.; Nilsson, L., On the Truncation of Long-Range Electrostatic Interactions in DNA. *Biophys. J.* **2000,** *79* (3), 1537-1553.
20. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-38.

21. Cheng, F.; Zhu, L.; Schwaneberg, U., Directed evolution 2.0: improving and deciphering enzyme properties. *Chem. Commun.* **2015,** *51* (48), 9760-9772.
22. Cui, H.; Cao, H.; Cai, H.; Jaeger, K. E.; Davari, M. D.; Schwaneberg, U., Computer-assisted Recombination (CompassR) teaches us how to recombine beneficial substitutions from directed evolution campaigns. *Chem. Eur. J.* **2019,** *26* (3), 643-649.
23. Vieille, C.; Zeikus, G. J., Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001,** *65* (1), 1-43.
24. Dotsenko, A. S.; Pramanik, S.; Gusakov, A. V.; Rozhkova, A. M.; Zorov, I. N.; Sinitsyn, A. P.; Davari, M. D.; Schwaneberg, U., Critical effect of proline on thermostability of endoglucanase II from Penicillium verruculosum. *Biochem. Eng. J.* **2019,** *152*, 107395.
25. Wang, K.; Luo, H.; Tian, J.; Turunen, O.; Huang, H.; Shi, P.; Hua, H.; Wang, C.; Wang, S.; Yao, B., Thermostability improvement of a Streptomyces xylanase by introducing proline and glutamic acid residues. *Appl. Environ. Microbiol.* **2014,** *80* (7), 2158-2165.
26. Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H.; Consortium, U., UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015,** *31* (6), 926-932.
27. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N., ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **2016,** *44* (W1), W344-W350.
28. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res.* **2004,** *14* (6), 1188-1190.