Protein Engineering, Design & Selection, 2016, vol. 29 no. 6, pp. 219–229
doi: 10.1093/protein/gzw011
Advance Access Publication Date: 28 April 2016
Original Article

Original Article

Design and characterizations of two novel cellulases through single-gene shuffling of Cel12A (EG3) gene from *Trichoderma reseei*

Aslı Yenenler[1,*] and Osman Uğur Sezerman[2,*]

[1]Faculty of Engineering and Natural Sciences, Molecular Biology, Genetics and Bioengineering, Sabanci University, 34956 Istanbul, Turkey, and [2]Department of Biostatistics and Medical Informatics, School of Medicine, Acibadem University, Atasehir, Istanbul, Turkey

*To whom correspondence should be addressed. E-mail: asliyenenler@sabanciuniv.edu (A.Y.); ugur.sezerman@acibadem.edu.tr (O.U.S.)

Edited by Gideon Schreiber

Received 24 December 2015; Revised 24 March 2016; Accepted 24 March 2016

Abstract

Cellulases have great potential to be widely used for industrial applications. In general, naturally occurring cellulases are not optimized and limited to meet the industrial needs. These limitations lead to demand for novel cellulases with enhanced enzymatic properties. Here, we describe the enzymatic and structural properties of two novel enzymes, EG3_S1 and EG3_S2, obtained through the single-gene shuffling approach of Cel12A(EG3) gene from *Trichoderma reseei*. EG3_S1 and EG3_S2 shuffled enzymes display 59 and 75% identity in protein sequence with respect to native, respectively. Toward 4-MUC, the minimum activity of EG3_S1 was reported as 5.9-fold decrease in native at 35°C, whereas the maximum activity of EG3_S2 was reported as 15.4-fold increase in native activity at 40°C. Also, the diminished enzyme activity of EG3_S1 was reported within range of 0.6- to 0.8-fold of native and within range of 0.5- to 0.7-fold of native toward CMC and Na-CMC, respectively. For EG3_S2 enzyme, the improved enzymatic activities within range of 1.1- to 1.4-fold of native and within range of 1.1- to 1.6-fold of native were reported toward CMC and Na-CMC, respectively. Moreover, we have reported 6.5-fold increase in the k_{cat}/K_m ratio of EG3_S2 with respect to native and suggested EG3_S2 enzyme as more efficient catalysis for hydrolysis reactions than its native counterpart.

Key words: Cel12A (EG3), endoglucanase, single-gene shuffling, *Trichoderma reseei*

Introduction

In modern biotechnology, genes provide a rich variety of starting materials that can be altered toward industrial needs. There have been great developments in powerful molecular biology techniques that enable the modification of these genes for biotechnological applications by expanding the gene diversity (Kurtzman *et al.*, 2001). DNA shuffling is one of the most powerful techniques, developed by Stemmer, and it has been effectively used for directed evolution of proteins. According to Stemmer (2002), all beautifully complex biological structures and sequences provided by nature have been designed in single process, called as natural evolution, and it becomes possible to mimic natural evolution in gene level by DNA shuffling with the aim of generating genes that encode proteins with new and improved functions (Jackel *et al.*, 2008). There are many successful examples that DNA shuffling, *in vitro* combination, is effectively used for directed evolution of proteins by improving enzyme stability (Song and Rhee, 2000; Akbulut *et al.*, 2013) and activity (Crameri *et al.*, 1996; De Groeve *et al.*, 2009), and altering their substrate specificity (Yano *et al.*, 1998; Chockalingam *et al.*, 2005).

There are two key steps in DNA shuffling process: (i) the fragmentation of parental sequences and (ii) the recombination of fragments by reassembly polymerase chain reaction (PCR). During reassembly,

the fragments are reassembled by an iterative cycle of denaturation, annealing and extension (Miyazaki, 2002) and this step has carefully been followed by screening and selection of the best variants at the protein level. The real aim of DNA shuffling is to create a library of chimeras with recombination of multiple DNA sequences, and the quality of this process has been controlled by (i) the length of the shuffled DNA, (ii) homology of sequences, (iii) the quality of sequence diversity, (iv) the crossover number and (v) the number and molar ratio of parents (Stemmer, 2002). In spite of the recombination of multiple DNA sequences, it is also possible to perform DNA shuffling with a single gene and it has a high potential to contribute the gene diversity (Stemmer, 2002). During the fragmentation step of single-gene DNA shuffling process, the usage of *DNase* I enzyme seems to be an advantage because the length of fragments obtained by *DNase* I varies greatly with minor changes in fragmentation conditions (Miyazaki, 2002) and it directly leads to diversity in single-gene shuffling.

Cellulases are one of the most important enzymes in the field of biotechnology and have had a great potential to be used in food, brewery and wine, animal feed, textile and laundry, biofuel and pulp and paper industry (Bayer *et al.*, 1994; Bajpai, 1999). Remarkably, the biofuel production field has gained considerable attention as a result of the world's growing energy demand. To date, many cellulase enzymes with various molecular structures and activities have been reported in the literature (Percival Zhang *et al.*, 2006). Among those, endoglucanase (EGs) and cellobiohydrolase (CBHs) have been extensively studied owing to their exo–endo synergistic action in defeating the biomass recalcitrance (Jalak *et al.*, 2012). The working mechanism of EGs is reported as they cleave the cellulose chain internally by hydrolyzing the internal β-1,4-glycosidic bonds with introducing two internal chains (Ulker and Sprey, 1990). Cel12A (EG3), belonging to a fungal GH family 12 enzyme (Sandgren *et al.*, 2001), is one of five EGs, Cel7B(EG1), Cel5A(EG2), Cel12A(EG3), Cel61A(EG4) and Cel45A (EG5) (Ulker and Sprey, 1990; Nidetzky *et al.*, 1994), produced by *Trichoderma reseei*. The 3D structure of Cel12A (EG3) has been already deposited to literature with PDB entry 1h8v, and its active and substrate binding sites have been well described (Sandgren *et al.*, 2001). Here, we present the results of the single-gene shuffling approach performed by Cel12A (EG3) and we demonstrate how the single-gene shuffling approach is a powerful method to obtain novel enzymes, EG3_S1 and EG3_S2, with altered structural and enzymatic properties for industrial applications.

Materials and methods

Amplification of EG3 and cloning

Trichoderma reseei (O00095) β-1,4-endoglucanase gene (EG3) (Okada *et al.*, 1998) was previously cloned into *pPICZαA* vector and kindly provided us by Dr Akcapinar for initial PCR amplifications. The PCR amplification of this EG3 clone was performed with *Pfu* polymerase (Thermo Scientific) by using the F_{EG3} and R_{EG3} primers, and the reaction conditions were set to 96°C for 120 s, 94°C for 30 s, 55°C for 30 s, 72°C for 240 s with 35 cycles and 72°C for 7 min.

For cloning of the shuffled genes, the amplification of shuffled products was also performed with F_{EG3} and R_{EG3} primers and the reaction conditions were set to 96°C for 120 s, 94°C for 30 s, 55°C for 30 s, 72°C for 240 s with 35 cycles and 72°C for 7 min. Then, the amplified genes were cloned into *pPICZαA* vector (Invitrogen, San Diego, CA, USA) with cloning approach. During cloning step, the shuffled DNA products were digested with EcoRI and XbaI restriction enzymes (Thermo Scientific) and ligated into EcoRI and XbaI sites of *pPICZαA* vector with rapid ligation kit (Thermo Scientific). Then, 20 µl of ligation mixtures was transformed into *Escherichia coli* XL-1 Blue cells and they were cultured on low-salt LB supplemented with 25 µg/ml of Zeocin. Positive transformants were selected through colony PCR and they were sequenced for bioinformatics analysis.

*Dnase*I digestion and reassembly PCR

After the amplification of EG3 clone, PCR purification and gel extraction steps (QIAGEN) were performed and the purified EG3 gene was subjected to *DNase*I digestion, carrying out at 37°C for 4 min with 800 ng DNA template. Then, the digested DNA fragments within desired range (50–500 bp) were recovered from 1.2% (w/v) agarose gel and they were subjected to reassembly PCR where the reaction conditions were set to 96°C for 90 s, 94°C for 30 s, 65–41°C with 3°C decrements for 30 s, 72°C for 120 s with 35 cycles and 72°C for 7 min (Arnold and Georgiou, 2003) and the amount of DNA template for this particular reaction was optimized as 450 ng/50 µl. Then, the reassembled DNA product was amplified with F_{EG3} and R_{EG3} primers through PCR whose reaction conditions were set to 94°C for 120 s, 94°C for 30 s, 55°C for 30 s, 72°C for 120 s with 35 cycles and 72°C for 7 min. Finally, the novel shuffled sequences were cloned into *pPICZαA* expression vector as described above.

Transformation and screening

First of all, the shuffled plasmids (first EG3_S1 and second EG3_S2) were linearized with SacI restriction enzymes (Thermo Scientific) to enable its integration to the AOX1 chromosomal locus of *Pichia pastoris* KM71H (aox:ARG4, arg4) (Invitrogen) (Bayram Akcapinar *et al.*, 2015). Then, the competent *P. pastoris* KM71H cells with slow methanol utilizing phenotype were prepared according to the procedure of Wu and Letchworth (2004) and 2 µg of linearized recombinant plasmids was electroporated into competent KM71H cells, using 1.5 kV charging voltage, 25 F capacitance and 200 Ω resistance. This transformation process yielded to formation of nine EG3_S1 and 13 EG3_S2 positive *P. pastoris* colonies in YPD-plate, which were selected against 100 µg/ml Zeocin and the colony PCR was carried out by priming AOX1 site of *pPICZαA* expression vector. Then, Azo-CMC plate assay (0.1% w/v) was performed to select the best transformants amount multiple copies (9 copies of EG3_S1 and 13 copies of EG3_S2). In Azo-CMC plate assay, the positive EG3_S1 (9 copies) and EG3_S2 (13 copies) were plated on 0.1% (w/v) Azo-CMC plates (Hughes *et al.*, 2006) and the plates were incubated at 30°C for 72 h. In every 24 h, the Azo-CMC plates were induced with 300 µl methanol. At the end of 72 h incubation at 30°C, the radii of clear zone around colonies were monitored with both eyes and under UV explosion, and the best transformants, one for EG3_S1 and one for EG3_S2, were selected according to the largest size of clear zones around transformants. During all expression processes, the best EG3_S1 and EG3_S2 transformants were expressed as described in the next section.

Pichia pastoris expression

Two shuffled (EG3_S1 and EG3_S2) and native (EG3_nat) transformants were inoculated into 50 ml YPD-Broth medium and incubated with 250 rpm shaking for 48 h at 30°C. When the optical density at 600 nm reached ~30 arbitrary units, the cells were collected by centrifugation (1534 *g* for 10 min at 4°C) and resuspended in 100 ml of 100 mM potassium phosphate buffer (pH 6.0), 1.34% YNB, 4×10^{-5}% biotin. In every 24 h, the culture was induced with 2% (v/v)

methanol. After the culture was grown at 30°C for 72 h, the cells were collected by centrifugation (1534 *g* for 20 min at 4°C). The supernatants of expression products were collected, concentrated and buffer-exchanged by using Sartocon Micro and Ultrafiltration system with 10 and 100 kDa cutoffs (Sartorius Stedim) with the buffers of enzyme assays, 50 mM sodium acetate (pH 4.7) for CMC and Na-CMC enzyme assays and 50 mM potassium phosphate buffer (pH 6.4) for 4-methylumberriferyl-β-D-cellobioside (4-MUC) enzyme assays.

Gel electrophoresis

After the collection of expression products, the supernatants of cell cultures were analyzed through sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) with 5% (w/v) stacking gel and 12% (w/v) separation gel. Fifteen microliters of the supernatant of expression products were boiled at 95°C for 5 min before loading. After the electrophoresis, SDS–PAGE gel was stained with Coomassie Brilliant Blue R-250.

Enzyme assays

First, the enzyme assays were performed with 4-MUC by using supernatants of expression products and the procedure outlined by Chernoglazov was followed (Chernoglazov *et al.*, 1989). Before performing the enzyme assays toward 4-MUC, samples were incubated with 50 mM potassium phosphate buffer (pH 6.4). The concentrations of enzymes (EG3_S1, EG3_S2 and EG3_nat) and 4-MUC were kept as 0.6 μM and 2 mM, respectively. This kinetic assay was performed at different reaction temperatures (30, 35, 40 and 45°C) for 1 h. The liberated fluorescence of 4-MUC was measured by fluorescence spectrophotometer with an excitation wavelength at 363 nm and emission at 435 nm. For this kinetic assay, the expression of EG3_S1, EG3_S2 and EG3_nat were performed in five batches that three samples from each batch were taken to perform each reaction, and the standard deviations were reported below 10% for each triplet. The maximum value among activity results of enzymes was set to 100% in order to calculate relative activities of the rests.

Also, the kinetic parameters of EG3_S2 and EG3_nat toward 4-MUC were calculated in 50 mM potassium phosphate buffer (pH 6.4) at 40°C. During the calculation of the kinetic constants (V_{max} and K_m) of EG3_S2 and EG3_nat, five different substrate (4-MUC) concentrations between 0.5 and 8 mM were used, and the kinetic constants were calculated by fitting the reaction rate data to the Michaelis–Menten equation. For this assay, the expressions of EG3_S2 and EG3_nat enzymes were performed in two different batches and the sampling was performed as described above. The standard deviations were reported below 10% for each triplet and the relative activity of enzymes was calculated by setting the highest value to 100%.

Moreover, the activity of shuffled and the native enzymes were determined by the 3,5-dinitrosalicyclic acid (DNS) method against 1.0% CMC (w/v) in 50 mM sodium acetate buffer (pH 4.7) as described by Gusakov *et al.* (2011) by keeping the concentration of enzymes as 0.48 μM. In addition to CMC, the activity tests of shuffled and native enzymes were performed toward 1.0% (w/v) Na-CMC that was provided to us by Acıselsan Company (Denizli-TURKEY). During Na-CMC enzyme assays, the enzyme concentrations were kept as 0.66 μM in 50 mM sodium acetate buffer (pH 4.7) and the enzyme concentrations. At the first step of enzyme assays toward CMC and Na-CMC [1.0% (w/v)], the enzymes and substrate were separately pre-incubated at assay temperature (40–60°C) for 5 min. After combination of substrate and enzymes, they were incubated at reaction temperature for 10 min, and DNS reagent was added to reaction mixture and boiled at 95°C for 5 min. Then, the reaction mixture was cooled to room temperature and the absorbance of the supernatant was measured at 540 nm. The percentages of reduced sugar were calculated at all assay temperatures after the subtraction of A_{540} of enzymes and substrate. For the enzyme activity tests toward CMC and Na-CMC, the expressions of shuffled and native enzymes were performed in three batches and the sampling was performed as described above. For each reaction, three samples from each batch were taken and the average value and standard deviation of them were calculated. The standard deviations were reported below 10% for each triplet and the relative activity of enzymes was calculated by setting the highest value to 100%.

Measurement of pH and thermal stability

These novel enzymes were incubated in the temperature range of 30–50°C with 5°C increments for 1 h and their residual activities were measured against 4-MUC substrate. During these measurements, the concentrations of enzyme and substrate were kept at 0.3 μM and 2 mM, respectively. The percentage residual activity of enzymes was calculated by setting the activity of enzymes without thermal incubation to 100%. Moreover, the pH activity profile of shuffled enzymes in 50 mM potassium phosphate buffer was constructed by kinetic assay of shuffled enzymes against 4-MUC within the pH range of 3–7 by keeping the enzyme and substrate concentration as 0.15 μM and 1 mM, respectively. For pH activity profile construction, the expressions of shuffled and native enzymes were done as five different batches and three samples from each batch were taken for each reaction. The average value and standard deviation belonging to five expression batches were calculated. The maximum activity value among all reactions was set to 100% and the relative activities of the rests were calculated according to it. All the results were displayed with normalized standard deviations.

Bioinformatics analysis of shuffled sequences

First of all, the deleted and resembled regions in the shuffled EG3_S1 and EG3_S2 sequences were determined and analyzed by using BLAST tool (Zhang *et al.*, 2000). Through the multiple alignments performed by ClustalW (Larkin *et al.*, 2007), the differences between shuffled and native nucleotide sequence were revealed. Through ExPaSy translate tool (Gasteiger *et al.*, 2003) and ExPaSy-pI/MW tool (Wilkins *et al.*, 1999), the protein sequence and the theoretical molecular weights of shuffled enzymes were predicted, respectively. After this prediction, the multiple alignments between shuffled and native protein sequences were again performed with ClustalW in order to assess the effects of the single-gene shuffling approach at protein level.

In addition to nucleotide/amino acid sequence-based comparison, the structural alignment of EG3_S1 and EG3_S2 with native was performed through TM-Align (Zhang and Skolnick, 2005) and the aligned regions were visualized by visual molecular dynamics (VMD) (Humphrey *et al.*, 1996) after performing the prediction of 3D structures of EG3_S1 and EG3_S2 shuffled enzymes. In order to provide a full structural comparison between shuffled enzymes and native, the missing residues, reported in the crystal structure of EG3_nat (PDB entry: 1h8v) (Sandgren *et al.*, 2001), were also predicted with I-TASSER. Lastly, the flexibility analysis of 3D structures of EG3_S1, EG3_S2 and EG3_nat enzymes was also performed by FlexPred web server (Kuznetsov and McDuffie, 2008) in order to demonstrate how single-gene shuffling process affected the rigidity or flexibility of residues in 3D structures of EG3_S1, EG3_S2 and native enzymes.

Results and discussion

Obtaining shuffled DNA sequences through reassembly PCR

At the first step of single-gene shuffling process, the randomly digested DNA fragments of Cel12A (EG3) gene were subjected to reassembly PCR where nine different annealing temperatures were used to obtain full-length products from digested DNA fragments (100–500 bp) (Joern *et al.*, 2002; Arnold and Georgiou, 2003). As shown in Supplementary Fig. S1, DNA smear where the lengths of DNA fragments were varied between 300 and 1400 bp was formed after reassembly PCR. The amplification of reassembly PCR's product has yielded to two novel shuffled sequences as displayed in Fig. 1; EG3_S1 (743 bp) and EG3_S2 (720 bp).

Analysis of shuffled genes through bioinformatics' tools

The results of ExPaSy translation tool show us that the translation of nucleotide sequences of EG3_S1 and EG3_S2 have yielded to shuffled enzymes in length of 207 and 239 aa, respectively. Even the nucleotide sequence of EG3_S1 was relatively longer than that of EG3_S2, the translation of EG3_S1 nucleotide sequences resulted in shorter protein sequences with respect to EG3_S2 due to introduction of the stop codon (5′-UAA-3′) into 624th position of nucleotide sequences of EG3_S1. In Table I, the length of protein and nucleotide sequences and the molecular weight predictions of EG3_S1, EG3_S2 and EG3_nat enzymes were shown.

As a further analysis, BLAST tool was used to identify the deleted and reassembled regions in EG3_S1 and EG3_S2 shuffled sequences. According to results of BLAST tool, 84, 240 and 419 bp DNA fragments and 80, 220 and 420 bp DNA fragments were reassembled during reassembly PCR and they constituted the nucleotide sequences of EG3_S1 and EG3_S2, respectively. Also, we have performed the multiple alignments and reported the similarity percentages between the nucleotide sequences shuffled and native plasmids as 94.89% in EG3_S1: EG3_nat, 98.82% in EG3_S2: EG3_nat and 94.43% in EG3_S1: EG3_S2 (Supplementary Fig. S2). Thus, we concluded that obtaining these high similarity percentages between native and shuffled sequences was an expected result of single-gene shuffling method.

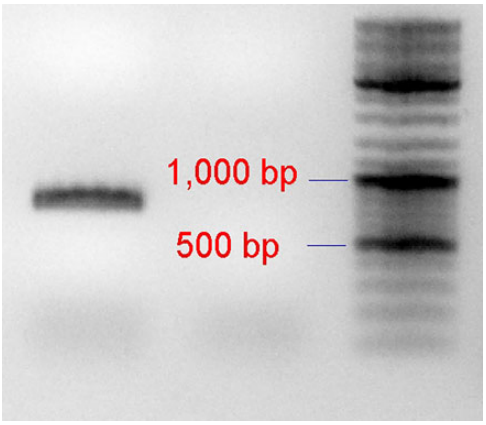

Fig. 1 The amplification results of reassembly PCR.50 and 25 ng of reassembly PCR's products were used as a template for amplification PCR. At the left side of Fig. 1, the amplification result with 50 ng templates was shown. Around 700–750 bp, the thick DNA band was observed. As a gene ladder, Gene Mix Ruler of Fermantas (100–10 000 bp) was used.

Table I. The length of protein and nucleotide sequences, and the molecular weight of EG3_S1, EG3_S2 and EG3_nat

	Protein sequence (aa)	Nucleotide sequence (bp)	Molecular weight (kDa)
EG3_S1	207	743	27
EG3_S2	239	720	39
EG3_nat	234	826	25

In Supplementary Fig. S2, it has been shown that 44 bp fragment between 462nd and 408th positions (native numbering) and 66 bp fragment between 684th and 750th positions (native numbering) were deleted both in nucleotide sequences of EG3_S1 and EG3_S2 during single-gene shuffling process. Although the initial and final positions of reassembled fragments are the same in nucleotide sequences of both EG3_S1 and EG3_S2, the deletions and insertions in nucleotide sequences of EG3_S1 and EG3_S2 (Supplementary Fig. S1) lead to the shift during translation process. Thus, we obtained great variations in amino acid level as shown in Fig. 2. Moreover, the multiple alignment results among protein sequences of shuffled and native enzymes were shown in Fig. 2 and the identity percentages were reported as 63% in EG3_S2: EG3_S1, 75% in EG3_S2: EG3_nat and 59% in EG3_S1: EG3_nat. As clearly seen in Fig. 2, the amino acid sequences of shuffled enzymes are highly conserved between 25th and 141th residues (native numbering) apart from few point mutations. Herein, we clearly observe the significant contributions of single-gene shuffling at the level of protein sequences, especially from the beginning to 25th residue (native numbering) and from 141th residue (native numbering) to end of the shuffled EG3_S1 and EG3_S2 protein sequences.

The structural analysis of EG3_S1 and EG3_S2

In order to assess the effects of a single-gene shuffling approach on 3D structures of shuffled enzymes (EG3_S1 and EG3_S2), TM-alignments have been performed between the predicted 3D structures of shuffled enzymes and native. The TM-scores of aligned regions EG3_S1: EG3_nat and EG3_S2: EG3_nat were reported as 0.786 with 58% identity and 0.918 with 65% identity, respectively. In Fig. 3a and b, the full atomic super-positions of structural alignment results of EG3_S1: EG3_nat and EG3_S2: EG3_nat are shown, respectively. These results tell us that even we observed the great variations in amino acid sequences of EG3_S1 and EG3_S2 with respect to native (Fig. 2), they have not lead to any dramatic change in overall folding pattern of EG3_S1 and EG3_S2 with respect to native, except the addition and/or deletion of few α-helices and β-sheets. In Fig. 3c, this information has been well supported with residue-based structural alignment results. As shown in Fig. 3c, the conformations of aligned residues are still in the range of 5 Å even the disruptive mutations occur. Hence, it becomes possible to conclude that the main reasons behind improved/diminished activities of EG3_S2 and EG3_S1 may be the introduced point mutations and the conservation or loss of critical residues.

Furthermore, a more detailed study about the 3D structures of native and shuffled enzymes has been performed with I-TASSER models, belonging to EG3_S1, EG3_S2 and EG3_nat. First, it has been revealed that there is a change in the number of secondary structure formations such that there are found 2 α-helices and 15 β-strands in native (Sandgren *et al.*, 2001), 13 β-strands and 1 α-helix in EG3_S1, and 16 β-strands and 3 α-helices in EG3_S2. Then, the numbers of residues involved in α-helical conformation have been

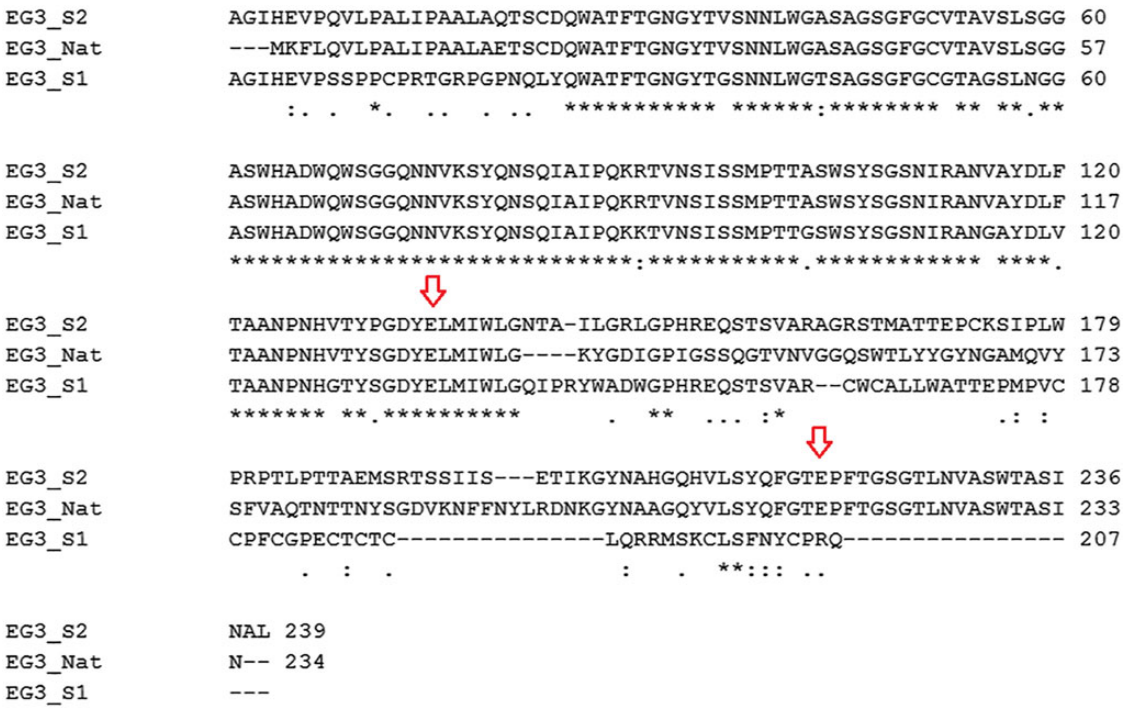

Fig. 2 The Multiple amino acid sequence alignments of shuffled proteins with EG3_nat are shown. '*' indicates the conserved amino acids; ':' indicates polar–polar, charge–charge and hydrophobic–hydrophobic mutations; '.' indicates hydrophobic–charge and charge–polar mutations. Red arrow indicates the invariant catalytic glutamate residues.

calculated for EG3_S1, EG3_S2 and native and it has been found that 40 residues out of 239 in EG3_S2, 18 residues out of 207 in EG3_S1 and 19 residues out of 234 are involved in α-helical conformations. Moreover, we reported some important conformational changes in the 3D structures of EG3_S1 and EG3_S2 based on structural comparisons performed in EG3_S1: EG3_nat and EG3_S2: EG3_nat. The structural comparison of EG3_S1 with native has displayed that C165-D173 in EG3_S1 is aligned with Q158-G168 in native, and two consecutive β-sheets are formed in EG3_S1 while the native has only one β-sheet. We have also observed the shortened β-sheet formations in EG3_S1 between P99-S107 compared with P96-N107 in native. Based on the structural comparison between EG3_S1 and EG3_nat enzymes, we can conclude that the single-gene shuffling approach has led to the formations of shortened or consecutive β-sheets in 3D structure of EG3_S1 by introducing the loop regions.

According to the structural alignment results of EG3_S2: EG3_nat, we have reported that the single-gene shuffling approach has led to more structural variations in 3D structure of EG3_S2 in terms of the addition or the deletion of secondary structures. For instance, the formation of extra α-helices structures has been observed between H4-P11, N39-G46 and R90-W104 in EG3_S2, where N39-G46 and R90-W104 in EG3_S2 are actually aligned to L37-G43 loop region of EG3_nat, and β-sheet and loop formations of EG3_nat, respectively. In addition to formation of extra secondary structures, we have observed the loss of secondary structures in 3D structure of EG3_S2 such that β-sheet formation in T179-D187 of EG3_nat is lost and this region aligns with the loop formation in L184-R192 of EG3_S2.

The conservation and loss of critical residues in EG3_S1 and EG3_S2

Previously, the crystal structure of Cel12A (EG3_nat) has been deposited to literature with a well description of critical residues, sugar binding site, catalytic triad, disulfide bond formation etc. (Sandgren *et al.*, 2001). It has been reported that there occurs a disulfide bond formation between C20 and C48 in EG3_nat (Sandgren *et al.*, 2001). The multiple alignments performed with amino acids sequences of EG3_S1, EG3_S2 and EG3_nat (Fig. 2) have revealed the fact that this particular disulfide bond formation is conserved in EG3_S2 but lost in EG3_S1 upon C20L mutation in protein sequence of EG3_S1. The conformations of these particular disulfide bond formations in EG3_S2 and EG3_nat enzymes are displayed in Fig. 4b and c, respectively and their distances are calculated as 4.56 and 2.04 Å in EG3_S2 and EG3_nat, respectively.

The more detailed analysis of 3D structures of shuffled enzymes have revealed that the single-gene shuffling approach leads to some critical changes in the sugar binding and the catalytic sites of EG3_S1 and EG3_S2. As previously reported, NAG (*N*-acetyl-D-glucosamine) has been found as covalently attached to D180, a part of an N-glycosylation amino acid sequence motif EG3_nat (Sandgren *et al.*, 2001). Through the multiple alignments of protein sequences of EG3_S1, EG3_S2 and EG3_nat, we report that single-gene shuffling process leads to Thr (charge to polar) and Glu (charge to charge) mutations of Asp180 (native numbering) in EG3_S1 and EG3_S2 shuffled enzymes, respectively. Thus, a part of N-glycosylation amino acid sequence motif has been lost in both EG3_S1 and EG3_S2 enzymes.

Moreover, the catalytic site of Cel12A (EG3_nat) has been well described with an indication of two catalytic and invariant glutamate residues, E132 and E216, in crevice (Sandgren *et al.*, 2001). As displayed in Fig. 2, the catalytic E132 residue has been conserved in both EG3_S1 and EG3_S2, yet the catalytic E216 residue has been just conserved in EG3_S2 enzyme. It has been figured out that the conformation of catalytic site has been totally changed in EG3_S1 shuffled enzymes due to E216R replacement. As clearly indicated in Fig. 4d, E216R is not anymore pointing toward the active site of EG3_S1. In

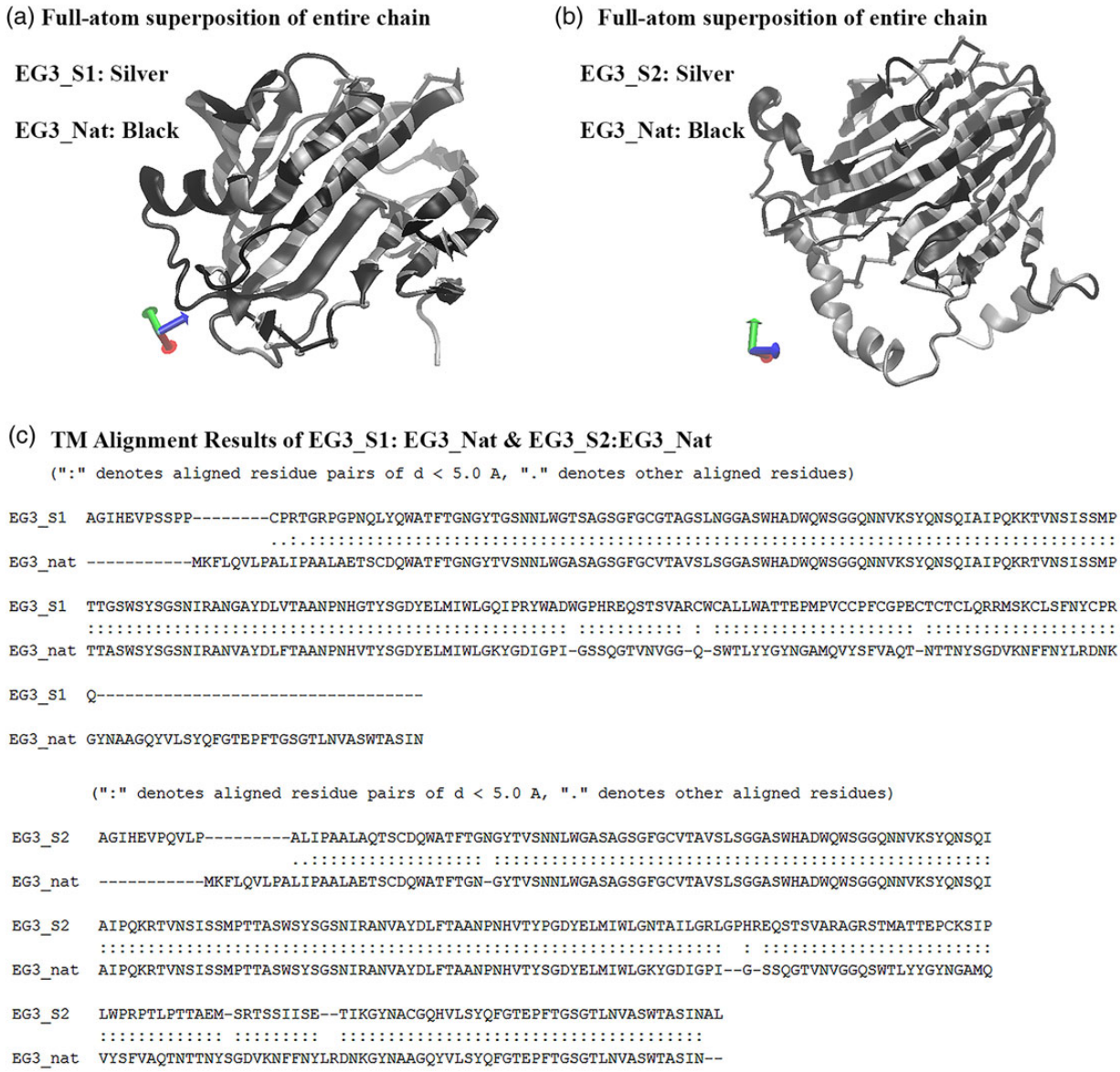

Fig. 3 The structural alignment results of EG3_S1: EG3_nat and EG3_S2: EG3_nat. (**a**) The structural alignment between EG3_S1 and EG3_nat has been performed through TM-Align with 3D models of EG3_S1 and EG3_nat enzymes, and the full-atom super position of entire chain is displayed. (**b**) The structural alignment between EG3_S2 and EG3_nat has been performed through TM-Align with 3D models of EG3_S2 and EG3_nat enzymes, and the full-atom super position of entire chain is displayed. The protein models are drawn in VMD as 'Cartoon' presentation. (**c**) The residue-based structural alignments of EG3_S1: EG3_nat and EG3_S2: EG3_nat are shown. In alignment results, ':' donates the aligned residue pairs smaller than 5 Å; '.' donates the other aligned residues.

addition to two invariant glutamate (E132 and E216) residues in catalytic site of native, there has been reported a third member of catalytic triad, D115 (Sandgren *et al.*, 2001). At the end of the single-gene shuffling approach, we have observed that D115 catalytic residue has been conserved in both EG3_S1 and EG3_S2 (Fig. 4d), and D115 is still pointing toward to the active sites of EG3_S1 and EG3_S2.

A substrate binding mechanism of EG3_nat has been described as there is an enclosed cellulose-binding tunnel containing the side-chains of tryptophan residues (Zou *et al.*, 1999), and these tryptophan rings are fully exposed and found in coordination with a pair of exposed tyrosine residues (Sandgren *et al.*, 2001). Hydrophobic residues, W23, W38, V73 and F218, are found in the upper edge of a crevice in EG3_nat and these residues are completed by the side chains of M170, N167 and I146 (Sandgren *et al.*, 2001). As shown in Fig. 5a, W23, W38, V73 and M170 in native are conserved as W26, W41, V76 and M175 in EG3_S1 yet F218 (native numbering) is lost due to the introduction of stop codon to nucleotide sequence of EG3_S1 via single-gene shuffling. In addition to these mutations, N167T and I146H (native numbering) mutations have been observed in EG3_S1 at the end of single-gene shuffling process. With these point mutations, the orientation of substrate binding site of EG3_S1 has been changed. For instance, T132 and H153 residues in EG3_S1 are not pointing toward to substrate binding site as if native.

Contrary to EG3_S1, we observe the high degree of conservation in substrate binding site of EG3_S2 as displayed in Fig. 5b. For instance, W23, W38, V73 and F218 residues in native structure are

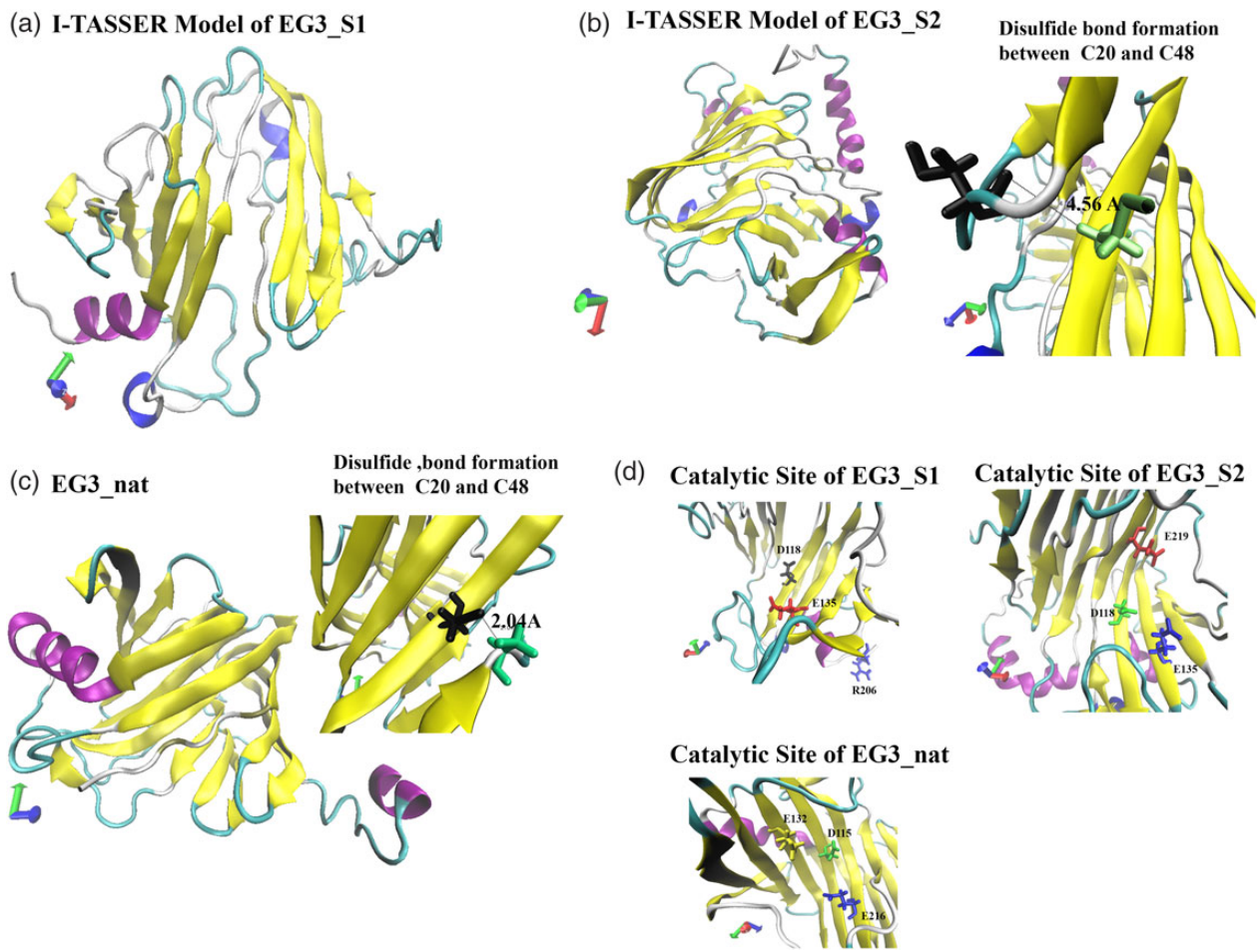

Fig. 4 The 3D models of EG3_S1, EG3_S2 and native enzymes. (**a**) The 3D model of EG3_S1 generated through I-TASSER is displayed. (**b** and **c**) The 3D models of EG3_S2 and EG3_nat generated through I-TASSER are displayed, respectively, with an indication of disulfide bond formation between C20 and C48. (**d**) The catalytic triad of shuffled enzymes and native are displayed. The 3D protein models and catalytic residues in active site of enzymes are drawn in VMD, and presented as 'Cartoon' and 'Licorice' model, respectively.

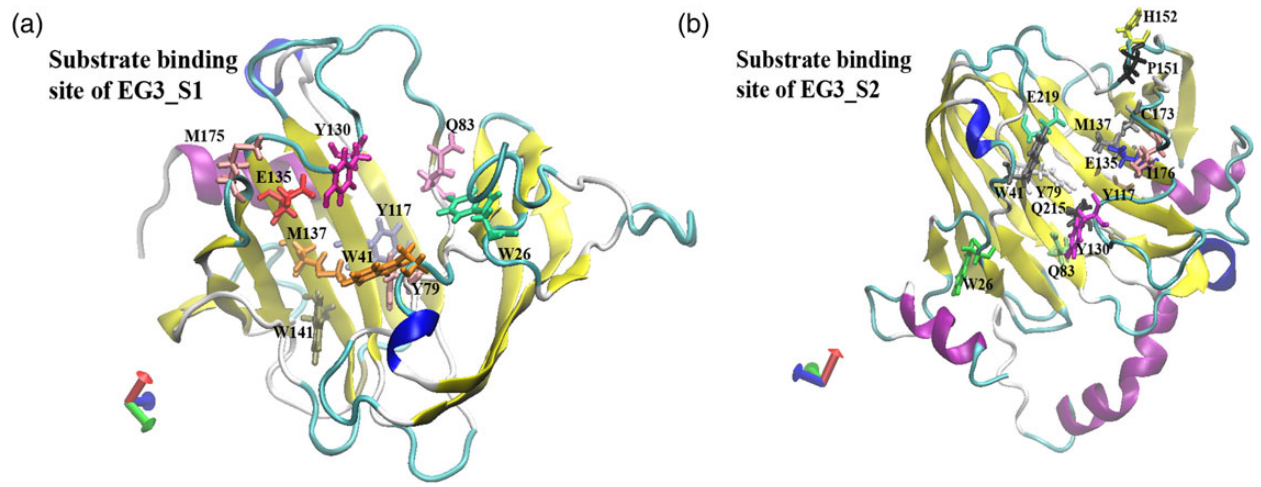

Fig. 5 The substrate binding sites of EG3_S1 and EG3_S2. (**a** and **b**) The substrate binding sites of EG3_S1 and EG3_S2 are displayed with indication critical residues for substrate coordination. The 3D protein models and critical residues of EG3_S1 and EG3_S2 shuffled enzymes are drawn in VMD, and presented as 'Cartoon' and 'Licorice' model, respectively.

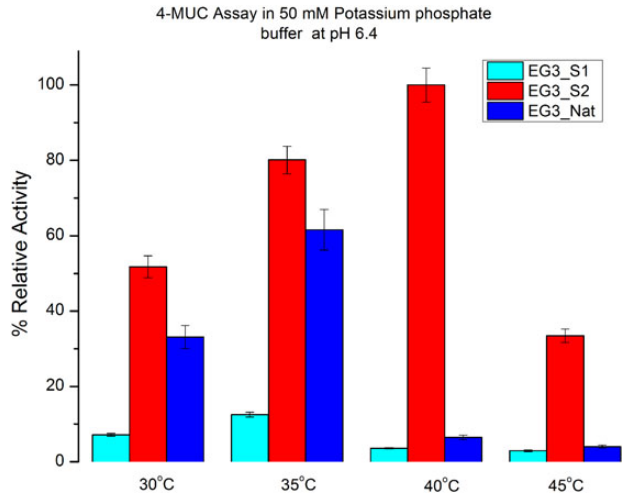

Fig. 6 4-MUC enzyme assay results at 30–45°C in K-phosphate. The kinetic assays have been performed with equal molar enzymes in 50 mM potassium phosphate buffer (pH 6.4). The maximum value of the enzyme activity was set to 100% in order to calculate the relative activity of rests. For each reaction, the standard deviations have been calculated below 10% and the normalized standard deviation values are presented in graphs with error bars.

conserved as W26, W41, V76 and F221 in EG3_S2. Also, we observe some particular mutations in substrate binding site of EG3_S2 such as M170I, N167C and I146H mutations (native numbering). Even these listed residues are mutated in EG3_S2, the single shuffling process has not lead to dramatic change in their conformations and they are still pointing toward the substrate binding site of EG3_S2.

The enzyme activity results of EG3_S1 and EG3_S2 toward 4-MUC and CMC substrates

In order to determine the activity of shuffled enzymes, we have first performed the activity test toward 4-MUC in 50 mM potassium phosphate buffer (pH 6.4). As displayed in Fig. 6, we observed both diminished and increased enzyme activity in EG3_S1 and EG3_S2, respectively. EG3_S1 has displayed a 4.6-fold decrease at 30°C; 4.9-fold decrease at 35°C; 1.8-fold decrease at 40°C and 1.4-fold decrease at 45°C in activity toward 4-MUC, compared with native (Fig. 6). Contrary to EG3_S1, EG3_S2 displayed the improved enzyme activity with respect to native and the reaction rates of EG3_S2 toward 4-MUC were reported as 1.6-fold of the native at 30°C; 1.3-fold at 35°C; 15.4-fold at 40°C and 8.3-fold at 45°C (Fig. 6). Finally, we reported the optimum working temperatures of EG3_S1 and EG3_S2 as 35 and 40°C, respectively.

As mentioned before, some critical changes were observed in the catalytic site of EG3_S1 and EG3_S2 as shown in Fig. 4d. For instance, two invariant catalytic Glutamate residues have been conserved in EG3_S2, but the catalytic E216 (native numbering) was mutated to R206 in EG3_S1. This structural information constituted a basis for us to explain the possible reason behind the diminished and improved enzyme activities of EG3_S1 and EG3_S2 toward 4-MUC, respectively. With the conservation of two invariant catalytic glutamate residues in catalytic site and the high degree of conservation in substrate binding site, the improved enzyme activity of EG3_S2 shuffled enzyme might be explained. Also, it might be possible to explain the diminished enzyme activity of EG3_S1 toward 4-MUC with E206R mutations occurred in catalytic site. Even Glu(E) to Arg(R) mutation is not considered as 'disruptive' mutation because of still being charged, it might lead to great decrease in activity of EG3_S1 toward 4-MUC (Fig. 6). As shown in Fig. 4d, there has happened a dramatic conformational change in the orientation of R206, and it is located on C-terminal and pointing away from the substrate binding site of EG3_S1.

Furthermore, the specific enzyme activity tests of shuffled enzymes and native were performed toward CMC and the results are shown in Fig. 7a. As a result, the diminished and improved enzyme activity of EG3_S1 and EG3_S2, respectively, was observed compared with native. The enzyme activity of EG3_S1 was reported as 0.6-fold of native at 40°C, 0.8-fold at 50°C and 0.6-fold at 60°C. For EG3_S2, the enzyme activity results toward CMC were reported as 1.4-fold of native at 40°C, 1.3-fold at 50°C and 1.1-fold at 60°C. Furthermore, the activity tests of shuffled and native enzymes were performed toward Na-CMC. According to the results displayed in Fig. 7b, we have reported that the activity of EG3_S1 was 0.7-fold of native at 40°C, 0.6-fold of native at 50°C and 0.5-fold of native at 60°C toward Na-CMC. However, the improved enzymatic activities in EG3_S2 enzyme toward Na-CMC were observed at all assay temperatures and its activities were reported as 1.6-fold of native at 40°C, 1.2-fold of native at 50°C and 1.1-fold of native at 60°C. Hence, these results suggested us that EG3_S2 shuffled enzyme seems to be more effective biocatalysts for hydrolysis reaction of Na-CMC than native.

It is also worthy to note that the activity pattern of shuffled and native enzymes toward CMC and Na-CMC is similar to that of toward 4-MUC, e.g. the diminished and improved enzyme activities have been captured in EG3_S1 and EG3_S2 shuffled enzymes compared with native enzyme. The conservation of two invariant catalytic Glu residues in EG3_S2 enzyme and E206 in EG3_S1 enzyme could be again a possible explanation for this similar pattern. Moreover, we want to talk about some possible reasons why EG3_S1 and EG3_S2 shuffled enzymes have displayed different fold of enzyme activities toward different substrates, e.g. 4-MUC, CMC and Na-CMC. As previously mentioned, the types of enzymes assays that performed to assess the activity of shuffled enzymes with respect to native enzyme are different from each other. For instance, the liberated fluorescence of 4-MUC was measured by fluorescence spectrophotometer during hydrolysis reaction, but the reducing sugar in reaction mixture was determined with DNS method against 1.0% CMC (w/v) and 1.0% Na-CMC (w/v) in CMC and Na-CMC enzyme assays. In addition to the differences in measurement methods of enzyme activity, the duration time of reactions, the reaction buffers, and enzyme and substrate concentrations were different from each other, e.g. 4-MUC assay was performed for 1 h in 50 mM potassium phosphate buffer (pH 6.4) by keeping enzyme concentration as 0.6 μM but CMC and Na-CMC enzyme assays were performed 10 min for in 50 mM sodium acetate buffer (pH 4.7) by keeping the enzyme concentrations as 0.48 and 0.66 μM, respectively. When the amount of enzyme was decreased to 0.48 μM to perform enzyme assay toward Na-CMC, we were not able to detect the reducing sugar in reaction mixture as considerable amount. Instead of extending the duration of chemical reaction, the higher amount of enzymes was used to perform the enzyme assay toward Na-CMC by keeping the duration of reaction as 10 min. Thus, the combination of all these details about enzyme assays with a general fact about different substrate specificity of cellulase (Percival Zhang *et al.*, 2006) could provide a plausible explanation why the increase in activity for EG3_S2 is much smaller for CMC and Na-CMC than 4-MUC. In a future computational study, we aim to understand the mechanism of the substrate selectivity of these modified enzymes.

In addition to the enzyme activity tests, the kinetic parameters of novel EG3_S1 and EG3_S2 enzymes were assessed through 4-MUC

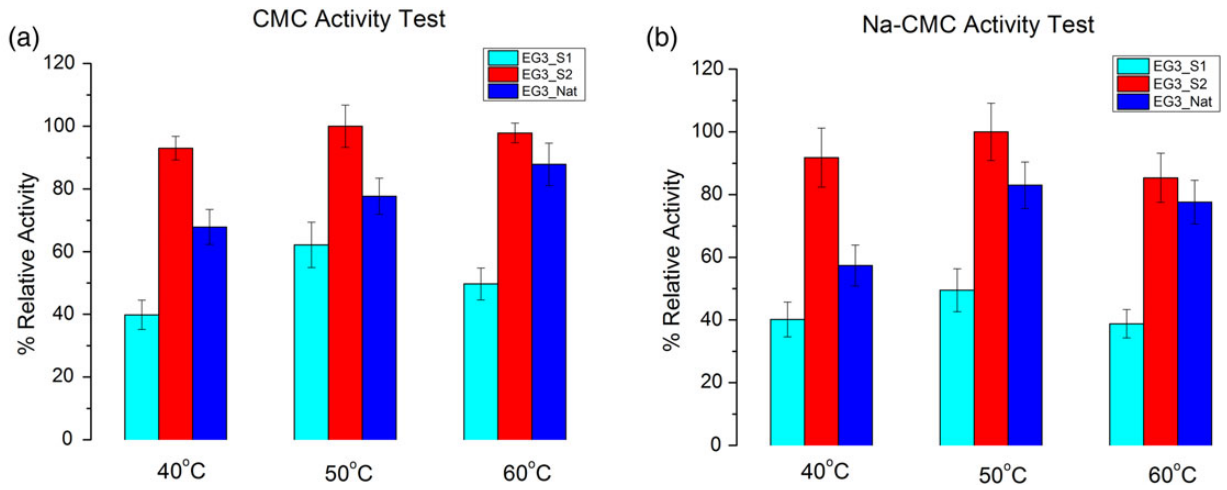

Fig. 7 The CMC and Na-CMC activity profile of EG3_S1 and EG3_S2. (**a** and **b**) The enzymatic activity of EG3_S1 and EG3_S2 has been performed in 50 mM sodium acetate buffer (pH 4.7) within 40–60°C temperatures range toward CMC and Na-CMC substrates. The maximum value of enzyme activity was set to 100% in order to calculate the relative activity of the rests. For each reaction, the standard deviations have been calculated below 10% and the normalized standard deviation values are presented in graphs with error bars.

enzyme assays. During the determination of kinetic parameters of EG3_S1, the decrease in reaction rate was observed upon increase in 4-MUC concentrations. Thus, this pattern could not be fitted to the Michaelis–Menten Equation and the kinetic parameters (V_{max} and K_m) of EG3_S1 could not be calculated. We have reported the kinetic parameters (V_{max} and K_m) of EG3_S2 and EG3_nat enzymes in Table II. According to these results, the K_m value of EG3_S2 was 1.2-fold higher than that of EG3_nat and this result suggests that EG3_S2 shows a lowered binding affinity towards 4-MUC in ground state, compared with native. Contrary to the decrease in binding affinity of EG3_S2, we have reported 12.8-fold improvements in V_{max} of EG3_S2 at 40°C with respect to native.

It has been already demonstrated in literature that the flexibility of residues in enzymes modulates the affinity and specificity of biomolecules through underlying energy landscapes (Chu and Wang, 2014). The increase in flexibilities of residues in biomolecules leads to a decrease in their binding affinity because the binding process with increased flexibility might lead disorder in bound state of protein (Spolar and Record, 1994; Dunker *et al.*, 1998). Hence, increased flexibility of the residues involved in binding that directly contributes to the entropy of protein might result a lowered binding affinity (Grunberg *et al.*, 2006). By taking this provided background in literature as a basis, we have performed the flexibility analysis with 3D models of EG3_S2 and EG3_nat enzymes. According to flexibility results of EG3_S2 and EG3_nat enzyme, there have been observed 2.50, 3.99 and 4.04% increase in flexibilities of E135, Y117 and M137 in EG3_S2 shuffled enzyme with respect to native (Supplementary Fig. S3a and b). Previously, the importance of Y117 and M137 has been described as they provide a support to the catalytic E132 residue in EG3_nat enzyme. Hence the increased flexibility around catalytic cleft of EG3_S2 enzyme and their contribution to entropy might lead to lowered binding affinity toward 4-MUC with respect to native. Comparing the dimensions of substrate binding site could make another possible explanation for lower K_m of EG3_S2 compared with native. This comparison shows us that the dimensions of substrate binding site of native (35 Å long, 8 Å wide and 15 Å deep) are increased to 38 Å long, 13 Å wide and 17 Å deep in EG3_S2 (Sandgren *et al.*, 2001). Due to this enlargement of the substrate binding site in EG3_S2, the entrance and exit and of substrate might become easier, also result in decrease in its binding affinity by increasing K_m value.

Table II. The kinetic parameters of EG3_nat and EG3_S2

	V_{max} (RFU/s)	K_m (mM)	k_{cat} ($\times 10^6$ s^{-1})	k_{cat}/K_m (s^{-1} μM^{-1})
EG3_Nat	0.706	5.814	0.235	40.5
EG3_S2	9.001	7.261	1.924	265.0

Along with the increase in K_m value of EG3_S2, we observed that much higher increase in V_{max} value of EG3_S2 enzyme compared with EG3_nat enzyme. Although increase in K_m value of enzymes might be explained by lowered binding affinity of the ground state of the substrate, the changes in V_{max} value of enzyme should be related to the binding affinity of the intermediate state of the reaction (Fersht, 1985). Given the changes occur within the flexibilities of the residues that are closed to catalytic E135 (EG3_S2 numbering) and the dimensions in the substrate binding site of EG3_S2, we have reported that novel EG3_S2 enzyme has been able to optimize the interactions of the intermediate state of the 4-MUC hydrolysis reaction better than EG3_nat enzyme. Thus, we have reported the higher V_{max} value of EG3_S2 toward 4-MUC compare with native in Table II. Overall, these changes that occur in the kinetic constants of EG3_S2 have resulted in a 6.5-fold increase in the k_{cat}/K_m ratio of the native enzyme toward 4-MUC substrate as shown in Table II. As a result, we could make a conclusion that EG3_S2 could be used as more efficient biocatalyst by replacing its native (EG3_nat) counterpart.

The thermal stability and pH activity profiles of EG3_S1 and EG3_S2 shuffled enzymes

Furthermore, the thermal stability and pH activity profiles of novel EG3_S1 and EG3_S2 enzymes were investigated through 4-MUC enzyme assays. As shown in Fig. 8a, we have reported the improved

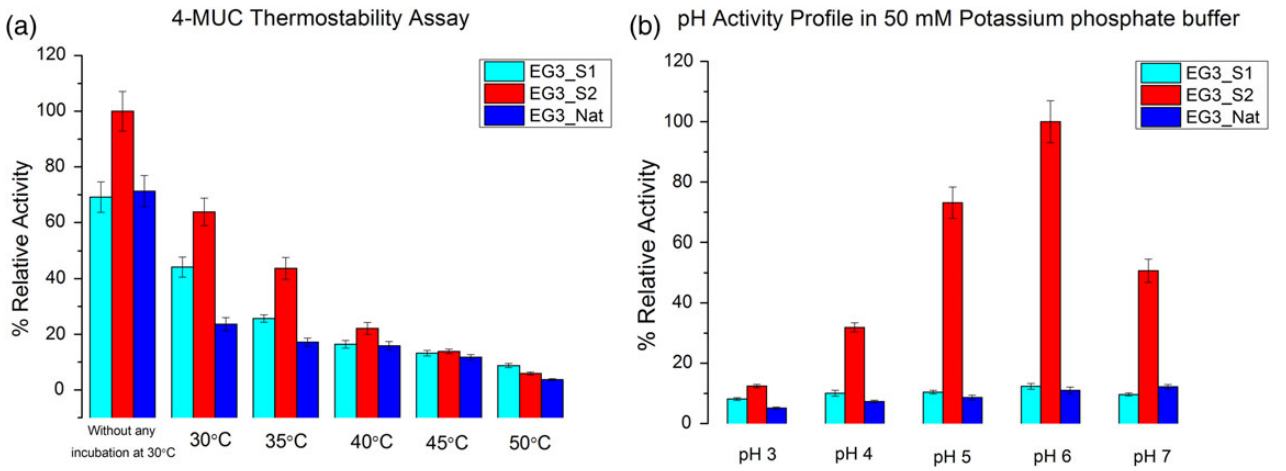

Fig. 8 Thermostability and pH activity profile of shuffled enzymes. (**a**) The thermal stabilities of EG3_S1 and EG3_S2 shuffled enzymes have been determined by measuring their residual activity toward 4-MUC at 30°C after 1 h incubation at the assay temperatures. Without any incubation, the reaction rates of EG3_S1, EG3_S2 and native were measured at 30°C. Among all enzyme activity values, the maximum value was set to100% in order to calculate the relative activity of the rests. (**b**) The pH activity profiles of EG3_S1 and EG3_S2 enzymes were determined in 50 mM potassium phosphate buffer at 35°C within pH 3–7 ranges. The reaction rates of EG3_S1, EG3_S2 and native enzymes were measured toward 4-MUC. Among all enzyme activity values, the maximum value was set to100% in order to calculate the relative activity of the rests. For each reaction, the standard deviations have been calculated below 10% and the normalized standard deviation values are presented in graphs with error bars.

thermal stabilities of EG3_S1 and EG3_S2 with respect to native at all assay temperatures. After 1 h incubation at each assay temperature, the thermal stability of EG3_S1 was reported as 1.9-fold at 30°C; 1.5-fold at 35°C; 1.0-fold at 40°C; 1.11-fold at 45°C and 2.3-fold at 50°C of native (Fig. 8a). For EG3_S2, the thermal stability was reported as 2.7-fold at 30°C; 2.5-fold at 35°C; 1.4-fold at 40°C; 1.20-fold at 45°C and 1.6-fold at 50°C of native (Fig. 8a). According to the thermal stability results of EG3_S1 and EG3_S2, we could make a conclusion that the deleted regions in nucleotide sequences of EG3_S1 and EG3_S2 did not lead to the loss of critical residues, having a control over thermal stability (Supplementary Fig. S2). Furthermore, the pH activity profiles of EG3_S1 and EG3_S2 were constructed toward 4-MUC (Fig. 8b). Whereas the optimum pH of the native was pH 7, the single-gene shuffling process has led to alterations in optimum pH values of EG3_S1 and EG3_S2, which were recorded as pH 6. It was also worthy to note that EG3_S2 displayed better enzymatic activity at all pH values with respect to EG3_S1 and EG3_nat, and it seems to be a better biocatalyst by replacing native enzyme at various pH values.

Conclusion

In this work, we present the results of the single-gene shuffling approach performed with Cel12A (EG3) gene from *T.reseei* and we have proved how the single-gene shuffling approach is as a powerful technique to generate novel enzymes, EG3_S1 and EG3_S2 with altered structural and enzymatic propertied. As a result of single-gene shuffling process, the novel EG3_S1 and EG3_S2 enzymes displayed 59 and 75% identity in protein sequence with respect to native and we have captured the great variation in their structures with respect to native enzyme. According to enzyme assay results toward 4-MUC, we have reported that the minimum enzyme activity of EG3_S1 was 4.9-fold of native at 35°C and the maximum enzyme activity of EG3_S2 was 15.4-fold of native at 40°C. Furthermore, the diminished enzyme activity of EG3_S1 was reported within range of 0.6- to 0.8-fold of native and within range of 0.5- to 0.7-fold of native toward CMC and Na-CMC, respectively. For EG3_S2, we have reported the improved enzyme activity toward CMC and Na-CMC within range of 1.1- to 1.4-fold of native and within range of 1.1- to 1.6-fold of native, respectively.

Furthermore, we have reported that the single-gene shuffling process has led to generation of shuffled EG3_S1 and EG3_S2 enzyme with improved thermal stability compare with native enzyme at all assay temperatures. The improved thermal stability of EG3_S1 and EG3_S2 shuffled enzymes compare with native showed us that the critical part of Cel12A (EG3) gene controlling over thermal stability was conserved during the single-gene shuffling process. Also, the kinetic parameters (V_{max} and K_m) of EG3_S2 and EG3_nat were calculated toward 4-MUC. Based on the comparison between kinetic parameters of EG3_S2 and EG3_nat, we have reported 6.5-fold increase in the k_{cat}/K_m ratio of EG3_S2 with respect to native and this result suggests us that EG3_S2 could be used as more efficient biocatalyst by replacing its native counterpart.

As a conclusion, we have demonstrated in this work that the single-gene shuffling approach is a powerful technique to generate the novel enzymes, EG3_S1 and EG3_S2, with altered structural and enzymatic properties with respect to native enzyme. We believe that this work could provide a basis for future studies that the further characterization would be performed through molecular dynamic simulations to understand the main reason behind the improved/diminished enzymatic activities and kinetic parameter, thermal stabilities etc.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

We have special thanks to Dr E. Timuçin for critically reading this manuscript.

Funding

The Scientific and Technological Research Council of Turkey supported this work with grant number (112T901).

References

Akbulut,N., Tuzlakoglu Ozturk,M., *et al.* (2013) *J. Biotechnol.*, **164**, 123–129.

Arnold,F.H. and Georgiou,G. (2003) *Directed Evolution Library Creation: Methods and Protocols*. Humana Press, Totowa, NJ.

Bajpai,P. (1999) *Biotechnol. Prog.*, **15**, 147–157.

Bayer,E.A., Morag,E., *et al.* (1994) *Trends Biotechnol.*, **12**, 379–386.

Bayram Akcapinar,G., Venturini,A., *et al.* (2015) *Protein Eng. Des. Sel.*, **28**, 127–135.

Chernoglazov,V.M., Jafarova,A.N., *et al.* (1989) *Anal. Biochem.*, **179**, 186–189.

Chockalingam,K., Chen,Z., *et al.* (2005) *Proc. Natl Acad. Sci. USA*, **102**, 5691–5696.

Chu,X. and Wang,J. (2014) *PLoS Comput. Biol.*, **10**, e1003782.

Crameri,A., Whitehorn,E.A., *et al.* (1996) *Nat. Biotechnol.*, **14**, 315–319.

De Groeve,M.R., De Baere,M., *et al.* (2009) *Protein Eng. Des. Sel.*, **22**, 393–399.

Dunker,A.K., Garner,E., *et al.* (1998) *Pac. Symp. Biocomput.*, 473–484.

Fersht,A. (1985) *Enzyme Structure and Mechanism*, 2nd edn. W.H. Freeman, New York.

Gasteiger,E., Gattiker,A., *et al.* (2003) *Nucleic Acids Res.*, **31**, 3784–3788.

Grunberg,R., Nilges,M., *et al.* (2006) *Structure*, **14**, 683–693.

Gusakov,A.V., Kondratyeva,E.G., *et al.* (2011) *Int. J. Anal. Chem.*, **2011**, 283658.

Hughes,S.R., Riedmuller,S.B., *et al.* (2006) *Proteome Sci.*, **4**, 10.

Humphrey,W., Dalke,A., *et al.* (1996) *J. Mol. Graph*, **14**, 33–38, 27–38.

Jackel,C., Kast,P., *et al.* (2008) *Annu. Rev. Biophys.*, **37**, 153–173.

Jalak,J., Kurasin,M., *et al.* (2012) *J. Biol. Chem.*, **287**, 28802–28815.

Joern,J.M., Meinhold,P., *et al.* (2002) *J. Mol. Biol.*, **316**, 643–656.

Kurtzman,A.L., Govindarajan,S., *et al.* (2001) *Curr. Opin. Biotechnol.*, **12**, 361–370.

Kuznetsov,I.B. and McDuffie,M. (2008) *Bioinformation*, **3**, 134–136.

Larkin,M.A., Blackshields,G., *et al.* (2007) *Bioinformatics*, **23**, 2947–2948.

Miyazaki,K. (2002) *Nucleic Acids Res.*, **30**, e139.

Nidetzky,B., Steiner,W., *et al.* (1994) *Biochem. J.*, **298** Pt 3, 705–710.

Okada,H., Tada,K., *et al.* (1998) *Appl. Environ. Microbiol.*, **64**, 555–563.

Percival Zhang,Y.H., Himmel,M.E., *et al.* (2006) *Biotechnol. Adv.*, **24**, 452–481.

Sandgren,M., Shaw,A., *et al.* (2001) *J. Mol. Biol.*, **308**, 295–310.

Song,J.K. and Rhee,J.S. (2000) *Appl. Environ. Microbiol.*, **66**, 890–894.

Spolar,R.S. and Record,M.T., Jr. (1994) *Science*, **263**, 777–784.

Stemmer,W.P.C. (2002) *Biotechnol. Bioprocess Eng.*, **7**, 121–129.

Ulker,A. and Sprey,B. (1990) *FEMS Microbiol. Lett.*, **57**, 215–219.

Wilkins,M.R., Gasteiger,E., *et al.* (1999) *Methods Mol. Biol.*, **112**, 531–552.

Wu,S. and Letchworth,G.J. (2004) *Biotechniques*, **36**, 152–154.

Yano,T., Oue,S., *et al.* (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5511–5515.

Zhang,Y. and Skolnick,J. (2005) *Nucleic Acids Res.*, **33**, 2302–2309.

Zhang,Z., Schwartz,S., *et al.* (2000) *J. Comput. Biol.*, **7**, 203–214.

Zou,J., Kleywegt,G.J., *et al.* (1999) *Structure*, **7**, 1035–1045.