*Protein Engineering, Design & Selection*, 2015, vol. 28 no. 5, pp. 127–135
doi: 10.1093/protein/gzv012
Advance Access Publication Date: 16 March 2015
Original Article

OXFORD

Original Article

# Modulating the thermostability of Endoglucanase I from *Trichoderma reesei* using computational approaches

**Gunseli Bayram Akcapinar[1], Alessandro Venturini[2], Pier Luigi Martelli[3,4], Rita Casadio[3,4], and Ugur O. Sezerman[5,*]**

[1]Sabanci University, Biological Sciences and Bioengineering, 34956 Istanbul, Turkey, [2]Institute for the Organic Synthesis and Photoreactivity, National Research Council of Italy, Via P. Gobetti 101, 40129 Bologna, Italy, [3]Biocomputing Group, Department of Biology, University of Bologna, CIRI-Health Science and Technology, 40126 Bologna, [4]Department of Computer Science and Engineering, University of Bologna, 40127 Bologna, Italy, and [5]Acibadem University, Department of Medical Informatics, Istanbul, Turkey

*To whom correspondence should be addressed. E-mail: ugur@sabanciuniv.edu
Edited by Gideon Schreiber

Received 19 November 2014; Revised 3 February 2015; Accepted 4 February 2015

## Abstract

In the last decades, effective cellulose degradation became a major point of interest due to the properties of cellulose as a renewable energy source and the widespread application of cellulases (the cellulose degrading enzymes) in many industrial processes. Effective bioconversion of lignocellulosic biomass into soluble sugars for ethanol production requires use of thermostable and highly active cellulases. The library of current cellulases includes enzymes that can work at acidic and neutral pH in a wide temperature range. However, only few cellulases are reported to be thermostable. In order to alleviate this, we have performed a hybrid approach for the thermostabilization of a key cellulase, Endoglucanase I (EGI) from *Trichoderma reesei*. We combined *in silico* and *in vitro* experiments to modulate the thermostability of EGI. Four different predictive algorithms were used to set up a library of mutations. Three thermostabilizer mutations (Q126F, K272F, Q274V) were selected and molecular dynamics simulations at room temperature and high temperatures were performed to analyze the effect of the mutations on enzyme structure and stability. The mutations were then introduced into the *endoglucanase 1* gene, using site-directed mutagenesis, and the effect of the mutations on enzyme structure and stability were determined. MD simulations supported the fact that Q126F, K272F and Q274V mutations have a thermostabilizing effect on the protein structure. Experimental studies validated that all of the mutants exhibited higher thermostability compared with native EGI albeit with a decrease in specific activity.

**Key words:** cellulase, endoglucanase 1, molecular dynamics, protein engineering, *Trichoderma reese*

## Introduction

Cellulose is the most abundant polymer on earth. The ability of cellulose as a renewable source of energy was well known after the detection of cellulases, the major cellulolytic enzymes possessing the ability of efficiently degrading cellulose to produce biofuel. Nearly all of the cellulases can work at neutral and acidic pH, and at a wide temperature range, but only few of them are thermostable (Bhat, 2000; Rastogi *et al.*, 2010). Some of the industrial applications of cellulase require heat stable enzymes that can withstand the elevated temperatures of the harsh industrial processes, while preserving their activity. Lignocellulosic biomass conversion is an important topic: one of the bottlenecks of bioconversion of lignocellulosic biomass is

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com


Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

the need of thermostable and highly active cellulases for the total conversion of cellulose to soluble sugars for ethanol production (Demain *et al.*, 2005). Lignocellulosic biomass is a highly complex substrate and its bioconversion requires synergistic action of different cellulase components (Gao *et al.*, 2011; Zhou *et al.*, 2011). Moreover, due to the complexity of this substrate, a pretreatment at high temperatures is needed to increase the accessibility of the lignocellulosic biomass to the enzyme attack. Thermostable enzymes are therefore key to effective lignocellulosic biomass conversion. Another limiting step in lignocellulosic biomass conversion is the cost of this process. Use of thermostable enzymes would: (i) decrease the cost of this process since with high temperatures, higher process rates would be achieved; (ii) more substrate would be converted to simple sugars; (iii) unwanted side reactions of the process would be eliminated at high temperatures and (iv) the enhanced stability of the cellulase would also increase the prolonged use of the enzyme (Demain *et al.*, 2005; Heinzelman *et al.*, 2009b; Gao *et al.*, 2011).

The filamentous fungus *Trichoderma reesei* is the key producer of the cellulases. Many commercially available cellulase formulations are derived from *T.reesei*. Endoglucanases, exoglucanases and β-glucosidases of *T.reesei* work synergistically for the conversion of cellulose to glucose. Endoglucanase I (EGI, CEL7B), the major endoglucanase of the celluloytic system of *T.reesei*, accounts for 5–10% of the total amount of cellulase produced by the organism. EGI is a member of the Glycosyl Hydrolase family 7 (based on the classification of the CAZY database (Henrissat, 1991)) and it is known to catalyze the hydrolysis of the cellulose chain with a net retention of the anomeric configuration. It contains a cellulose binding domain (CBD) and an *o*-glycosylated flexible linker which connects the CBD to the catalytic domain. The three-dimensional structure of the catalytic core domain of EGI was determined in an extended, open substrate-binding cleft which originates after the deletion of the active site tunnel forming loops in its homologous cellobiohydrolase, CBHI from the same organism. This open active site architecture enables the enzyme to function as an endoglucanase (Kleywegt *et al.*, 1997). For lignocellulosic mass bioconversion, rate limiting enzymes were known to be exoglucanases. Interestingly, only few studies addressed the problem of making Cellobiohyrolase I and II of *T.reese*i thermostable. However, none is available on EGI, which is also an important enzyme for biomass conversion in the same organism (Heinzelman *et al.*, 2009a,b; Lantz *et al.*, 2010).

The problem of conferring thermostability to a protein has been often addressed. However, no single universal stabilization rule or mechanism can be applicable to all proteins. Depending on the different proteins, electrostatic–hydrophobic interactions, hydrogen bonding, disulfide bonding, overall rigidity and compactness, the presence of stable α helices and loops, of certain amino acids, glycosylation and metal binding sites are included among the factors that modulate their thermostability (Vetriani *et al.*, 1998; Kumar *et al.*, 2000; Sterner and Liebl, 2001; Li *et al.*, 2005). Modification of the properties of a protein could be achieved through the use of different approaches such as directed enzyme evolution or rational design (Arnold and Moore, 1997; Hellinga, 1997; Zhao and Arnold, 1997; Jäckel *et al.*, 2008; Arnold, 2009).

Our strategy for the thermostabilization of EGI involves first the use of a combination of predictive algorithms and molecular dynamics simulations to compute which lateral side chain substitution is more suited to increase the protein thermostability. The double check on the effectiveness of our predictions, done with two completely independent computational methods, allows a better definition of the experimental strategy. Indeed, we use both machine learning-based methods for the selection of the more suited substitutions and verified them with a traditional approach concerning high-temperature MD simulations to induce protein unfolding.

The computational step allowed the selection of three mutations (Q126F, K272F, Q274V) that were then introduced into the *endoglucanase 1* gene, using site-directed mutagenesis. The effect of the mutations on enzyme structure and stability were determined. Experimental studies validated that all of the mutants exhibited higher thermostability compared with native EGI, and for one of them (Q274V) we find that activity of the thermostable protein is similar to the native mesophilic enzyme.

## Materials and methods

### Computational methods

Destabilizing/stabilizing protein variations were first detected by means of predictive tools previously described and aiming at computing the free energy change after residue substitutions. For this purpose, four different servers are used for the prediction of stabilizing and destabilizing mutations: I mutant (http://folding.biofold.org/i-mutant/i-mutant2.0.html), I-Mutant Suite (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) (Capriotti *et al.*, 2005a, b), PoPMuSiC (http://babylone.ulb.ac.be/popmusic/) (Dehouck *et al.*, 2009) and SDM (http://www-cryst.bioc.cam.ac.uk/~sdm/sdm.php). Three protein stabilizing mutations were selected as candidates for MD simulations and site-directed mutagenesis.

Molecular dynamics simulations were performed on EGI protein (PDB ID: 1EG1) and its mutants with NAMD v2.9 program (Phillips *et al.*, 2005) CHARMM27 all-atom force field were used (MacKerell *et al.*, 1998), and water was described using the TIP3P model (Jorgensen *et al.*, 1983). This strategy takes advantage of the kinetic relationship between the inactivation rate constant ($k_{ina}$) and the temperature ($T$):

$$k_{ina} = \frac{k_B T}{h} \exp\left(\frac{\Delta G_{ina}}{RT}\right).$$

To avoid that folding and unfolding occur on a too long time scale for the current computational power, a more traditional computational approach was used where the MD simulations were carried out at high temperatures to study the stability of the enzyme (Day and Daggett, 2005a,b; Phillips *et al.*, 2005; Eleftheriou *et al.*, 2006; Gao *et al.*, 2009).

EGI protein (5228 atoms) was initially solvated in a water box of dimensions 65 Å × 65 Å × 80 Å ($x$, $y$, $z$) with the minimum solute-wall distance of 10 Å and ionized with 16 $Na^+$ atoms to neutralize the system (net charge of $3.11e^{-6}$). In total, 12 229 water molecules were added. Single mutations were built from this initial system using Visual Molecular Dynamics (VMD) software.

Minimization for 20 000 steps and initial 500 ps of simulation at constant volume, NVT ensemble, equilibrated all the systems. Then, simulations were run in the NPT ensemble for 10 ns. The simulations were carried out at 310 K with constant pressure (1.0123 bar) using the Langevin piston with an oscillation period of 200 fs. Electrostatic interactions were calculated with PME (Darden *et al.*, 1993; Essmann *et al.*, 1995). Another temperature (550 K) to induce unfolding was applied. Starting coordinates of additional runs at the higher temperature value were obtained by extending the equilibration run and taking the coordinates after 10 ns. These high-temperature MD simulations were carried out at constant temperature and volume (NTV ensemble, time step 1 fs) and interrupted when protein had completely lost the original features (~5 ns).

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

## Microorganisms, enzymes and chemicals

*Trichoderma reesei* (QM9414) *endoglucanase 1* gene previously cloned and expressed in *Pichia pastoris* (Bayram Akcapinar *et al.*, 2012) was used in the site-directed mutagenesis experiments. *Escherichia coli* XL1-Blue was used as host for the propagation of the plasmids and subcloning. *Pichia pastoris* KM71H (aox1::ARG4, arg4) (Invitrogen, San Diego, USA) was used for the recombinant protein expression. pPICZαA (Invitrogen, San Diego, USA) plasmid was used for cloning and protein expression. pPICZαA plasmid contains an α factor secretion signal for extracellular secretion of cloned gene products, a Zeocin resistance gene for selection in *E.coli* and *P.pastoris*, and an alcohol oxidase (AOX) promoter for methanol-induced expression of the cloned gene (Cregg *et al.*, 1985; Cereghino and Cregg, 2000). *Pwo* polymerase (Roche) and rapid ligation kit (Fermentas) were used for cloning purposes. *Taq* polymerase (Qiagen) was used for colony PCR. 4-Methylumbelliferyl-β-D-cellobioside (4-MUC) was purchased from RPI Corp. All chemicals used were of analytical grade.

## Site-directed mutagenesis

The overlap PCR extension method was applied to introduce the point mutations Q126F, K272F and Q274V. Overlap PCR extension primers were designed according to the literature (Vallejo *et al.*, 1994) (Table I).

## Cloning

The mutations were introduced to a codon optimized (for *P.pastoris*) endoglucanase 1 synthetic gene (*egl1*) carrying EcoRI and XbaI restriction sites on each arm. Protein sequence of the synthetic gene is identical to *T.reesei* endoglucanase 1 (GenBank Accession Number: M15665) but EcoRI and XbaI sites were added to the gene flanking regions. The mutant genes were digested with EcoRI and XbaI and re-ligated into EcoRI–XbaI site of pPiczαA vector using rapid ligation kit. *egl1* and the mutants were subcloned in *E.coli* cells. *Escherichia coli* cells were cultured on law salt LB plates in the presence of 25 μg/ml Zeocin. Zeocin positive colonies were selected and colony PCR with 5′ AOX and 3′ AOX primers was performed. Colony PCR positive colonies were selected and recombinant plasmids were isolated using MiniPrep kit (Qiagen). The fidelity of the constructs was confirmed by sequencing of the mutant plasmids.

## Transformation and screening

The recombinant plasmids pPiczαA-*egl1-Q126F*, pPiczαA-*egl1-K272F* and pPiczαA-*egl1-Q274V* were linearized with SacI before transformation. This process resulted in the stable integration of one or multiple copies of the linearized vector at 5′ AOX1 chromosomal locus of *P.pastoris* KM71H by homologous recombination. All the obtained transformants were MutS (slow methanol utilizing phenotype). Competent *P.pastoris* KM71H cells were prepared according to a procedure combining chemical transformation and electroporation (Wu and Letchworth, 2004). Approximately 1 μg linearized recombinant plasmid was mixed with competent KM71H cells. The mixture was immediately transferred to a pre-chilled 0.2 cm electroporation cuvette and incubated on ice for 5 min. About 1 ml of ice-cold 1 M sorbitol was immediately added to the cuvette after electroporation. The charging voltage, capacitance and resistance were 1.5 kV, 25 F and 200 Ω, respectively. The transformation mixture was spread onto YPD plates containing 100 μg/ml Zeocin. The plates were incubated at 28°C until the appearance and growth of colonies (~3 days). Zeocin positive colonies were selected and colony PCR with 5′ AOX and 3′ AOX primers was performed. Colony PCR positive colonies were selected.

Multicopy transformants were selected on BMM-agar plates containing blue colored Azo-CMC (carboxymethylcellulose) as a substrate. More active transformants expressing the recombinant enzymes were selected according to the relative radii of the clear zones around the colonies. When the enzyme is active, it degrades Azo-CMC and clear zones are produced around the colonies as a result of enzymatic hydrolysis.

## Expression of recombinant *P.pastoris* strains

The colonies selected on Azo-CMC agar plates (for Q126F mutant Colony 9, for K272F mutant Colony 15 and for Q274V mutant colony 31) were inoculated into 500 ml BMG medium (100 mM potassium phosphate, pH 6.0, 1.34%YNB, $4 \times 10^{-5}$% biotin and 1% glycerol) and shaken (250 rpm) at 29°C overnight. When $OD_{600}$ reached 10, the cells were collected by centrifugation (3000 × *g*, 5 min). The cell pellet was resuspended in 150 ml BMM medium (BMG with 0.5% methanol instead of 1% glycerol). The culture was grown for ~24 to 48 h at 28°C. Methanol was added to a final concentration of 0.5% (v/v) at every 24 h. Cell culture supernatants were collected every 24 h for 48 h. Supernatants were filtered, buffer exchanged to 50 mM sodium acetate buffer (pH 5) and concentrated using Sartocon Micro and Ultrafiltration System (Sartorius-Stedim). Sartocon Slice 200 HydroSart membranes with 0.45 μm cutoff and polyether sulfone membranes with 100 and 10 kDa cutoff were used.

## Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

Collected cell culture supernatants were analyzed by sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS-PAGE) with a 5% (w/v) stacking gel and 12% (w/v) separating gel. About 15–20 μl of the supernatant was loaded into each well of the gel. After electrophoresis,

**Table I.** Sequences of primers used in Overlap Extension PCR reaction

| Primer name | Sequence |
|---|---|
| Forward *egl1* | 5′-CCG*GAATTC*CAGCAACCGGGTACCAGCACC-3′ |
| Reverse *egl1* | 5′-CCG*TCTAGA*GCAAGGCATTGCGAGTAGTAG-3′ |
| Forward *Q126F* | 5′-TTGAAGTTGAACGGTTTTGAGTTGTCTTTCGAC-3′ |
| Reverse *Q126F* | 5′-GTCGAAAGACAACTCAAAACCGTTCAACTTCAA-3′ |
| Forward *K272F* | 5′-GTTTCCATCACTAGATTTTACCAGCAGAACGGT-3′ |
| Reverse *K272F* | 5′-ACCGTTCTGCTGGTAAAATCTAGTGATGGAAAC-3′ |
| Forward *Q274V* | 5′-TGGTTTCCATCACTAGAAAGTACGTTCAGAACGGTGTTG-3′ |
| Reverse *Q274V* | 5′-CAACACCGTTCTGAACGTACTTTCTAGTGATGGAAACCA-3′ |

Sequences in italics are the restriction sites introduced for directional cloning. Underlined sequences are the mutation sites for the mutants.

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

the gel was stained with Coomassie Brilliant Blue R-250. All gel photographs were documented using Gel-Doc (BioRad).

### Purification of recombinant proteins

EGI and the mutant proteins were purified using regenerated amorphous cellulose (RAC) as an affinity chromatography matrix. Batch affinity purifications of both enzymes were obtained. Hong *et al.*'s purification methodology was employed with slight modifications (Hong *et al.*, 2008). Avicel Ph 105 was used for the synthesis of RAC. Purification was performed at room temperature. Recombinant cellulases were eluted with ethylene glycol and ethylene glycol was further removed and exchanged with 50 mM sodium acetate buffer (pH 5) due to its interference with BCA protein assays by ultrafiltration through 10 kDa cutoff Vivaspin 500 membrane spin filters (Sartorius-Stedim).

### Protein and enzyme assays

Protein concentrations were determined either using BCA Protein Assay Reagent (Pierce) according to the manufacturer's instructions or using the NanoDrop. Bovine serum albumin was used as the protein standard for the BCA assays.

All CMC activity assays were performed in triplicate with a standard deviation of <10%. All 4-MUC activity assays were performed in duplicate or triplicate with a standard deviation <10%.

The activity of each enzyme at different temperatures (25–65°C) was determined by the 3,5-dinitrosalicylic acid (DNS) method against 0.5% CMC (w/v) in 50 mM sodium acetate buffer (pH 5). Each enzyme and substrate were preincubated for 5 min separately at each assay temperature, and then combined and incubated at the assay temperature for 10 min. Reducing sugars produced were measured at 550 nm. Glucose was used as a standard and enzyme activity was calculated as CMC units/ml according to IUPAC method for measurement of cellulase activity (Ghose, 1987). Percent enzyme activity was calculated by taking the maximum activity of the enzyme at the determined temperature as 100%.

Activity of each enzyme at different pHs (pH 3–7) was determined using the DNS method against 0.5% CMC (w/v) in different pH buffers for 10 min at 55°C. Citrate-phosphate buffer (100 mM citric acid, 200 mM $Na_2HPO_4$) at pH 3; 50 mM sodium acetate buffer at pH 4, 50 mM sodium acetate buffer at pH 5 and 50 mM potassium phosphate buffer at pH 6, pH 7 were used. Reducing sugars produced were measured at 550 nm. Glucose was used as a standard. Percent enzyme activity was calculated by taking the maximum activity of the enzyme at the determined pH as 100%.

The activities of the recombinant enzymes against 1% (w/v) phosphoric acid swollen cellulose were determined by the DNS method in 50 mM sodium acetate buffer (pH 5) and at 45°C. Each enzyme and substrate were incubated for 1 h. Reducing sugars produced were measured at 550 nm. Glucose was used as a standard and enzyme activity was calculated as CMC units/ml according to IUPAC method for measurement of cellulase activity (Ghose, 1987).

4-MUC assay was performed for enzyme supernatants according to the assay outlined by Chernoglazov *et al.* (1989). Enzyme samples were incubated with 0.5 mg/ml 4-MUC in 50 mM sodium acetate buffer at pH 5. Kinetic analysis was performed at 45°C for 30 min. 4-Methylumbelliferone (4-MU) was used as the standard. Liberated 4-MU was measured with a fluorescence spectrophotometer with excitation at 363 nm and emission at 435 nm. Enzyme activities were calculated as RFU liberated per minute, RFU liberated per second and mM 4-MU liberated per minute.

Kinetic constants ($k_{cat}$ and $K_m$) for the recombinant enzymes were determined against 4-MUC at 45°C in 50 mM sodium acetate buffer at pH 5. Ten different substrate concentrations (0.001–1 mM) were used for the kinetic assay. The enzyme concentrations were kept at 200 nM throughout the assay and all measurements were performed in triplicate. The kinetic constants, $K_m$ and $k_{cat}$ were calculated by fitting the initial rate data to the Michaelis–Menten equation with OriginPro software.

### Measurement of thermal stability

Percent remaining enzyme activity was determined by incubating 15 µM of each enzyme at 45 and 65°C for 0–8 h, chilling on ice for 10 min and assaying the enzyme activities against 4-MUC at 45°C and pH 5 in 50 mM sodium acetate buffer for 60 min. Enzyme activity with no incubation (time 0) was taken as 100%. At each incubation time, recombinant enzymes were assayed in duplicate with a standard deviation <10%.

Thermal shift assay was used to determine the thermal stability of the recombinant enzymes. Proteostat thermal shift stability assay kit (Enzo Life Sciences) was used to determine aggregation temperature ($T_{agg}$) of the recombinant enzymes. All the experiments were performed in quadruplicates following the instructions supplied by the manufacturer. One micromolar of each enzyme was used in the assays in a BioRad i-Cycler real-time PCR machine by heating the samples 1°C/min from 20 to 90°C at pH 5. Fluorescence readings at each temperature were normalized and plotted using OriginPro Software and first derivative of the graphs were calculated to determine the aggregation temperature.

## Results and discussion

### Computational selection of residues for site-directed mutagenesis

Protein structure stabilizing variations were first selected based on the consensus of at least three of four well performing different predictors (see above), and then tested with a classical molecular dynamics approach. To benchmark the high-temperature unfolding, MD simulations were carried out at 310 K for the native EGI protein and for the corresponding mutated systems Q126F, K272F and Q274V. Root Mean Squared Deviation (RMSD) (Cα atoms) for EGI and the three mutants indicate that at 310 K the protein variants are stable (Fig. 1). EGI RMSD diagram peaked at around 2.4 Å and increased slowly over the 10 ns of simulation. In general, mutants are more stable than the EGI native protein having by the end of the simulation lower RMSD values K272F peaked at 2.25 Å after 6 ns and Q274V at 2.2 after 9 ns of simulation. In conclusion, the more stable mutant is Q126F which peaked at 1.66 Å.

#### Protein flexibility

To gain insight on flexibility, we compare the RMS fluctuations for each residue (Fig. 2). Figure 4 shows that the flexibility of the systems is quite similar. Averages of the Cα fluctuations at 310 K for the wild type of the protein and the mutant systems are small (0.97, 0.87, 0.74 and 0.84) and follow this trend EGI > K272F > Q274V > Q126F.

The slight differences in the flexibility of the wild type protein are in the region around the 98–99 residues, together with the K272F mutant (the turn SSGGY region). Small differences are present in other coil structures (region of the residues 79, 231, 159 and 298).

#### Unfolding MD simulations

The unfolding of the native EGI protein and of the corresponding variants was addressed in 5 ns MD simulations at 550 K. The simulations

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

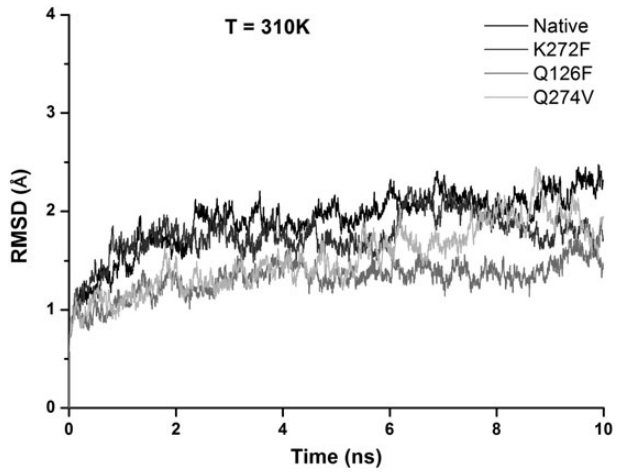



**Fig. 1** Root Mean Squared Deviations (RMSDs) of the Cα atoms for the EGI protein and the three mutated proteins over all trajectories in the 310 K simulations (10 ns).

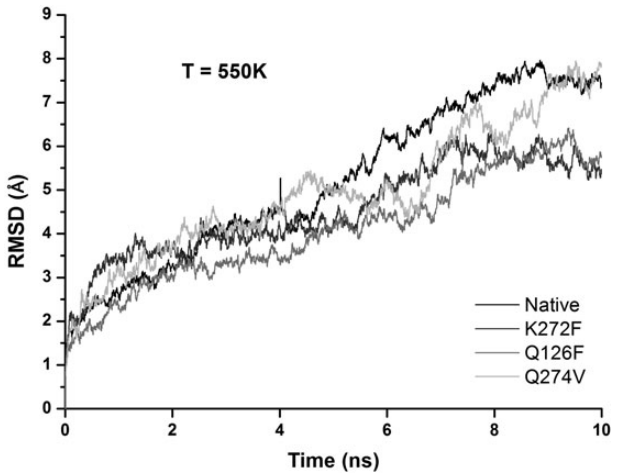



**Fig. 3** RSMDs of the Cα atoms for the EGI protein and the three mutated proteins over all trajectories in the 550 K simulations (10 ns).

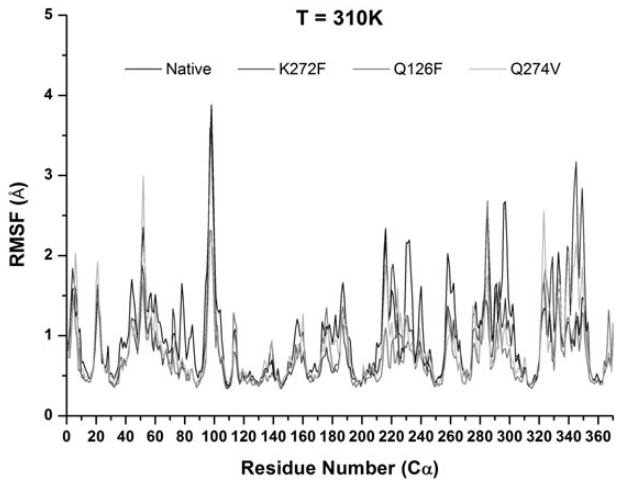



**Fig. 2** RSMDs per residue averaged over the last 5 ns of MD simulations at 310 K for EGI and the mutated proteins.

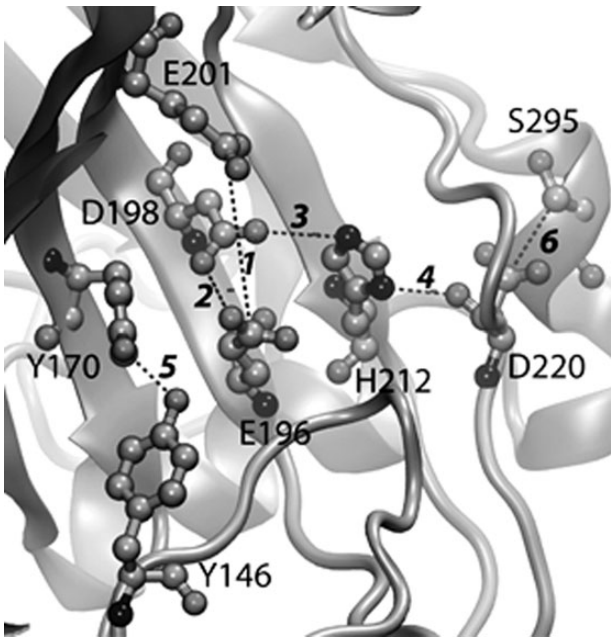



**Fig. 4** Three-dimensional representation of the active site of the EGI protein with the more important residues. In dashed line, the distances discussed in the text.

were started from structures obtained at the end of the trajectories of the 310 K simulations. In Fig. 3, the RMSD from the initial configuration of each mutant and the native enzyme at 550 K is given.

At this high temperature, Q126F and K272F are more stable than native EGI and Q274V mutant. EGI peaked at 7.27 Å by the end of the simulation together with Q274V. K272F and Q126F stabilize around 6 Å.

### Analysis of the distances in the active site

The distances between the catalytic cluster residues might be used to understand the impact of mutations on the structure of catalytic domain and the thermal unfolding. Eight residues in the catalytic site form a cluster which is one more than the catalytic site description published in Kleywegt *et al.* (1997) (Fig. 4): E197, E201, D198, E197, H212, Y146, Y170 and S295. In Table II given are the more important average distances of the catalytic cluster at 310 and 550 K. The distance between E196 and E201 (No. 1 in Fig. 3) is crucial for catalytic activity, and the average Cd–Cd distance is 8.6 Å for the wild type protein, 8.95 Å for K272F, EGI and Q126F meanwhile the distance is increased for K272F, 10.24 and Q274V. D198 is a secondary residue for catalytic action and forms an hyrogen bond with E196 (No. 2). EGI has an average Oe–Od distance of 4.44 Å. This distance increases slightly only for Q126F, 5.30 Å. H212 is also a secondary residue and forms another hydrogen bond with D198 (No. 3). The average distance is 6.97 Å for EGI. Together with D198, H212 roles in trafficking of the protons in the active site for proper hydrolysis. For K272F and Q274V, the value increases (8.05 and 9.23 Å) meanwhile it is lower for Q126F, 5.46 Å. H212 forms another hydrogen bond with the main chain carbonyl of D220 (No. 4). The two hydrogen bonds in the imidazole ring of H212 enables it to protonate the active site residues for catalytic activity. For EGI, the value is 5.44 Å which is

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

**Table II.** Summary of the mean distances (in Å) between the active site residues from the simulations

| Residues distances | EGI | | K272F | | Q126F | | Q274V | |
|---|---|---|---|---|---|---|---|---|
| | 310 K | 550 K | 310 K | 550 K | 310 K | 550 K | 310 K | 550 K |
| Glu196-Glu201(Cd-Cd) | 8.58 | 7.28 | 8.95 | 10.24 | 9.14 | 8.24 | 8.31 | 8.80 |
| Glu196-Asp198(Oe-Od) | 4.44 | 5.10 | 4.91 | 5.22 | 5.30 | 5.29 | 4.70 | 5.38 |
| Asp198-His212(Od-Ne2) | 6.97 | 5.87 | 8.05 | 7.71 | 5.46 | 6.06 | 9.23 | 8.18 |
| His212-Asp220(Nd1-O) | 5.44 | 5.21 | 3.96 | 7.03 | 2.93 | 8.19 | 5.64 | 9.32 |
| Tyr146-Tyr170(Oh-Oh) | 2.87 | 3.38 | 2.88 | 3.99 | 2.89 | 3.63 | 2.89 | 4.19 |
| Asp220-S295 (Cg-Cb) | 9.17 | 14.28 | 5.60 | 17.34 | 6.94 | 12.90 | 10.22 | 13.50 |

more or less the same for the mutated protein Q274V. The same values for K272F and Q126F are 3.96 and 2.93 Å, respectively. Two tyrosines at the bottom of the catalytic cluster, Y146 and Y170 also interact with each other with a hydrogen bond at a distance of 2.8 Å, which determines the border of the cleft for substrate binding (No. 5). This average value does not change in all the systems. S295 is part of the α-helix in the region between the 294 an 299 residues. Through the hydrogen bonding with D220, this helix is directly connected with the catalytic site (No. 6). Table II shows that this distance changes drastically at high temperatures. The mean value changes in this order Q274V < EGI < Q126F < K272F). For example, EGI has an average distance of 9.17 Å at 310 K and 14.28 Å at 550 K with an increase of 5.11 Å meanwhile K272 has an average value of 5.60 Å at 310 K and 17.34 Å at 550 K with a large increase of the mean value, 11.74 Å (see Table II).

**Comparison between 310 and 550 K**

The biggest change in the case of EGI is only for the D220–S295 (Cg–Cb) distance, 5.11 Å. Stride analysis shows that this is connected with the unfolding of the helix in the region of the residue 295. For K272F, big changes are for the distances between H212–D220 (Nd1–O) (3.7 Å) and D220–S295 (Cg–Cb). In this system, the α-helix in the region 294–299 does not even exist at 310 K. The same happens to Q126F and Q274 with the values becoming 5.26, 5.96, 3.68 and 3.28, respectively. Meanwhile the distances for EGI, at high temperatures, have large fluctuations without significant changes except for the distance between D220–S295 (Cg–Cb) residues. In the other mutants, these changes even concern the H212–D220 (Nd1–O) distance responsible for the structural intregrity of the active site.

**Comparison between EGI and the other variants**

At 310 K, K272F differs only in the D220–S295 (Cg–Cb) distance. Q126F also in the H212–D220 (Nd1–O) where both the distances become shorter. For Q274V at 310 K, however, the D198–H212 (Od–Ne2) distance also changes. Both D198 and H212 role in trafficking of the protons in the active site for proper hydrolysis. At 550 K, K272F differs also in the Q196–Q201 (Cd–Cd) distance meanwhile Q126F only in the H212–D220 (Nd1–O). Q274V has a big change in H212–D220 (Nd1–O) distance and a smaller change in the D198–H212 (Od–Ne2) distance.

This variation of the hydrogen bonding between D220 and S295 seems particularly important with regard to the unfolding at high temperature. In fact, if we compare the protein secondary structure assignment, using STRIDE algorithm, the unfolding of the protein at high temperature begins from the α-helix in the 294–299 residues region. Simulations at 550 K show that for EGI, unfolding starts in the 1–3 ns interval, for K272F and Q126 after 6 ns while for Q274V this helix is unstable before 1 ns of simulation.

In summary, at higher temperatures these average distances change only slightly for the wild-type protein EGI suggesting a quite stable active site cluster structure. This is not the same in the case of the mutated systems. For K272F, the distance between H212–D220 changes substantially. The same behavior can be seen in the other mutated systems where, consequently, the structural integrity of the active site is probably compromised (Kleywegt *et al.*, 1997). Moreover, big fluctuations are present in the distance between the α-helix (S295) and the D220 residue. Analysis of the secondary structure assignment shows that the beginning of unfolding of protein is due to this part of the molecule. In general, we believe that this part of the protein is too close to the active site to be irrelevant to the decrease in activity due to mutations and to the increase of the temperature.

## Experimental validation

Site-directed mutagenesis was performed on codon optimized (for *P.pastoris*) synthetic *endoglucanase 1* (*egl1*) gene of *T.reesei* using Overlap Extension primers enlisted in Table I to introduce *in silico* predicted Q126F, K272F and Q274V point mutations. Protein sequence of the synthetic gene is identical to *T.reesei* endoglucanase 1 (GenBank Accession Number: M15665). Overlap PCR extension method was used to introduce the point mutations. The results of the Overlap PCR extension were shown in Supplementary Figs S1 and S2. *egl1* and the mutant genes were subcloned in *E.coli* cells. Colony PCR positive colonies were selected (Supplementary Fig. S3). The fidelity of the constructs was confirmed by sequencing of the mutant plasmids as shown in Supplementary Fig. S4. The selected plasmids were linearized and transformed into *P.pastoris* KM71H cells.

The colonies selected on Azo-CMC agar plates (for Q126F mutant colony 9, for K272F mutant colony 15 and for Q274V mutant colony 31) as shown in Fig. 5 were used for expression studies in BMM medium at 29°C. The activities of the produced recombinant enzymes were determined at 45°C and pH 5 against 4-MUC substrate (Fig. 6). Q126F activity was found to be decreased with increasing fermentation time. The produced enzymes were found to be glycosylated by *P.pastoris* as no single bands could be observed in the SDS-PAGE gels (Supplementary Fig. S5).

pH activity assays were performed at 55°C from pH 3 to pH 7. Activity results are shown in Fig. 7. The native enzyme and the mutants were found to exhibit a pH optimum around pH 5. Q126F mutant has retained ~25% of its activity at pH 4 whereas it has lost its activity completely at pH 6. K272F mutant has exhibited no activity at pH 3 and 4 but retained ~35% of its activity at pH 6 and 7. Q274V mutant has kept ~70% of its activity intact at pH 4 and continued to exhibit ~20 and ~10% of its activity at pH 6 and pH 7, respectively.

Activity of the native and mutant enzymes at different temperatures (from 25 to 65°C) was determined against 1% CMC substrate at pH 5 (Fig. 8). Q126F mutant was found to exhibit a native like temperature activity profile and Q274V mutant was found to have an

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

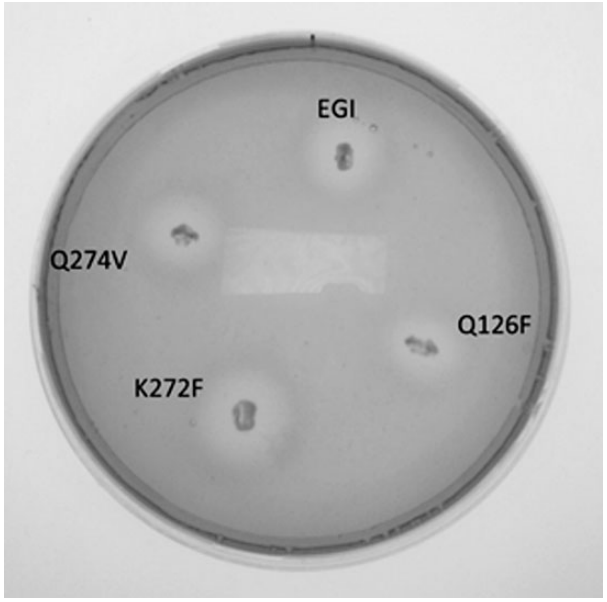



**Fig. 5** Enzyme production on Azo-CMC containing BMM-agar medium for Q126F, K272F andQ274V. Native EGI producing colony was also grown for comparison.

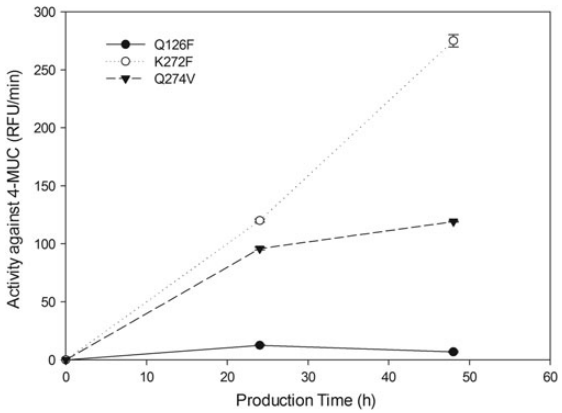



**Fig. 6** Activity of the produced mutant enzymes from the culture supernatants against 4-MUC at 45°C.

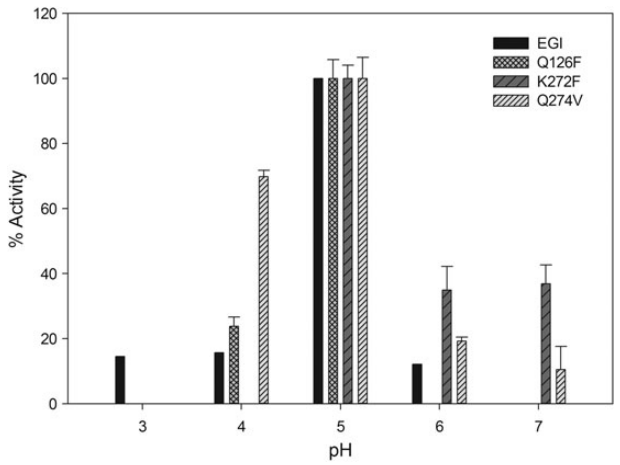



**Fig. 7** pH activity profile of EGI and EGI mutants against 0.5% CMC at 55°C. Maximum activities were taken as 100% and other activities were calculated accordingly. Standard deviations were <10% and all the experiments were performed in triplicates.

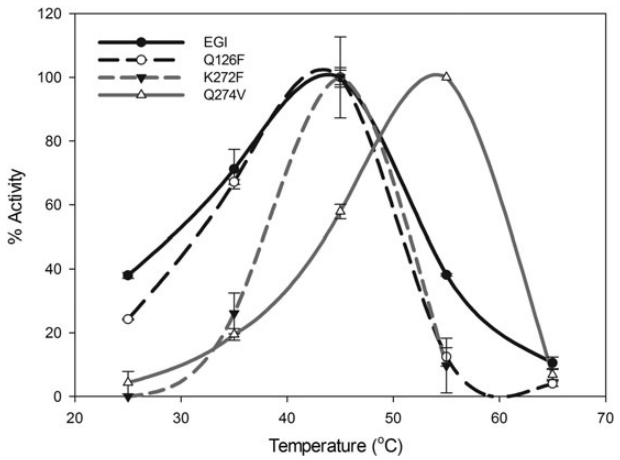



**Fig. 8** Temperature activity profile of EGI and EGI mutants against 1% CMC at pH 5. Maximum activities were taken as 100% and other activities were calculated accordingly. Standard deviations were <10% and all the experiments were performed in triplicates.

increased temperature optimum with respect to other mutants and EGI. The mutants were found to exhibit lower specific activity toward CMC compared with the native enzyme (Table III).

Kinetic parameters for EGI and the mutants were determined by assaying the enzyme activities against 4-MUC substrate at pH 5 and 45°C for an hour (Table IV). Ten different substrate concentrations (from 0.01 to 1 mM) were used. The turnover numbers ($k_{cat}$), the Michealis–Menten constant ($K_m$) and the specificity constants ($k_{cat}/K_m$) for the mutants were found to be decreased with respect to EGI. The mutant enzymes were found to exhibit increased affinity toward 4-MUC in comparison with EGI. A drastic decrease was observed in the turnover numbers of the mutants and hence, the $k_{cat}/K_m$ values indicate that the catalytic efficiency of the mutant enzymes was decreased compared with EGI. The decrease in the specific activity of the mutants is also reflected in the decrease in the catalytic efficiency of the mutant enzymes in comparison with the native enzyme. The mechanistic basis of the decrease in the substrate specificities and turnover rates remains unclear. Although the spatial arrangement of the mutations is not in close proximity to the active site cleft, these residues located on the surface of the enzyme, could play a role in binding and orienting the substrate to the catalytic site along with the CBD. It has been long recognized that mutations in residues far away from the active site might change the substrate specificity of the enzymes (Fong *et al.*, 2000; Nardini *et al.*, 2000). Fong *et al.* changed the substrate specificity of the *E.coli* D-2-keto-3-deoxy-6-phosphogluconate aldolase using directed evolution and all of the substitutions they have found were located far away from the active site. For the temperature stability tests, EGI and the mutants were incubated at 45 and 65°C for 30 min, 1 h and 8 h in duplicates and then after incubation, the enzymes were immediately put on ice and then all the incubated and unincubated enzymes were assayed for their activity against 4-MUC at 45°C for 1 h (Fig. 9). All three mutants were found to be more stable with respect to EGI upon prolonged incubations at 45 and 65°C, Q274V mutant being the most stable and Q126F mutant being the least.

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

**Table III.** Specific activities of EGI vs.mutants against 1% CMC at pH 5 and 45°C

| | EGI | Q126F | K272F | Q274V[a] |
|---|---|---|---|---|
| Specific enzyme activity (mM reducing sugars produced/μM protein) | 0.034563 | 0.013186 | 0.004059 | 0.011513 |

[a]At 55°C.

**Table IV.** Kinetic parameters of EGI and mutants for hydrolysis of 4-MUC at pH 5 and 45°C

| | $k_{cat}$ ($s^{-1}$) | $K_m$ (mM) | $k_{cat}/K_m$ ($s^{-1}$ mM) |
|---|---|---|---|
| EGI | 0.010 | 0.125 | 0.08 |
| Q126F | 0.003 | 0.114 | 0.02 |
| K272F | 0.001 | 0.111 | 0.01 |
| Q274V | 0.001 | 0.100 | 0.01 |

A thermal stability shift assay was used to compare the temperature stability of mutant enzymes with respect to EGI based on aggregation temperature ($T_{agg}$). $T_{agg}$ is known to be an indicator of protein stability. The fluorescent probe used in the assay was known to bind to the aggregated version of the proteins arising from thermally induced protein denaturation. EGI and the mutants have exhibited multi-phasic aggregation. Three different aggregation curves were obtained upon thermal denaturation of the proteins. Aggregation temperatures are shown in Table V. The second and the third aggregation temperature of the mutants were found to be increased compared with EGI and the Q274V mutant was the most stable. Thermal stability shift assay results confirmed the thermal stability test results. The increase in thermal stability of the mutant enzymes with a cost of reduction in the enzyme activity and the spatial arrangement of the mutated residues with the substrate-binding domain indicate that the 'stability–function' theory may also be valid for EGI. The 'stability–function' theory states that there is a balance between protein function and stability. The residues involved in catalysis or ligand binding may not be optimal for protein stability. As a result, thermostabilizing mutations on these residues might lead to decreased activity. Shoichet *et al.* mutated catalytic and substrate-binding residues of T4 lysozyme and shown that the mutations on substrate-binding residues lead to increased thermostability of the enzyme with a cost of reduction in the activity (Shoichet *et al.*, 1995).

In conclusion, we have found three mutations that confer thermostability to EGI using predictive algorithms. We have performed molecular dynamics simulations to corroborate the thermostabilizing effect of these selected mutations on protein structure and understand the underlying mechanisms of thermostability. Both high temperature and room temperature MD simulations supported the fact that Q126F, K272F and Q274V mutations have a stabilizing effect on the protein structure. The computational studies on the stabilizing mutations were validated by experimental studies. All of the mutants were shown to exhibit higher thermostability compared with native EGI. However, the specific activity of the mutants was found to be lower than the native EGI. This reduction in the specific activity of the mutant enzymes can be compensated by increasing the amount of the enzyme for high-temperature processes. We believe that the increase in the thermostability of the mutants was due to the changes in the distances of the catalytic cluster residues as a result of the mutations. This also accounts for the reduced specific activity in the mutants compared with EGI. Active site integrity is changed in the three mutants as shown by the drastic changes in the distances between H212–D220 and D220–S295 at high temperatures simulations.

Further experiments would be performed to understand and alleviate the decrease in the specific activity.

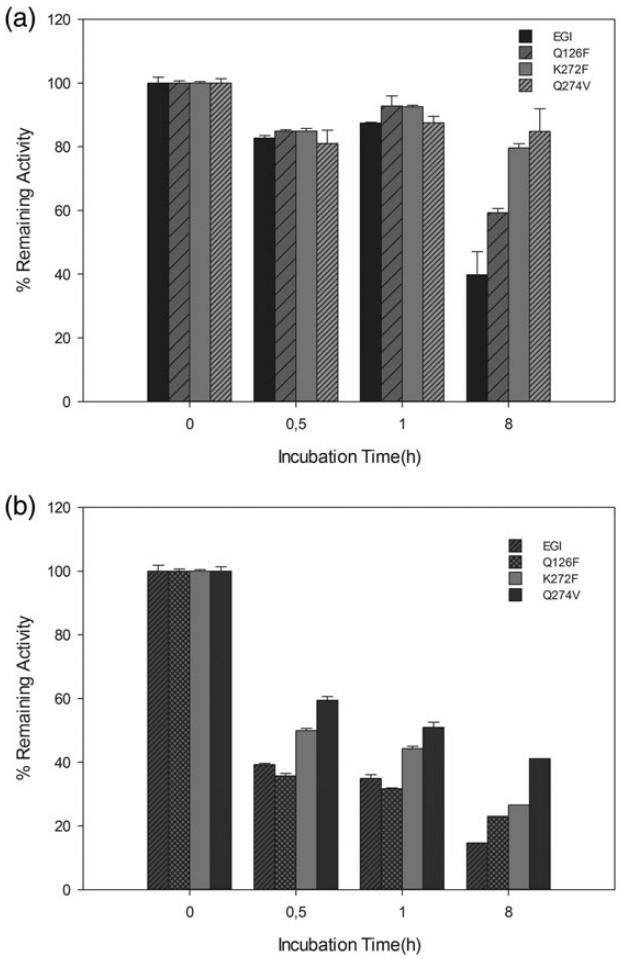



**Fig. 9** Thermal stability test results for EGI, Q126F, K272F and Q274V: (**a**) at 45°C and (**b**) 65°C for 0–8 h. 1% CMC at pH 5 is used as the substrate.

**Table V.** Prediction of $T_{agg}$ of EGI and mutants using thermal shift stability assay

| | $T_{agg}1$ (°C) | $T_{agg}2$ (°C) | $T_{agg}3$ (°C) |
|---|---|---|---|
| EGI | 48.5 | 64.5 | 82 |
| Q126F | 49 | 69 | 82.5 |
| K272F | 48.5 | 65.5 | 82 |
| Q274V | 48 | 69.5 | 84.5 |

1 μM of each protein is used in the experiments.

## Supplementary data

Supplementary data are available at *PEDS* online.

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022

## Acknowledgements

We would like to thank Dr Emel Durmaz Timucin for her invaluable comments. A.V. thanks the 'Progetto bandiera' N-CHEM and Mr Giovanni Bragaglia for his valuable assistance in the Computer Laboratory.

## Funding

This research was co-funded by Turkish Scientific and Technologic Research Agency (TUBITAK) (Grant no. 111T846) and National Research Council of Italy (CNR), in the framework of the Bilateral Agreements of Scientific and Technological Cooperation between Turkey and Italy (2011–2013).

## References

Arnold,F.H. (2009) *Cold Spring Harb. Symp. Quant. Biol.*, **74**, 41–46.
Arnold,F.H. and Moore,J.C. (1997) *Adv. Biochem. Eng. Biotechnol.*, **58**, 1–14.
Bayram Akcapinar,G., Gul,O. and Sezerman,U.O. (2012) *J. Biotechnol.*, **159**, 61–68.
Bhat,M.K. (2000) *Biotechnol. Adv.*, **18**, 355–383.
Capriotti,E., Fariselli,P., Calabrese,R. and Casadio,R. (2005a) *Bioinformatics*, **21**(Suppl 2), ii54–ii58.
Capriotti,E., Fariselli,P. and Casadio,R. (2005b) *Nucleic Acids Res.*, **33**, W306–W310.
Cereghino,J.L. and Cregg,J.M. (2000) *FEMS Microbiol Rev*, **24**, 45–66.
Chernoglazov,V.M., Jafarova,A.N. and Klyosov,A.A. (1989) *Anal. Biochem.*, **179**, 186–189.
Cregg,J.M., Barringer,K.J., Hessler,A.Y. and Madden,K.R. (1985) *Mol Cell Biol*, **5**, 3376–85.
Darden,T., York,D. and Pedersen,L. (1993) *J. Chem. Phys.*, **98**, 10089–10092.
Day,R. and Daggett,V. (2005a) *Proc. Natl Acad. Sci. U.S.A.*, **102**, 13445–13450.
Day,R. and Daggett,V. (2005b) *Protein Sci.*, **14**, 1242–1252.
Dehouck,Y., Grosfils,A., Folch,B., Gilis,D., Bogaerts,P. and Rooman,M. (2009) *Bioinformatics*, **25**, 2537–2543.
Demain,A.L., Newcomb,M. and Wu,J.H.D. (2005) *Microbiol. Mol. Biol. Rev.*, **69**, 124–154.
Eleftheriou,M., Germain,R.S., Royyuru,A.K. and Zhou,R. (2006) *J. Am. Chem. Soc.*, **128**, 13388–13395.
Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) *J. Chem. Phys.*, **103**, 8577–8593.
Fong,S., Machajewski,T.D., Mak,C.C. and Wong,C. (2000) *Chem. Biol.*, **7**, 873–883.
Gao,D., Narasimhan,D.L., Macdonald,J., Brim,R., Ko,M.C., Landry,D.W., Woods,J.H., Sunahara,R.K. and Zhan,C.G. (2009) *Mol. Pharmacol.*, **75**, 318–323.
Gao,D., Uppugundla,N., Chundawat,S., *et al.* (2011) *Biotechnol. Biofuels*, **4**, 5.
Ghose,T.K. (1987) *Pure Appl. Chem.*, **59**, 257–268.
Heinzelman,P., Snow,C.D., Smith,M.A., *et al.* (2009a) *J. Biol. Chem.*, **284**, 26229–26233.
Heinzelman,P., Snow,C.D., Wu,I., Nguyen,C., Villalobos,A., Govindarajan,S., Minshull,J. and Arnold,F.H. (2009b) *Proc. Natl Acad. Sci. U.S.A.*, **106**, 5610–5615.
Hellinga,H.W. (1997) *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 10015–10017.
Henrissat,B. (1991) *Biochem. J.*, **280**, 309–316.
Hong,J., Wang,Y., Ye,X. and Zhang,Y.H. (2008) *J. Chromatogr. A*, **1194**, 150–154.
Jäckel,C., Kast,P. and Hilvert,D. (2008) *Annu. Rev. Biophys.*, **37**, 153–173.
Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) *J. Chem. Phys.*, **79**, 926.
Kleywegt,G.J., Zou,J., Divne,C., *et al.* (1997) *J. Mol. Biol.*, **272**, 383–397.
Kumar,S., Tsai,C.J. and Nussinov,R. (2000) *Protein Eng.*, **13**, 179–191.
Lantz,S., Goedegebuur,F., Hommes,R., Kaper,T., Kelemen,B., Mitchinson, C., Wallace,L., Stahlberg,J. and Larenas,E. (2010) *Biotechnol. Biofuels*, **3**, 20.
Li,W.F., Zhou,X.X. and Lu,P. (2005) *Biotechnol. Adv.*, **23**, 271–281.
MacKerell,A.D., Bashford,D., Bellot,M., *et al.* (1998) *J. Phys. Chem. B*, **102**, 3586–3616.
Nardini,M., Lang,D.A., Liebeton,K., Jaeger,K.E. and Dijkstra,B.W. (2000) *J. Biol. Chem.*, **275**, 31219–31225.
Phillips,J.C., Braun,R., Wang,W., *et al.* (2005) *J. Comput. Chem.*, **26**, 1781–1802.
Rastogi,G., Bhalla,A., Adhikari,A., Bischoff,K.M., Hughes,S.R., Christopher, L.P. and Sani,R.K. (2010) *Bioresour. Technol.*, **101**, 8798–8806.
Shoichet,B.K., Baase,W.A., Kuroki,R. and Matthews,B.W. (1995) *Proc. Natl Acad. Sci. U.S.A.*, **92**, 452–456.
Sterner,R. and Liebl,W. (2001) *Crit. Rev. Biochem. Mol. Biol.*, **36**, 39–106.
Vallejo,A.N., Pogulis,L.J. and Pease,L.R. (1994) *Genome Res.*, **4**, 123–130.
Vetriani,C., Maeder,D.L., Tolliday,N., Yip,K.S.P., Stillman,T.J., Britton,K.L., Rice,D.W., Klump,H.H. and Robb,F.T. (1998) Proc. Natl Acad. Sci.*U.S.A.*, **95**, 12300–12305.
Wu,S. and Letchworth,G.J. (2004) *Biotechniques*, **36**, 152–154.
Zhao,H. and Arnold,F.H. (1997) *Curr. Opin. Struct. Biol.*, **7**, 480–485.
Zhou,C.H., Xia,X., Lin,C.X., Tong,D.S. and Beltramini,J. (2011) *Chem. Soc. Rev.*, **40**, 5588–5617.

Downloaded from https://academic.oup.com/peds/article/28/5/127/1525464 by guest on 15 March 2022