Functional Characterization and Low-Resolution Structure of an Endoglucanase Cel45A from the Filamentous Fungus Neurospora crassa OR74A: Thermostable Enzyme with High Activity Towa...

ARTICLE *in* MOLECULAR BIOTECHNOLOGY · FEBRUARY 2015

Impact Factor: 1.88 · DOI: 10.1007/s12033-015-9851-8 · Source: PubMed

4 AUTHORS:

Marco Kadowaki

University of São Paulo

9 PUBLICATIONS **22** CITATIONS

SEE PROFILE

Cesar Camilo

Centro de Tecnologia Canavieira

12 PUBLICATIONS **41** CITATIONS

SEE PROFILE

Amanda Bernardes

University of São Paulo

17 PUBLICATIONS **90** CITATIONS

SEE PROFILE

Igor Polikarpov

University of São Paulo

263 PUBLICATIONS **5,101** CITATIONS

SEE PROFILE

Mol Biotechnol
DOI 10.1007/s12033-015-9851-8

RESEARCH

Functional Characterization and Low-Resolution Structure of an Endoglucanase Cel45A from the Filamentous Fungus *Neurospora crassa* OR74A: Thermostable Enzyme with High Activity Toward Lichenan and β-Glucan

Marco Antonio Seiki Kadowaki · Cesar Moises Camilo · Amanda Bernardes Muniz · Igor Polikarpov

Abstract Biomass is the most abundant and short-term renewable natural resource on Earth whose recalcitrance toward enzymatic degradation represents significant challenge for a number of biotechnological applications. The not so abundant but critically necessary class of GH45 endoglucanases constitutes an essential component of tailored industrial enzyme cocktails because they randomly and internally cleave cellulose molecules. Moreover, GH45 glucanases are core constituents of major-brand detergent formulations as well as enzymatic aid components in the cotton processing industry, clipping unwanted cellulosic fibers from cotton (cellulosic)-based tissues. Here we report on a recombinant high-yield *Neurospora crassa* OR74A NcCel45A production system, a single-band GH45 endoglucanase purification, and a complete enzyme functional characterization. NcCel45A is a bimodular endoglucanase showing maximum activity at pH 6.0 and 60 °C, while most active against lichenan and β-glucans and lesser active toward filter paper, carboxymethylcellulose, and phosphoric acid-swollen cellulose. Gluco-oligosaccharide degradation fingerprinting experiments suggest cellopentaose as the minimal length substrate and ThermalFluor studies indicate that NcCel45A displays excellent stability at elevated temperatures up to 70 °C and pHs ranging from 5 to 9. Remarkably, we show that NcCel45A is uniquely resistant to a wide-range of organic solvents and small-angle X-ray scattering show a monkey-wrench molecular shape structure in solution, which indicates, unlike to other known cellulases, a non-fully extended conformation, thus conferring solvent protection. These NcCel45A unique enzymatic properties maybe key for specific industrial applications such as cotton fiber processing and detergent formulations.

Keywords Glycoside hydrolases · GH45 · Endoglucanase · *Neurospora crassa* · Thermostability

Abbreviations

CBM	Carbohydrate-binding module
CMC	Carboxymethyl cellulose
DAM	Dummy atom models
DNS	Dinitrosalycylic acid
D_{max}	Maximum particle size
EDTA	Ethylenediaminetetraacetic acid
GH	Glycoside hydrolase
HCCA	Alpha-cyano-4-hydroxy cinnamic acid
PASC	Phosphoric acid-swollen cellulose
PDB	Protein Data Bank
R_g	Radius of gyration
pNPG	*p*-Nitrophenyl-β-D-glucoside
Rpm	Rotations per minute
SAXS	Small-angle X-ray scattering
T_m	Melting temperature

Introduction

Lignocellulosic biomass is the most abundant biomaterial produced on Earth which is predominantly found in plant cell walls [1]. Cellulose, composed by D-anhydro-glucopyranose units joined together into a linear high-degree

Electronic supplementary material The online version of this article (doi:10.1007/s12033-015-9851-8) contains supplementary material, which is available to authorized users.

M. A. S. Kadowaki · C. M. Camilo · A. B. Muniz · I. Polikarpov (✉)
Instituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador São Carlense 400, São Carlos, SP 13566-590, Brazil
e-mail: ipolikarpov@ifsc.usp.br

Published online: 26 February 2015

polymerization fiber through β-1,4-glycosidic linkages represent the major component (30 %) of a typical plant cell wall [2]. Cellulose crystallinity is due to extensive intermolecular fiber hydrogen bonding that results in a highly stable molecular aggregate recalcitrant toward natural degradation by single-source microorganisms, such as fungi and/or bacteria. However, cellulose recalcitrance toward degradation is overcome by fungi that produce collectively endoglucanases (E.C. 3.2.1.4), exoglucanases/cellobiohydrolases (E.C. 3.2.1.91 and 3.2.1.176), β-glucosidases (E.C. 3.2.1.21) as well as some auxiliary enzymes [3–5] that have the potential to be applied in biomass-to-biofuel technological processes [6, 7].

Endoglucanases, the enzymes responsible for endo hydrolysis of β-1,4-glucosidic linkages in cellulose, have been classified according to Carbohydrate-Active enZYmes (CAZY) into glycoside hydrolase (GH) families: GH5, GH6, GH7, GH8, GH9, GH10, GH12, GH26, GH44, GH45, GH48, GH51, GH74, and GH124 [8]. GH45 endoglucanases are being widely used in laundry detergents due to their neutral optimal pHs and high activities [9–11]. These enzymes are also known for its low molecular weight and inverting stereo chemical mechanism of action [12]. Based on phylogenetic analysis, fungal GH45 enzymes cluster into three subfamilies that also include expansions and swollenins [13, 14]. Interestingly, some GH45 genes are found in coleopteran species and molecular evolution studies suggest horizontal gene transfer and a relationship with fungi [15, 16]. At present, this family contains 329 different members in CAZY [8], with 15 bacterial, 307 eukaryotic, and 7 unclassified genes. Crystal structures are available for only four proteins: the ones from *Humicola grisea* var. *thermoidea* (PDBid: 1HD5) and *Humicola insolens* DSM 1800 [17] as well as *Melanocarpus albomyces* [18] and *Mytilus edulis* (PDBid: 1WC2; unpublished).

Neurospora crassa is a filamentous fungus that has been widely studied as a model organism for eukaryotic biology [19] and is found in Nature growing on carbohydrate-rich foodstuffs such as residues of sugarcane processing [19]. The genome sequence [20] of *N. crassa* makes it a potential cellulase, hemicellulase, and a variety of oxidases and laccases [21–23] producer.

Although several fungal GH45 enzymes have already been characterized biochemically [9, 13, 14, 24–29], more biochemical and structural information about this family of GHs is necessary, particularly on the multidomain GH45s organization. So far, only a handful of multidomain GH45 s, including Mce1-2 from *Mucor circinelloides* [30], EGL4-CBD from *H. grisea* [31], and 5(CBM)-PpCel45A from *Pichia pastoris* [13] have been studied biochemically.

Here, we report on the biochemical characterization and low-resolution structural model of endoglucanase GH45 from *N. crassa* OR74A (NCU05121). The NCU05121 gene was cloned into *Aspergillus nidulans* with heterologous over-expression/secretion and NcCel45A was purified. Biochemical characterization of NcCel45A shows that the enzyme has an optimum pH and temperature at 6.0 and 60 °C, respectively. Furthermore, the enzyme is highly stable up to a temperature of 70 °C and its activity is enhanced by the presence of organic solvents (such as ethanol, for example) and divalent ions (for example Ca^{2+} and Co^{2+}). The enzyme specificity and a pattern of oligosaccharide hydrolysis by NcCel45A have also been determined. Finally, we obtained the small-angle X-ray scattering (SAXS)-based molecular envelope of NcCel45A in solution and compared it with other studied cellulases.

Materials and Methods

Sequence Analysis

The uncharacterized gene NCU05121 (GenBank: XM_952014.1) from *N. crassa* OR74A genome was identified, compared, and analyzed using BLAST tools [32] from National Center for Biotechnology Information database. The amino acid sequence alignment was performed with T-Coffee [33] and the graphic generated by ESPript 3.0 [34]. The biochemical properties of NcCel45A were determined by ProtParam tools [35]. The signal peptide position was determined using the SignalP 4.1 [36]. Potential *O*- and *N*-glycosylation sites were predicted using NetOGlyc 4.0 [37] and NetNGlyc 1.0 on ExPASy [38].

Materials, Microorganism Strains, and Culture Conditions

All chemicals were Molecular Biology grade and purchased from Merck (Germany), Sigma-Aldrich (USA), J.T.Baker (Netherlands), or Thermo Scientific (USA). Oligonucleotides were purchased from Sigma (USA). The high maltose corn sirup was purchased from Cargill (Brazil). The substrates Avicel PH-101, carboxymethylcellulose (CMC), *p*-nitrophenyl-β-D-glucoside (pNPG), cellopentaose, and cellotetraose were obtained from Sigma-Aldrich (USA); arabinan, galactomannan, barley β-glucan, lichenan, 1,4-β-D-manan, xyloglucan and cellohexaose were obtained from Megazyme (Ireland). Phosphoric acid-swollen cellulose (PASC) was prepared from Avicel as described [39].

Escherichia coli DH5a (Invitrogen, USA) was used for cloning and plasmid propagation. *A. nidulans* strain A773 (*pyrG89*; *wA3*; *pyroA4*) and *N. crassa* OR74A (wild type) were obtained from the Fungal Genetic Stock Center (University of Missouri, Kansas City, Missouri).

A. nidulans A773 was propagated in minimal medium supplemented with uracil, uridine, and pyridoxine at 37 °C [40], and *N. crassa* OR74A was grown in Volgel's medium at 30 °C as described [41].

Total RNA Isolation and cDNA Synthesis from *N. crassa* OR74A

10^8 conidia/mL from 10-day-old petri dishes were collected and inoculated in 200 mL of Vogel's medium supplemented with 1.5 % sucrose, 3 % Avicel or 3 % sugarcane bagasse (w/v) as carbon sources to induce cellulase expression. The culture was incubated at 30 °C and 150 rpm for 7 days. After NcCel45A expression induction, the mycelium was separated from culture medium by filtration, washed twice with water, and quickly frozen using liquid nitrogen for RNA extraction. RNA isolation was carried out using the TRIzol reagent (Invitrogen) and the first strand cDNA synthesized using ProtoScript Kit (Neb) and stored at −80 °C.

Cloning of NcCel45A from *N. crassa* OR74A

The target gene NCU05121 was cloned into pEXPYR vector [40], which was adapted for a Ligation-Independent Cloning protocol (LIC) [42, 43]. The vector was modified by a two-step PCR amplification that inserted LIC cloning sites and a sequence encoding a polyhistidine tag to be expressed fused to the N-terminal of target proteins. Briefly, the pEXPYR plasmid was used as template for the first amplification using primers pEXPYR/LIC_Fw1 (5′-GATATCGTAATCGTGATGGTGATGGTGATGGTAGT ACGAGCGCTTGGAAATCACATTTG-3′) and pEXPY R/LIC_Rv (5′-CCGCGTCGGGTCAAACGCTCTAGAGA ACAAAAACTCATC-3′). The PCR product was purified and used as a template for the second amplification using primers pEXPYR/LIC_Fw2 (5′-TGGCGCCCTGAAAA-TACAGGTTTTCGGTCGTTGGGATATCGTAATCGTG ATGG-3′) and pEXPYR/LIC_Rv2 (5′-CCGCGTCGGGT CA-3′). The second PCR product was further used for LIC cloning procedure as described [43].

The NCU05121 gene from *N. crassa* (Uniprot Q872Q1) was PCR amplified from total cDNA without the original signal peptide. The PCR product comprising 64-882 bp from the original sequence was amplified using the oligonucleotide primers NcCel45A-Fw (5′-CAGGGCGCC ATGGCTTCCGGATCCGG-3′) and NcCel45A-Rv (5′-GA CCCCGACGCGGTCAGGCACACTGATGGTAATA-3′) with LIC extensions (underlined) and cDNA as template. The clone with the correct DNA sequence was confirmed by sequencing and transformed in *A. nidulans* A773 as described [40].

Heterologous Expression/Secretion and Protein Purification

Approximately 10^7–10^8 spores/mL were inoculated in liquid minimal medium supplemented with 3 % high maltose corn sirup (45 % maltose, 35 % glucose) and maintained in static culture at 37 °C for 2 days. The mycelial mat was carefully removed and the culture medium filtrated using Miracloth membrane (Biocompare). The secreted proteins were concentrated twenty times and washed twice with 150 mM NaCl and 20 mM Tris–HCl pH 8.0 by tangential flow filtration (Hollow fiber, GE HealthCare). The protein solution was applied to a DEAE-Sephadex column and the flow-thought collected and concentrated by ultrafiltration (10,000 Da cutoff Centricon). The purification was completed using a size exclusion chromatography on HiLoad 16/60 Sephadex75 column (GE Healthcare) with a running buffer consisting of 150 mM NaCl and 20 mM Tris–HCl pH 8.0. The total protein was quantified by the Bradford method [44] and the purified protein spectrophotometrically at 280 nm using a molar extinction coefficient. The protein purity was analyzed by SDS-PAGE [45] stained with Coomassie blue G-250 and the protein identity was confirmed by mass spectrometry.

Mass Spectrometry MALDI-TOF of Purified NcCel45A

The protein band was removed from the SDS-PAGE 15 % gel and was submitted to in-gel trypsin (20 ng/μL) digestion [46]. An aliquot of digested product (1 μL) was mixed with 1 μL of alpha-cyano-4-hydroxy cinnamic acid (HCCA) matrix at 10 mg/mL and allowed to dry over the sample plate. The sample was analyzed using the mass spectrometer Microflex LT MALDI-TOF (Bruker Daltonics). The measurement was done in the linear positive-ion mode within a range of 400–3300 m/z. Average masses were assigned and processed using flexAnalysisTM software (Bruker Daltonics). The mass fingerprint search was done using BioToolsTM (Bruker Daltonics) and the peptide m/z list generated in silico by computational tryptic digestion of protein amino acid sequence. The peptide-fingerprinting match was also performed using the MASCOT (Matrix Science Inc.).

Thermal Stability Analysis Using ThermoFluor Assay

ThermoFluor assay was applied to assess the best conditions for NcCel45A structural integrity, as a function of pH and buffer type. Several buffers that are likely to stabilize a

majority of the proteins were included in the buffer screen matrix. It consisted of a set of 30 different buffers (Additional file 1: Table S1), each at a concentration of 50 mM and with a pH ranging from 3.0 to 10.0.

Solutions of 13 μM of the enzyme, diluted in each buffer and added of SYPRO Orange dye (Molecular Probes) to a final concentration of 1/2000 were inserted to the wells of a 96-well thin-wall PCR plate (Bio-Rad). The plates were sealed with Optical-Quality Sealing Tape (Bio-Rad) and heated in an iCycler iQ Real-Time PCR Detection System (Bio-Rad).

The temperature range was from 25 to 90 °C, with stepwise increments of 1 °C per minute and 30 s hold step for every point, followed by the fluorescence reading with excitation/emission wavelengths at 490/530 nm. Analyses of the curves, including the melting temperature (T_m) determination, were carried out using GraphPad Prism software (version 5.0).

Enzymatic Assays

Enzymatic activity was assayed by measuring the release of reducing sugars using 3,5-dinitrosalicylic acid (DNS) [47]. One unit (U) of enzyme activity was defined as the amount of enzyme that catalyzes the liberation of 1.0 μmol glucose/min. All experiments were performed in triplicate.

Hydrolytic specific activity was evaluated in various substrates at 0.5 % (w/v), including lichenan, β-glucan, filter paper, CMC, PASC, xyloglucan, avicel, 1,4-β-D-manan, galactomanan, arabinan, and pNPG.

The optimal temperature and pH were determined under standard conditions using barley β-glucan as substrate. The pH assay was performed using citrate-glycine-phosphate buffer. The effect of various metal ions, solvents, and surfactants on activity of the enzyme was determined incubating 1 μM of enzyme and 0.5 % barley β-glucan for 1 h at 60 °C.

Thermostability of the purified enzyme was monitored by pre-incubating the enzyme under standard conditions at different times at 70, 80, and 90 °C.

The cellohexaose, cellopentaose, and cellotetraose hydrolysis by NcCel45A was analyzed by MALDI-TOF MS. Substrates at 1 mM concentration were incubated with 1 μM of NcCel45A in 20 mM ammonium formate buffer (pH 5.0) for 15 h at 60 °C in a total volume of 100 μL. The reaction products were analyzed by MALDI-TOF mass spectrometry in positive-ion mode on a Microflex LT MALDI-TOF (Bruker Daltonics). Samples were supplemented with NaCl to a final concentration of 20 mM and mixed with 2,5-dihydroxybenzoic acid (10 mg/mL) (1:1) and spotted on the target plate. The spectra were generated by 300 laser shots at 60 % potency.

Small-Angle X-ray Scattering (SAXS) Data Collection and Analysis

NcCel45A samples were measured at concentrations of 0.44 and 0.88 mg/mL and 25 °C on the D02A-SAXS2 beamline at the Synchrotron Light National Laboratory (LNLS, Campinas, Brazil). The measurements were performed using a monochromatic X-ray beam with a wavelength of $\lambda = 1.488$ Å and the X-ray scattering patterns were recorded using a two-dimensional CCD detector (MarResearch, USA). The sample to detector distance was 955 mm, providing a scattering vector q interval from 0.008 to 0.34 $Å^{-1}$, where q-vector is defined by $q = 4\pi/\lambda \sin\theta$ (2θ is the scattering angle). The samples in 20 mM Tris–HCl buffer pH 8.0 and 150 mM NaCl were centrifuged for 30 min at 23,500×g and 4 °C prior to measurements to remove any aggregates. Each sample was recorded twice for 300 s to monitor radiation damage. The 2D scattering data were corrected for the detector response and scaled by the incident beam intensity. SAXS patters were integrated using Fit2D [48] after buffer subtraction.

The radius of gyration (R_g) was determined by the Guinier equation and by the indirect Fourier transform method using GNOM software [49]. The distance distribution function P(r) was also analyzed by GNOM and the maximum diameter (D_{max}) was assigned. The experimental molecular weight of NcCel45A based on SAXS data was calculated using the SAXS MoW [50].

SAXS Ab Initio Modeling

Dummy atom models (DAMs) were calculated from the scattering curves using DAMMIN package [51]. Ten low-resolution envelope models were generated and averaged with the program DAMAVER [52]. The superposition of the low-resolution and high-resolution homology models was performed by the program SUPCOMB [53]. Theoretical scattering profiles of ab initio model and homology models were compared to experimental scattering profiles using FoXS [54].

3D Structure and EOM Modeling

The structure models of NcCel45A catalytic and carbohydrate-binding domains were generated independently and automatically using SwissModel [55], based on the structures available at the Protein Data Bank (PDB).

Ensembles of full-length NcCel45A models with constant core domains (catalytic and carbohydrate-binding domains) and a variable linker region were generated using the EOM method [56]. The RANCH program [56] generate a pool of 10,000 structures and GAJOE [56] selected the best ensemble fitting to the experimental SAXS data using

a genetic algorithm. All figures were produced using Pymol (PyMOL Molecular Graphics System, Delano Scientific, Portland, OR)

Results

Sequence Analysis of NcCel45A from *N. crassa*

The NCU05121 gene (GenBank: XM_952014.1) found on chromosome 6 of the *N. crassa* genome encodes a GH45 endoglucanase, subfamily A also known as endoglucanase V. NcCel45A is a modular protein with a *N*-terminal catalytic domain connected to a *C*-terminal carbohydrate-binding module (CBM1) through a heavily *O*-glycosylated linker (Fig. 1). In the linker region, one putative *N*-glycosylation site and twelve *O*-glycosylation sites were identified. The CBM1 was classified as CBM type 1 which was almost exclusively found in fungi and known to be important in cellulose binding. A putative signal peptide, composed by the first 21 amino acids of the *N*-terminal region was also found. Interestingly, NcCel45A has twenty cysteine residues (6.7 % of the total amino acid content) and amino acid sequence alignments with GH45 from *M. albomyces* (76 % identity) and *H. insolens* (71 % identity) suggested that all cysteine residues are making disulfide bonds (Fig. 2).

Cloning, Expression, and Purification of the Recombinant NcCel45A from *N. crassa*

The NcCel45A open reading frame was PCR amplified without its signal peptide and inserted into the modified pEXPYR vector that allows the protein expression and secretion using the glucoamylase signal peptide and an *N*-terminal $(His)_6$ tag. The recombinant enzyme was successfully expressed after 48 h but the protein did not adsorb to the nickel affinity chromatography column probably because the His-tag was buried or was removed. Therefore, the protein was purified by anion exchange and size exclusion chromatography resulting in a final yield of 2.8 mg of pure protein per liter of culture medium.

The SDS-PAGE shows a main protein band with a migration pattern that agrees with the predicted molecular weight of 31.2 kDa and additional diffuse bands that suggest multiple glycosylation moieties (Fig. 3a). To confirm protein identity, the purified enzyme was analyzed by MALDI-TOF/MS. The MASCOT results match perfectly to endoglucanase V from *N. crassa* OR74A (NCU05121) with a score of 29 and a sequence coverage of 40 %, thus confirming protein identity (Fig. 3b).

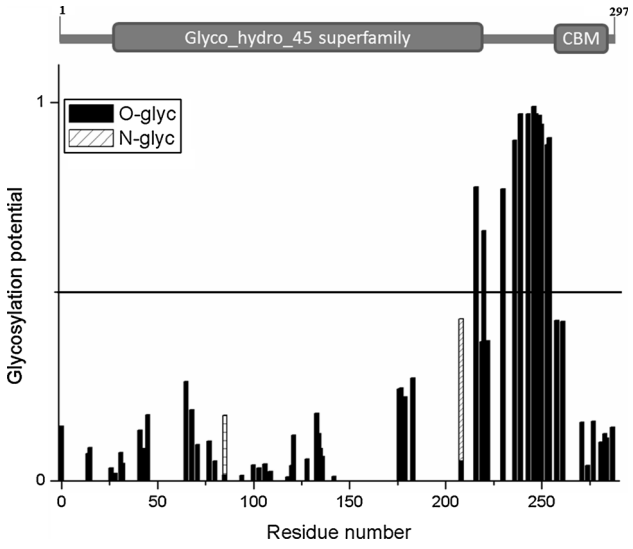

Fig. 1 Schematic diagram of putative NcCel45A glycosylations. Putative glycosylation decorations based on NetOGlyc 4.0 [37] and NetNGlyc 1.0 [38] are shown

pH and Temperature Stability Accessed by ThermoFluor Assay

The ThermoFluor analysis was used to determine NcCel45A temperature and pH stability. All tested buffers resulted in measurable transitions and T_m values were determined using the Boltzmann model (Fig. 4a). The highest T_m was 72 °C, recurrent for several tested conditions and persistent through a wide pH range (5.0–9.0) (Fig. 4b and Additional file 1: Table S1). Under extreme pH conditions, pHs ranging from 3.0 to 4.5 or 9.0 and above, a decrease in thermal induced melting occurred, although they remained higher than 55 °C. These results highlight the elevated pH and thermal stability of NcCel45A.

Biochemical Characterization of NcCel45A

The optimum temperature and pH for NcCel45A activity utilizing β-glucan substrates were 60 °C and 6.0, respectively (Fig. 5). The substrate specificity of the recombinant enzyme was determined using a panel of eleven soluble and insoluble substrates (Table 1). NcCel45A exhibited a preference for lichenan and β-glucan with specific activities of 1.05 ± 0.01 and 1.02 ± 0.03 U/mg prot., respectively. These results point to higher specific activity toward substrates with β-1,4-glucosidic linkages (Table 1). The enzyme also showed measurable activity toward filter paper (0.54 ± 0.01 U/mg prot.), CMC (0.21 ± 0.04 U/mg prot.), and PASC (0.16 ± 0.01 U/mg prot.) (Table 1).

Next, we set out to obtain insights into the NcCel45A mechanism of cellulose depolymerization based on cellohexaose (C6), cellopentaose (C5), and cellotetraose (C4) hydrolysis. Degradation products of short defined

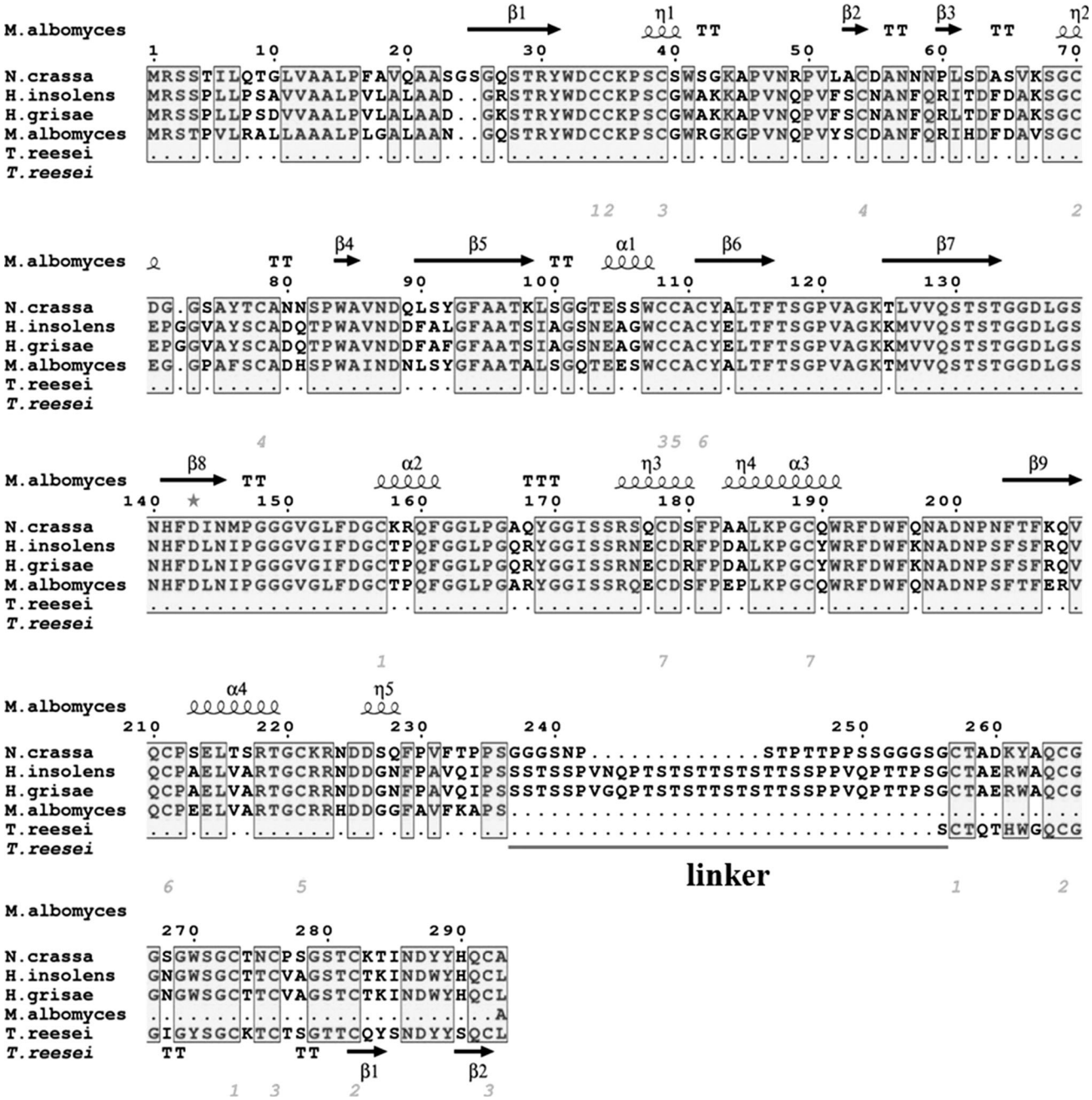

Fig. 2 Amino acid sequence alignment of Cel45 from *Neurospora crassa* (Uniprot Q872Q1), *Humicola grisea var. thermoidea* (PDB code 1HD5), *Humicola insolens* DSM 1800 (PDB code 2ENG), *Melanocarpus albomyces* (PDB code 1L8F), and CBM from *Trichoderma reesei* (PDB code 4BMF). The secondary structure elements are shown by arrows (strands), spirals (helices), and turns (T) from the crystal structure of Cel45 from *Melanocarpus albomyces* (PDB code 1L8F) and CBM from *Trichoderma reesei* (PDB code 4BMF). Conserved catalytic residues are marked with asterisks. *Green* numbers indicate disulfide bonds

gluco-oligosaccharides (C6, C5, and C4) were analyzed by MALDI-TOF/MS. Cellobiose (C2) and cellotriose (C3) were released from the longer C6 and C5 substrates, whereas cellotetraose (C4) was observed only using cellohexaose (C6) as substrate (Fig. 6). Remarkably, C(N-1) oligosaccharides (C5 in the case of cellohexose and C4 in the case of cellopentaose) have not been observed for neither substrates. This means that the enzyme is unable to release a single glucose molecule from any substrate. Moreover, NcCel45A was unable to hydrolyze C4. A similar mode of action has been previously reported for Cel45 from *P. pastoris* and *Reticulitermes speratus* using thin-layer chromatography and anion exchange chromatography, respectively [13, 57].

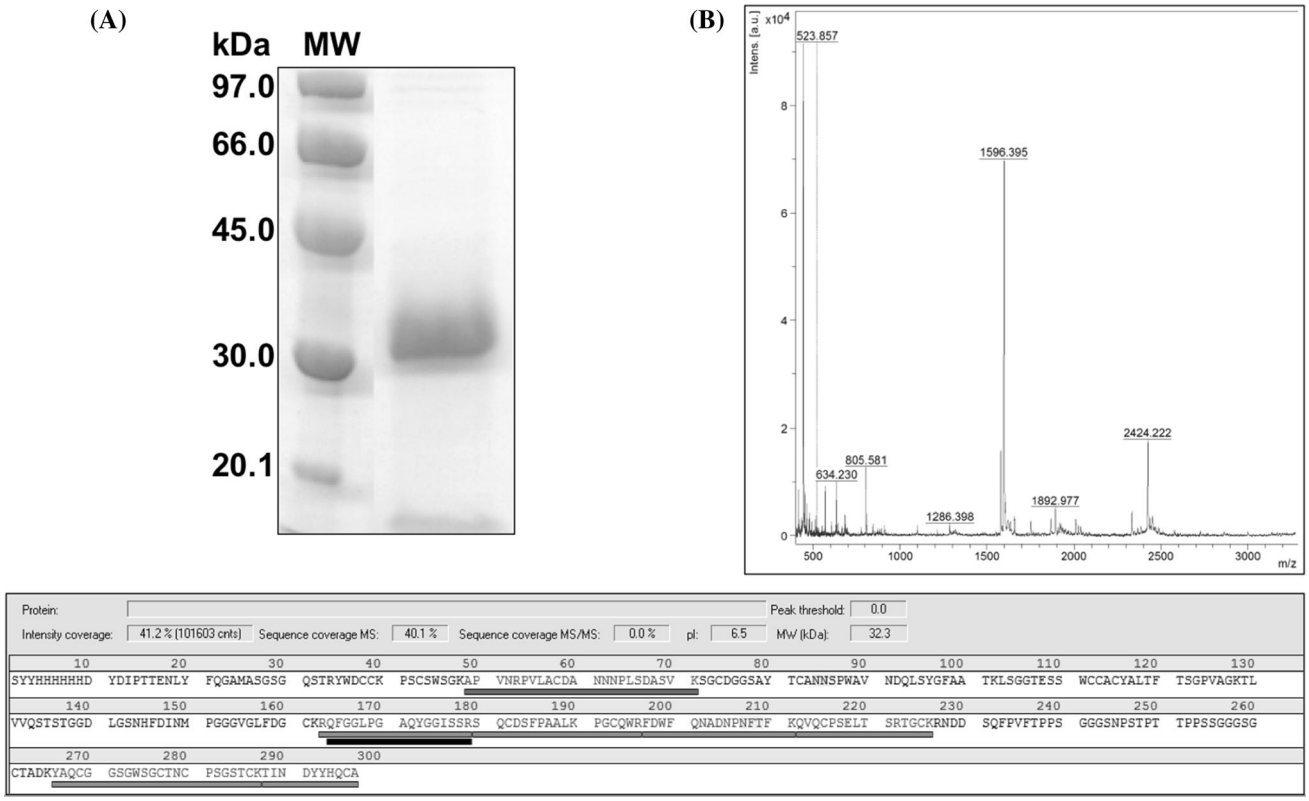

Fig. 3 SDS-PAGE and identification by mass spectrometry of the NcCel45A heterologous expression in *Aspergillus nidulans*. **a** SDS-PAGE 15 % of the purified enzyme. *MW* molecular mass marker. The gel was stained with Coomassie *blue*. **b** Mass spectrogram of NcCel45A digested with trypsin. Protein identification and sequence coverage after m/z list analysis

Furthermore, we analyzed the temperature stability and the effects of metals, solvents, reducing agents, and surfactants on NcCel45A.

Thermal stability results (Fig. 7) indicate that the recombinant enzyme retained at least 80 % of its original activity when incubated at 70 °C for 72 h. However, when incubated at 80 or 90 °C for 10 h less than 20 % of the original activity was retained (Fig. 7). These results are consistent with our ThermoFluor analysis (see "Result" section above) thus further corroborating the favorable high-thermo and wide-pH-range stability of NcCel45A.

The influence of metals as well as other chemicals on the relative activity of NcCel45A shown in Table 2. The results show that NcCel45A activity is strongly inhibited by Ni^{2+}, Mn^{2+}, Zn^{2+}, Mg^{2+}, Cu^{2+}, SDS, and Triton X-100, while Ca^{2+} and Co^{2+} significantly enhance enzyme activity. ##NcCel45A activity was unaffected by methanol, however acetone, isopropanol, and ethanol enhanced specific enzymatic activity (up to 20 %).

Low-Resolution Molecular Structure and Structure-Based Homology Modeling

Structural information about full-length multidomain cellulases is not forthcoming. Therefore, we subjected NcCel45A to SAXS analysis aiming the determination of its molecular shape in solution. The X-ray scattering curve obtained at 0.88 mg/mL and 25 °C is shown in Fig. 8a. The Guinier plot exhibited a linear behavior indicating a monodispersity of NcCel45A (Fig. 8a). The gyration radius (R_g) using the Guinier approximation was 23.40 ± 0.485 Å and the asymmetric pair distance distribution function (P(r)) profile (Fig. 8b) suggested an elongated protein shape with a maximum dimension (D_{max}) of 75 Å.

NcCel45A global compactness and flexibility in solution was evaluated from SAXS experimental data using Kratky plot analysis. The Kratky curve presented a well-defined maximum (Fig. 8c) but did not decay to zero at higher q values, suggesting some degree of flexibility that can be justified by the two-domain architecture of the enzyme, connected by a flexible inter-domain linker peptide. As estimated by SAXS MoW [50] relying on SAXS real data, the molecular weight of the protein was 32.3 kDa, very close to the theoretical MW of 31.2 kDa, computed on the basis of its amino acid sequence. The difference between these two values might be the result of protein glycosylation and/or methodological imprecisions [58]. SAXS data collection and analysis parameters are given in Table 3.

The low-resolution molecular model for NcCel45A using DAMMIN showed an excellent agreement with

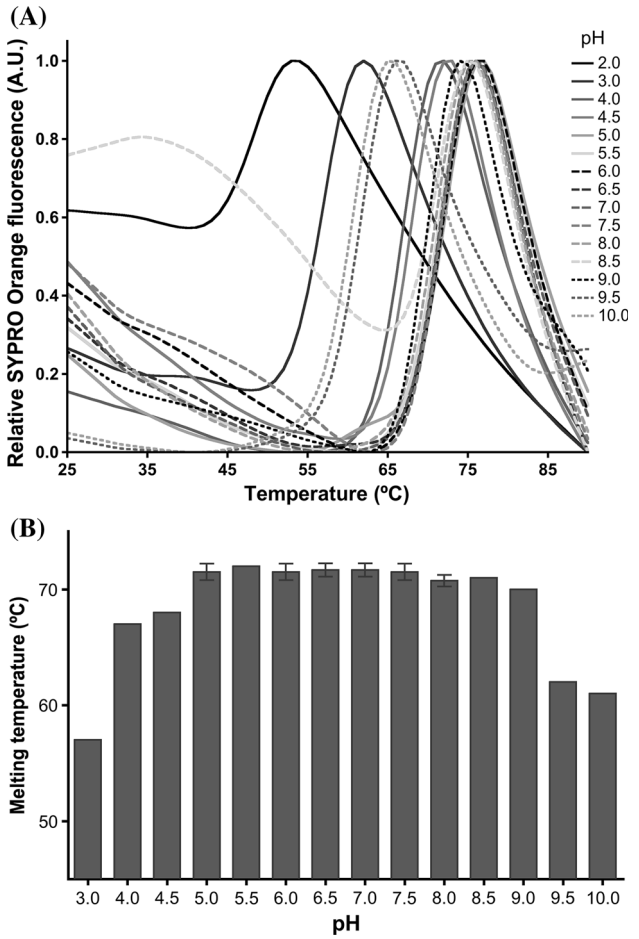

Fig. 4 Enzyme thermal stability determined by a ThermoFluor assay. **a** Normalized thermal denaturation curves allowing to visualize the thermal shift between different pH conditions. **b** Variation of the melting temperature (T_m) determined using the Boltzmann model for all tested pHs (3–10)

Table 1 Hydrolytic specific activity of the purified NcCel45A

Substrate	Specific activity (U/mg protein)
Lichenan	1.05 ± 0.01
β-glucan	1.02 ± 0.03
Filter paper	0.54 ± 0.01
CMC	0.21 ± 0.04
PASC	0.16 ± 0.01
Xyloglucan	0.09 ± 0.03
Avicel	0.09 ± 0.03
1,4-β-D-manan	<0.01
Galactomanan	<0.01
Arabinan	<0.01
pNPG	<0.01

The activity values are the mean of three replicates

experimental SAXS data (χ value of 2.5). The molecular envelope suggested a monkey-wrench-shaped molecule indicating relative positions for the *N*-terminal (catalytic) and the *C*-terminal (carbohydrate-binding) domains (Fig. 9).

Since a NcCel45A high-resolution tridimensional structure was not available, we built homology-based structural models for each domain (catalytic and carbohydrate binding) and employed ensemble optimization modeling (EOM) to fit molecular models with experimental SAXS data. The resulting homology models having catalytic and cellulose binding domains were treated as independent rigid bodies connected by a flexible linker of 27 dummy residues. The R_g and D_{max} distributions showed major peaks at 21.4 and 73.6 Å, respectively (Fig. 8d).

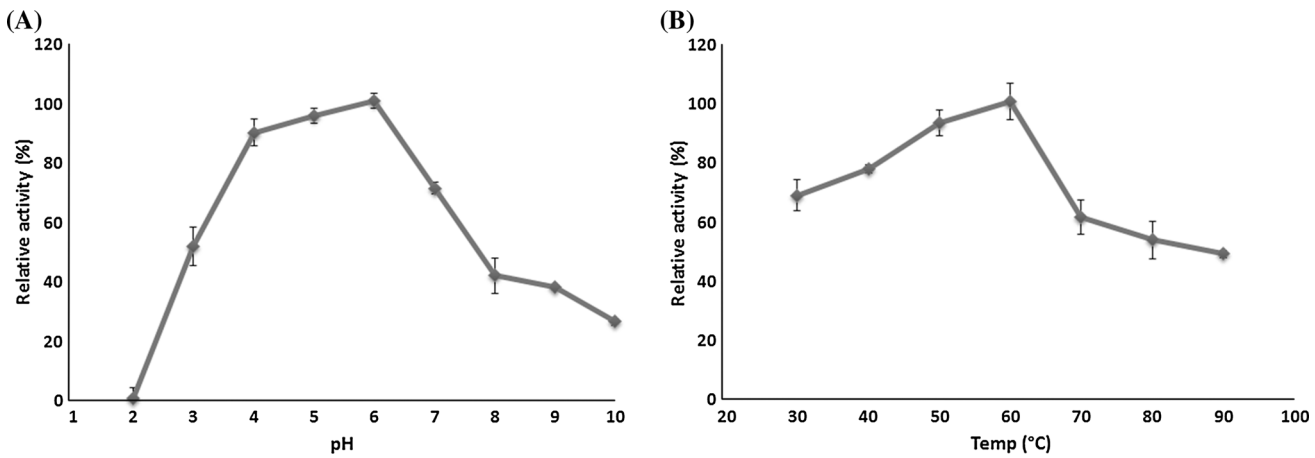

Fig. 5 Effect of pH (**a**) and temperature (**b**) on enzymatic activity. The NcCel45A endoglucanase was incubated with 0.2 % β-glucan for 1 h at 60 °C (**a**) and with β-glucan (0.2 %) for 1 h at pH 6.0 (**b**). Results are the average of three replicates and the *bars* indicate the experimental errors

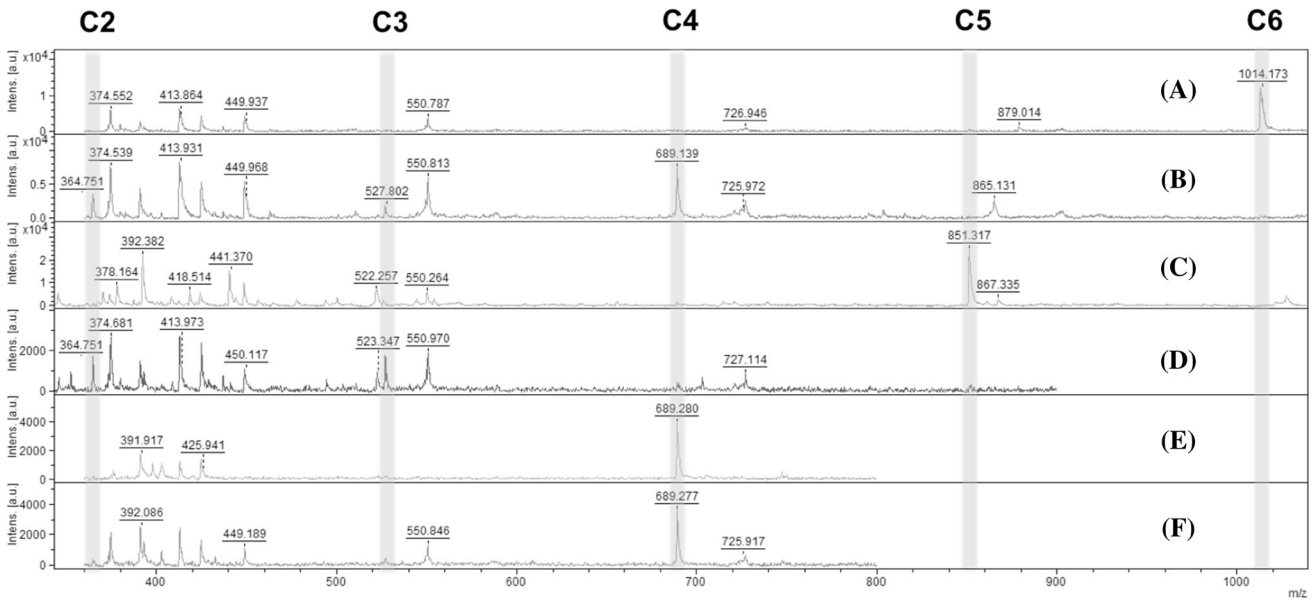

Fig. 6 MALDI-TOF mass spectrum of cellohexaose, cellopentaose, and cellotetraose hydrolyzed products generated by GH45. Formulas of acidic oligosaccharides were deduced from the observed mass-to-charge signals of the corresponding $[M + Na]^+$ ions. Spectrum of cellohexaose, cellopentaose, and cellotetraose substrates before (*a*, *c*, *e*) and after (*b*, *d*, *f*) enzyme hydrolysis are indicated. *C6* cellohexaose, *C5* cellopentaose, *C4* cellotetraose, *C3* cellotriose, and *C2* cellobiose

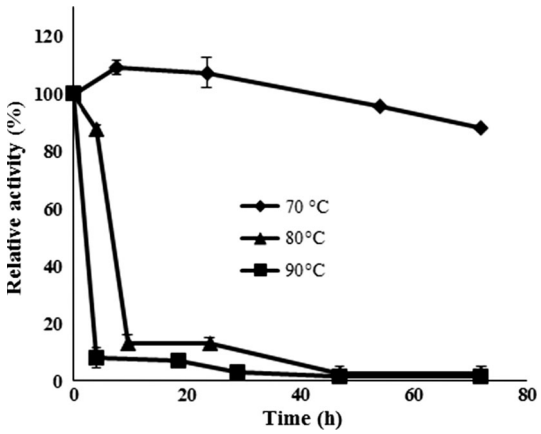

Fig. 7 Thermostability of the purified NcCel45A. The enzymatic assay was performed using β-glucan (0.5 %) as substrate after each incubation time. Results are the average of three replicates and the *bars* indicate the experimental errors

This type of homology-guided structural modeling of the catalytic and carbohydrate-binding domains along with linker length assumptions is in excellent agreement with our X-ray scattering values (Table 3). An ensemble containing just two models fits the experimental data well ($\chi = 2.9$), which suggests a limited diversity of the enzyme conformations (Fig. 8d). However, the linker conformation does not fit well within the molecular envelope of the enzyme. This might steam from the absence of glycosylations in our model since we were unable to ascertain the positions and conformations of glycan decorations. Similar results were observed by Guttman and co-workers [59] for a set of glycosylated proteins. Nevertheless, the most representative EOM model shows a good agreement with the low-resolution model (Fig. 9) allowing to assign the relative domain positions but not their orientation. Taken together, these results support the predominantly compact structure with a flexibility restricted to the linker region. A monkey-wrench molecular shape of NcCel45A might be relevant to its activity and substrate specificity.

Discussion

Prospection of novel enzymes for biotechnological application remains a field to be expanded. In particular, in recent years, cellulolytic enzymes received considerable attention powered by a necessity to develop biomass-to-biofuels technologies for example, for second-generation (defined as extracting carbon from a lignocellulosic source, e.g. biomass) ethanol production. Enzymes belonging to family GH45 are typical endoglucanases relevant to lignocellulosic biomass transformation, cotton fiber processing, and additives in detergents [28, 30]. Here, we report on biochemical and structural characterization of NcCel45A, a GH45 endoglucanase from *N. crassa*.

Based on sequence analysis, NcCel45A is a bimodular enzyme with an *N*-terminal catalytic domain connected to a

Table 2 Effect of metal ions, solvents, and surfactants on enzymatic activity

Effectors	Concentration	Relative activity (%)
Control	–	100.0
EDTA	1 mM	99.1 ± 0.0
Ni^{2+}	1 mM	82.2 ± 0.5
Ca^{2+}	1 mM	110.9 ± 4.5
Mn^{2+}	1 mM	76.8 ± 0.0
Zn^{2+}	1 mM	72.0 ± 4.7
Fe^{2+}	1 mM	98.5 ± 0.0
Mg^{2+}	1 mM	65.5 ± 9.1
Cu^{2+}	1 mM	69.4 ± 0.0
Co^{2+}	1 mM	140.4 ± 2.1
Acetone	20 %	122.0 ± 1.6
Isopropanol	20 %	118.8 ± 1.3
Methanol	20 %	103.4 ± 8.4
Ethanol	20 %	119.6 ± 0.8
SDS	0.1 %	4.4 ± 6.2
Triton X-100	1 %	82.8 ± 12.3
Tween-20	1 %	105.3 ± 8.9

The activity values are the mean of three replicates

C-terminal cellulose binding module through a glycosylated linker. The optimal pH and temperature of the enzyme, 6.0 and 60 °C, are similar to the ones observed for GH45s from *Staphylotrichum coccosporum* [27], *H. grisea* (EGL4-CBD) [31], *H. grisea* (EGL3), *Chaetomium brasiliensis*, and *Trichoderma reseei* [60–62]. The slightly acidic optimum pH of NcCel45A maybe be attractive for the laundry detergent applications [30].

Among all *N. crassa* endoglucanases only endo-IV shares the same optimal conditions with NcCel45A [23]. Out of all the tested substrates, NcCel45A showed the highest activity against lichenan and β-glucan. Similar specificity was also reported for GH45 from *P. pastoris* [13] and *Phanerochaete chrysosporium* [14] confirming the preference of these class of enzymes for β-1,4-glucosidic bonds. NcCel45A showed also considerable activity toward filter paper so far only reported for Cel45 from *T. reesei* [63]. This activity can be related to high defibrillation of lyocell fabric showed by Cel45 from *Mucorales* [28, 64]. The loose cellulosic fibers present in filter paper may be similarly recognized by NcCel45A [28, 64]. The gluco-oligosaccharide degradation pattern reinforced the endoglucanase mode of action of Cel45 enzymes and suggested C5 as the minimum substrate of NcCel45A. The GH45 endoglucanase mechanism was also corroborated by a typical open active-site cleft observed in crystallographic structures [17, 65]. Moreover the inability to hydrolyze cellotetraose was not observed in well-studied cellobiohydrolases [66, 67], and a cellotetraose accumulation has been observed using larger substrates like PASC and CMC [25, 60]. These results indicate that at least +2 to −3 catalytic subsites must be occupied for the hydrolysis to take place. In line with this observation, Davis et al. [68] reported four subsites before the aspartic catalytic residues. Therefore substrates smaller then C4 cannot be efficiently recognized and cleaved by the enzyme.

Another attractive biochemical property of NcCel45A is its quite remarkable thermostability [69]. The enzyme maintained more than 80 % of its hydrolytic activity after 72 h of pre-incubation at 70 °C. These results were reinforced by an experimentally determined melting temperature of 72 °C as revealed by ThermoFluor assay. The latter assay also revealed NcCel45A structure integrity in a pH range from 5 to 9. Enzyme thermostability is a characteristic that might be attractive for the biotechnological and biofuel applications. Enzymatic hydrolysis of (hemi) cellulosic substrates performed at high temperatures frequently is more efficient. In addition, high temperatures tend to decrease substrate viscosity and microbial contaminations. Other GH45 s, such as GH45 from *P. pastoris* [13], *H. grisea* [61] and *Acremonium thermophilum* [70], also have elevated thermostability.

There is a consensus that understanding structural factors responsible for thermostability of enzymes still represent a considerable challenge [69, 71]. However, in the case of NcCel45A some underlying molecular reasons seem to be clear. Based on amino acid sequence comparisons with structurally characterized proteins, the enzyme has 7 disulfide bonds within the catalytic domain and 3 disulfide bonds within the carbohydrate-binding module. A high number of disulfide bonds also stands out in other GH45s. Thus, the large number of disulfide bonds tightly locks the tertiary structure of NcCel45A, increasing its compactness and preventing from incorrect refolding after heating. From a thermodynamical point of view, these intra-molecular interactions stabilize the tridimensional structure by reducing the entropy of the denatured state [72]. Disulfide bonds rich proteins have also been reported for many thermophilic organisms, and the covalent interaction present in the proteins structures have been considered to play a major role in protein stability [73, 74]. Interestingly, with exception of mesophilic GH45s from *N. crassa* and *P. pastoris*, all other GH45 reported as so far are from thermophilic fungi.

In addition to temperature stability, resistance to solvents and detergents is also crucial for a number of industrial applications. All the tested solvents at 20 % concentration did not affect the enzymatic activity of NcCel45A and in the case of acetone, isopropanol, and

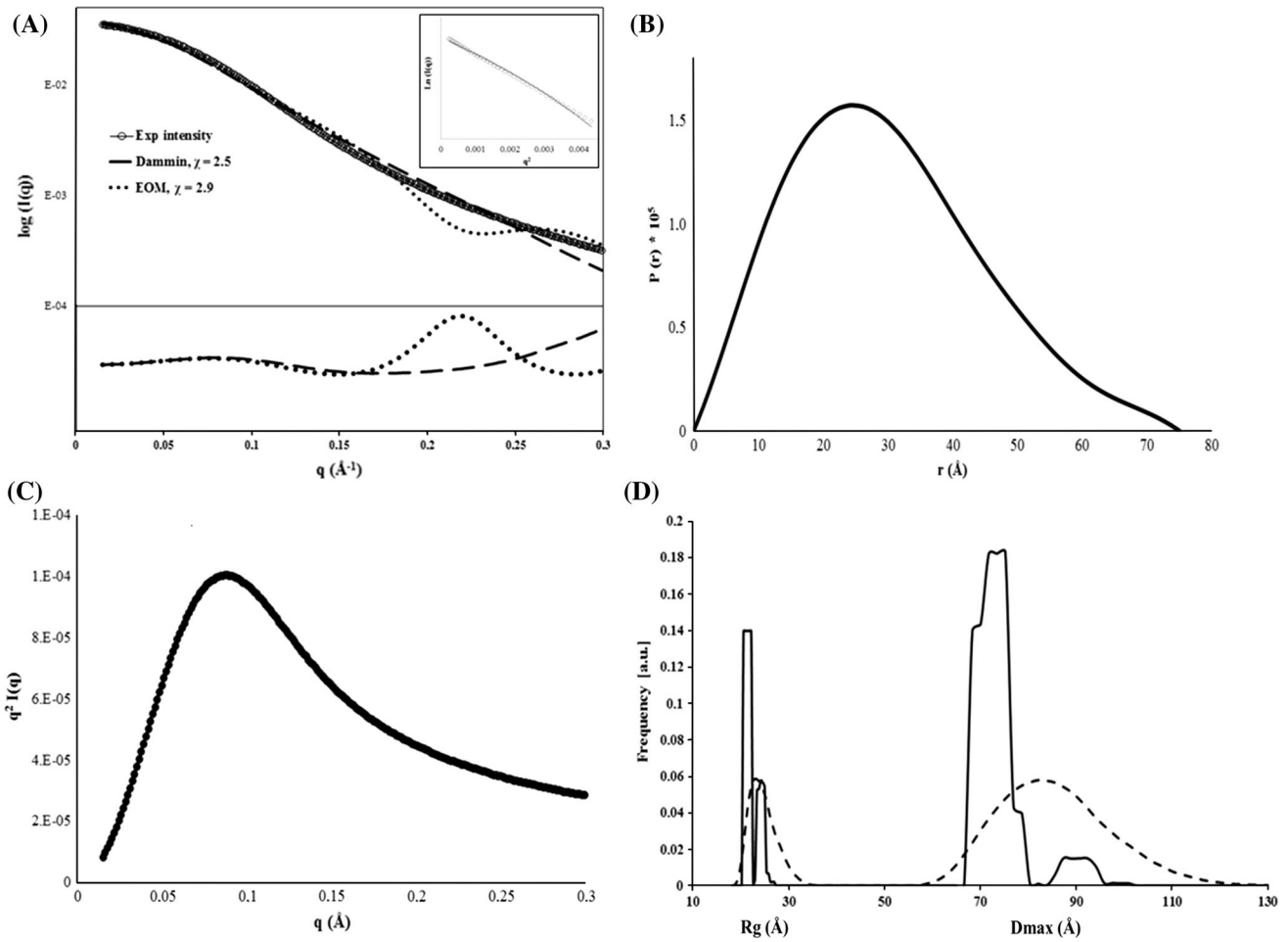

Fig. 8 SAXS data analysis of NcCel45A. **a** Scattering curves differences between the experimental data and theoretical profiles of DAMMIN (*dashed line*) and EOM models (*dotted line*). The *inset* shows the Guinie plot. Residuals for the fit are shown below. **b** Normalized pair distribution function P(r). **c** Kratky plot. **d** Distribution of radius of gyration (Rg) and maximum diameter (D_{max}) of the generated model ensemble pool of 10,000 models (*dashed lines*) and the best ensemble fitting (*fill lines*) generated by EOM

ethanol an increase superior of 10 % in enzyme activity was observed (Table 2). Notably, ethanol at 20 % concentration not only had no detrimental effect on the enzymatic activity, but also increased activity in 20 %. This feature might be quite attractive for the cellulosic ethanol process of simultaneous saccharification and fermentation, the process during which the hydrolytic enzymes are subjected to increased concentrations of ethanol as the fermentation proceeds. Ethanol tolerance is an essential factor under these conditions [75], which makes NcCel45A potentially promising enzyme for the ethanol production from lignocellulosic biomass. The positive influence of ethanol on enzymatic activity was also observed for β-glucosidase from *Myceliophthora thermophila* and was justified by the effect of environment polarity changes induced by solvents and resulting in stabilization of protein conformation [76]. A high number of cysteine residues present in NcCel45A might also have a positive impact on its stability toward organic solvents.

EDTA had no effect on NcCel45A activity suggesting that the reaction is not dependent on divalent ions. However, the enzyme activity decreased significantly using many divalent ions except for calcium and cobalt. NcCel45A is strongly inhibited by SDS but not by nonionic detergents such as TritonX-100.

Our SAXS envelope analysis confirms that NcCel45A has an extended multidomain structure. The raw SAXS data show that NcCel45A is somewhat elongated with a maximum dimension of 75 Å, and our structural modeling suggests that NcCel45A does not adopt a fully extended conformation in solution. Amino acid sequence analysis of the linker region reveals conformational restricted peptides

Table 3 SAXS data collection and scattering-derived parameters

Data collection parameters	
Beamline	LNLS D02A-SAXS2
Wavelength (Å)	1.488
q range ($Å^{-1}$)	0.008–0.34
Exposure time per frame (s)	300
Concentration range (mg/mL)	0.44
Temperature (°C)	25
Structural parameters	
R_g (Å) (from Guinier) (± SE)	23.40 ± 0.485
R_g (Å) (from GNOM) (± SE)	23.33 ± 0.008
D_{max} (Å)	75
Resolution (Å)	18.5
MW^{SAXS} (kDa)	32.3
MW^{Teo} (kDa)	31.2
Ab initio modeling	
Number of models	10
NSD	0.645 ± 0.023
χ	2.590 ± 0.068
Software employed	
Primary data reduction	Fit2D
Data processing	PRIMUS
Modeling	DAMMIN
	EOM

Resolution: $2\pi/q_{max}$

"TPPS" and "PSTP" intercalated with two "GGG" strings, which presumably introduce flexible hinges and/or bends in the linker region. Partial flexibility of the enzyme is also supported by Kratky plot. The monkey-wrench shape of NcCel45A with partially extended linker is quite different from *H. insolens* Cel45. The later heavily *O*-glycosylated enzyme has a tadpole molecular shape and its linker is in extended conformation, leading to a maximum enzyme dimension of 125 Å [58]. This enzyme shares 60 % sequence identity with NcCel45A. Although *H. insolens* Cel45 has only a 13 amino acid longer linker, its maximum molecular dimension is almost 50 Å greater (75 Å for NcCel45A vs 125 Å for *H. insolens* Cel45). This seems to imply that *H. insolens* Cel45 has more rigid and more heavily glycosylated linker, which favors an extended, tadpole-shaped conformation of the enzyme. This is in agreement with the general notion that linker *O*-glycosylations favor a more extended conformation in cellulases [58, 77]. The differences in molecular organization observed for these two enzymes may translate into differences in substrate specificity. Opposite to NcCel45A, which is highly active toward lichenan and β-glucan, *H. insolens* Cel45 was reported to have higher activity against polymeric cellulose substrates (such as PASC) and lower activity on hemicellulosic substrates [63].

In summary, we characterized a GH45 endoglucanase from the mesophilic fungus *N. crassa*. Our results demonstrate that this enzyme has high stability at elevated temperatures (70 °C) and retains its activity at high concentration to solvents (such as ethanol, isopropanol, and acetone). These characteristics are considered attractive for the industrial enzymes which makes NcCel45A an interesting candidate for biotechnological and bioindustrial applications.

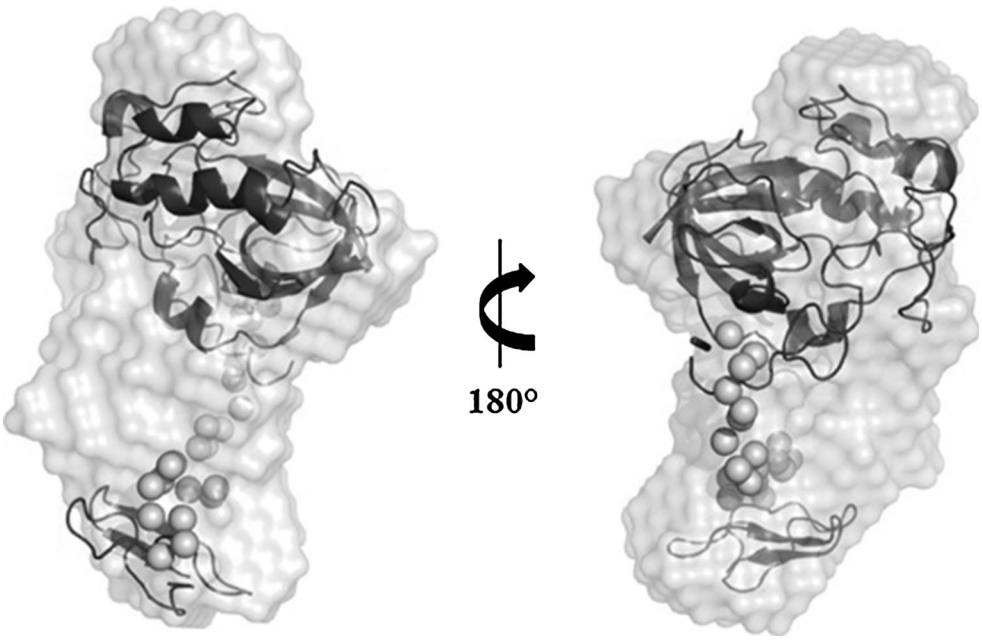

Fig. 9 Structural modeling of NcCel45A based on SAXS data. EOM-generated model representing the most representative NcCel45A structure (*purple*) is overlaid with the average molecular envelope shown in *gray*

Acknowledgments This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via grants 2008/56255-9, 2009/54035-4, 2009/11536-3, 2010/08680-2, 11/20505-4 and 09/11536-3; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via grants 490022/2009-0, and 301981/2011-6; and the University of São Paulo (USP) via NAP de Bioenergia & Sustentabilidade (NAPBS) e NAP de Instrumentação (CIEA_MNB). We would like to thank Livia Manzine, Alexandre Antoniazzi, Maria Auxiliadora Morim Santos and Mariana Ortiz de Godoy for technical support and the Brazilian Synchrotron Light Laboratory (LNLS, Campinas) for access to the SAXS beam line.

References

1. Jarvis, M. (2003). Chemistry—Cellulose stacks up. *Nature, 426*, 611–612.
2. Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology, 6*, 850–861.
3. Henrissat, B. (1994). Cellulases and their interaction with cellulose. *Cellulose, 1*, 169–196.
4. Zhang, Y. H., & Lynd, L. R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering, 88*, 797–824.
5. Horn, S. J., Vaaje-Kolstad, G., Westereng, B., & Eijsink, V. G. (2012). Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels, 5*, 45.
6. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., & Foust, T. D. (2007). Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science, 315*, 804–807.
7. Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R., & Monot, F. (2009). New improvements for lignocellulosic ethanol. *Current Opinion in Biotechnology, 20*, 372–380.
8. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Research, 37*, D233–D238.
9. Hirvonen, M., & Papageorgiou, A. C. (2003). Crystal structure of a family 45 endoglucanase from *Melanocarpus albomyces*: Mechanistic implications based on the free and cellobiose-bound forms. *Journal of Molecular Biology, 329*, 403–410.
10. Nakamura, Y., Moriya, T., Baba, Y., Yanai, K., Sumida, N., Nishimura, T., Murashima, K., Nakane, A., Yaguchi, T., Koga, J., Murakami, T. and Kono, T. (2000) Endoglucanases and cellulase preparations containing the same. *WO patent* 2000024879.
11. Jakobsen, T. S., Lindegaard, P., & Chan, M. (1998). Color-care cellulases: Fabric color shield. *Inform, 9*, 788–792.
12. Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., et al. (1993). Structure and function of endoglucanase V. *Nature, 365*, 362–364.
13. Couturier, M., Feliu, J., Haon, M., Navarro, D., Lesage-Meessen, L., Coutinho, P. M., & Berrin, J. G. (2011). A thermostable GH45 endoglucanase from yeast: Impact of its atypical multimodularity on activity. *Microbial Cell Factories, 10*, 103.
14. Igarashi, K., Ishida, T., Hori, C., & Samejima, M. (2008). Characterization of an endoglucanase belonging to a new subfamily of glycoside hydrolase family 45 of the basidiomycete *Phanerochaete chrysosporium*. *Applied and Environment Microbiology, 74*, 5628–5634.
15. Eyun, S. I., Wang, H., Pauchet, Y., Ffrench-Constant, R. H., Benson, A. K., Valencia-Jiménez, A., et al. (2014). Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*). *PLoS ONE, 9*, e94052.
16. Palomares-Rius, J. E., Hirooka, Y., Tsai, I. J., Masuya, H., Hino, A., Kanzaki, N., et al. (2014). Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. *BMC Evolutionary Biology, 14*, 69.
17. Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C., & Schülein, M. (1995). Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1.9 A resolution. *Biochemistry, 34*, 16210–16220.
18. Valjakka, J., & Rouvinen, J. (2003). Structure of 20 K endoglucanase from *Melanocarpus albomyces* at 1.8 angstrom resolution. *Acta Crystallographica D, 59*, 765–768.
19. Davis, R. H., & Perkins, D. D. (2002). Timeline: *Neurospora*: A model of model microbes. *Nature Reviews Genetics, 3*, 397–403.
20. Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature, 422*, 859–868.
21. Romero, M. D., Aguado, J., González, L., & Ladero, M. (1999). Cellulase production by *Neurospora crassa* on wheat straw. *Enzyme and Microbial Technology, 25*, 244–250.
22. Eberhart, B. M., Beck, R. S., & Goolsby, K. M. (1977). Cellulase of *Neurospora crassa*. *Journal of Bacteriology, 130*, 181–186.
23. Yazdi, M. T., Woodward, J. R., & Radford, A. (1990). The cellulase complex of *Neurospora crassa*: Activity, stability and release. *Journal of General Microbiology, 136*, 1313–1319.
24. Nakamura, A., Ishida, T., Fushinobu, S., Kusaka, K., Tanaka, I., Inaka, K., et al. (2013). Phase-diagram-guided method for growth of a large crystal of glycoside hydrolase family 45 inverting cellulase suitable for neutron structural analysis. *Journal of Synchrotron Radiat., 20*, 859–863.
25. Liu, G., Wei, X., Qin, Y., & Qu, Y. (2010). Characterization of the endoglucanase and glucomannanase activities of a glycoside hydrolase family 45 protein from *Penicillium decumbens* 114-2. *Journal of General and Applied Microbiology, 56*, 223–229.
26. Sakamoto, K., & Toyohara, H. (2009). Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam *Corbicula japonica*. *Comparative Biochemistry and Physiology, 152*, 390–396.
27. Koga, J., Baba, Y., Shimonaka, A., Nishimura, T., Hanamura, S., & Kono, T. (2008). Purification and characterization of a new family 45 endoglucanase, STCE1, from *Staphylotrichum coccosporum* and its overproduction in *Humicola insolens*. *Applied and Environment Microbiology, 74*, 4210–4217.
28. Shimonaka, A., Koga, J., Baba, Y., Nishimura, T., Murashima, K., Kubota, H., & Kono, T. (2006). Specific characteristics of family 45 endoglucanases from *Mucorales* in the use of textiles and laundry. *Bioscience, Biotechnology, and Biochemistry, 70*, 1013–1016.
29. Bukhtojarov, F. E., Ustinov, B. B., Salanovich, T. N., Antonov, A. I., Gusakov, A. V., Okunev, O. N., & Sinitsyn, A. P. (2004). Cellulase complex of the fungus *Chrysosporium lucknowense*: Isolation and characterization of endoglucanases and cellobiohydrolases. *Biochemistry (Biokhimiia), 69*, 542–551.
30. Baba, Y., Shimonaka, A., Koga, J., Kubota, H., & Kono, T. (2005). Alternative splicing produces two endoglucanases with one or two carbohydrate-binding modules in *Mucor circinelloides*. *Journal of Bacteriology, 187*, 3045–3051.
31. Takashima, S., Iikura, H., Nakamura, A., Hidaka, M., Masaki, H., & Uozumi, T. (1999). Comparison of gene structures and enzymatic properties between two endoglucanases from *Humicola grisea*. *Journal of Biotechnology, 67*, 85–97.
32. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research, 25*, 3389–3402.

33. Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology, 302*, 205–217.
34. Gouet, P., Robert, X., & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research, 31*, 3320–3323.
35. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) *Protein identification and analysis tools on the ExPASy server.* Totowa: Humana Press.
36. Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*, 785–786.
37. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., et al. (2013). Precision mapping of the human *O*-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal, 32*, 1478–1488.
38. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research, 40*, W597–W603.
39. Wood, T. M. (1988). Preparation of crystalline, amorphous, and dyed cellulase substrates. In S. T. K. Willis & A. Wood (Eds.), *Methods Enzymol* (pp. 19–25). New York: Academic Press.
40. Segato, F., Damásio, A. R., Gonçalves, T. A., de Lucas, R. C., Squina, F. M., Decker, S. R., & Prade, R. A. (2012). High-yield secretion of multiple client proteins in *Aspergillus*. *Enyzme and Microbial Technology, 51*, 100–106.
41. Vogel, H. J. (1956). A convenient growth medium for *Neurospora (Medium N)*. *Microbial Genetics Bulletin*, *13*, 42–43.
42. Aslanidis, C., & de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research, 18*, 6069–6074.
43. Camilo, C. M., & Polikarpov, I. (2014). High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC). *Protein Expression and Purification, 99*, 35–42.
44. Bradford, M. M. (1976). Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Analytical Biochemistry, 72*, 248–254.
45. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*, 680–685.
46. Westermeier, R., & Naven, T. (2002). *Proteomics in practice: A laboratory manual of proteome analysis*. Weinheim: Weiley.
47. Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chem, 31*, 426–428.
48. Hammersley, A. P. (1997) FIT2D: An introduction and overview. *ESRF Internal Report.*
49. Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography, 25*, 495–503.
50. Fischer, H., Oliveira Neto, M. D., Napolitano, H. B., Polikarpov, I., & Craievich, A. F. (2009). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography, 43*, 101–109.
51. Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution. *Biophysical Journal, 76*, 2879–2886.
52. Volkov, V. V., & Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography, 36*, 860–864.
53. Kozin, M. B., & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography, 34*, 33–41.
54. Schneidman-Duhovny, D., Hammel, M., & Sali, A. (2010). FoXS: A web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research, 38*, W540–W544.
55. Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics, 22*, 195–201.
56. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society, 129*, 5656–5664.
57. Otagiri, M., Lopez, C. M., Kitamoto, K., Arioka, M., Kudo, T., & Moriya, S. (2013). Heterologous expression and characterization of a glycoside hydrolase family 45 endo-beta-1,4-glucanase from a symbiotic protist of the lower termite, *Reticulitermes speratus*. *Biotechnology and Applied Biochemistry, 169*, 1910–1918.
58. Receveur, V., Czjzek, M., Schulein, M., Panine, P., & Henrissat, B. (2002). Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *Journal of Biological Chemistry, 277*, 40887–40892.
59. Guttman, M., Weinkam, P., Sali, A., & Lee, K. K. (2013). All-atom ensemble modeling to analyze small-angle X-ray scattering of glycosylated proteins. *Structure, 21*, 321–331.
60. Karlsson, J., Siika-aho, M., Tenkanen, M., & Tjerneld, F. (2002). Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. *Journal of Biotechnology, 99*, 63–78.
61. Murashima, K., Shimonaka, A., Nishimura, T., Baba, Y., Koga, J., Kubota, H., & Kono, T. (2006). Exploring amino acids responsible for the temperature profile of glycoside hydrolase family 45 endoglucanase EGL3 from *Humicola grisea*. *Bioscience, Biotechnology, and Biochemistry, 70*, 2205–2212.
62. Vlasenko, E., Schulein, M., Cherry, J., & Xu, F. (2010). Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresource Technology, 101*, 2405–2411.
63. Saloheimo, A., Henrissat, B., Hoffren, A. M., Teleman, O., & Penttila, M. (1994). A novel, small endoglucanase gene, egl5, from *Trichoderma reesei* isolated by expression in yeast. *Molecular Microbiology, 13*, 219–228.
64. Shimonaka, A., Murashima, K., Koga, J., Baba, Y., Nishimura, T., Kubota, H., & Kono, T. (2006). Amino acid regions of family 45 endoglucanases involved in cotton defibrillation and in resistance to anionic surfactants and oxidizing agents. *Bioscience, Biotechnology, and Biochemistry, 70*, 2460–2466.
65. Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure, 3*, 853–859.
66. Goto, M., Furukawa, K., & Hayashida, S. (1992). An Avicel-affinity site in an Avicel-digesting exocellulase from a *Trichoderma viride* mutant. *Bioscience, Biotechnology, and Biochemistry, 56*, 1523–1528.
67. Zverlov, V. V., Velikodvorskaya, G. A., & Schwarz, W. H. (2002). A newly described cellulosomal cellobiohydrolase, CelO, from *Clostridium thermocellum*: Investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology, 148*, 247–255.
68. Davies, G. J., Dodson, G., Moore, M. H., Tolley, S. P., Dauter, Z., Wilson, K. S., et al. (1996). Structure determination and refinement of the *Humicola insolens* endoglucanase V at 1.5 angstrom resolution. *Acta Crystallographica D, 52*, 7–17.
69. Yennamalli, R. M., Rader, A. J., Kenny, A. J., Wolt, J. D., & Sen, T. Z. (2013). Endoglucanases: insights into thermostability for biofuel applications. *Biotechnology for Biofuels, 6*, 136.
70. Vehmaanperä, J., Alapuranen, M., Puranen, T., Siika-aho, M., Kallio, J., Hooman, S., Voutilainen, S., Halonen, T. and Viikari, L. (2013) Treatment of cellulosic material and enzymes useful therein. *US patent 20130224801*.

71. Yennamalli, R. M., Rader, A. J., Wolt, J. D., & Sen, T. Z. (2011). Thermostability in endoglucanases is fold-specific. *BMC Structural Biology, 11*, 10.
72. Betz, S. F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Science, 2*, 1551–1558.
73. Beeby, M., O'Connor, B. D., Ryttersgaard, C., Boutz, D. R., Perry, L. J., & Yeates, T. O. (2005). The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biology, 3*, e309.
74. Jorda, J., & Yeates, T. O. (2011). Widespread disulfide bonding in proteins from thermophilic archaea. *Archaea*. doi:10.1155/2011/409156.
75. Wu, Z., & Lee, Y. Y. (1997). Inhibition of the enzymatic hydrolysis of cellulose by ethanol. *Biotechnology Letters, 19*, 977–979.
76. Karnaouri, A., Topakas, E., Paschos, T., Taouki, I., & Christakopoulos, P. (2013). Cloning, expression and characterization of an ethanol tolerant GH3 beta-glucosidase from *Myceliophthora thermophila*. *PeerJ, 1*, e46.
77. von Ossowski, I., Eaton, J. T., Czjzek, M., Perkins, S. J., Frandsen, T. P., Schulein, M., et al. (2005). Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase. *Biophysical Journal, 88*, 2823–2832.