

Mutations from a family-shuffling-library reveal amino acid residues responsible for the thermostability of endoglucanase CelA from *Clostridium thermocellum*

Zhuo-Lin Yi · Zhong-Liu Wu

Received: 4 July 2010 / Accepted: 21 July 2010 / Published online: 31 July 2010
© Springer Science+Business Media B.V. 2010

Abstract We constructed a library of chimeras from the major endoglucanase, CelA, of *Clostridium thermocellum* and a less stable endoglucanase CelB from *Clostridium josui* with multiple point mutations using low-fidelity family-shuffling method. Mutations that inactivated the enzyme were rapidly eliminated with high-throughput screening. The activities and thermostabilities of selected variants were evaluated, and four amino acid substitutions, K249R, P258S, S329N and E355G, were identified as having significant impact on the thermostability of CelA without affecting enzymatic activity. In the crystal structure of CelA, most of them are away from

the activity cleft and are responsible for the stabilization of secondary structures.

Keywords CelA · *Clostridium thermocellum* · Endoglucanase · Family shuffling · Thermostability

Introduction

Cellulosic biomass is the most abundant renewable resource in the world. In the pursuit of cellulose-based biofuel production, *Clostridium thermocellum* is the most extensively studied thermophilic anaerobic bacteria. It exhibits one of the fastest growth rates on crystalline cellulose (Lynd et al. 2002) by virtue of an efficient extracellular enzyme-complex, the large cellulosome, which consists of a large noncatalytic scaffolding protein and numerous catalytic modules exhibiting varied activities of endoglucanase, exoglucanase, hemicellulase and chitinase etc (Lynd et al. 2002).

The endoglucanase CelA was one of the first cellulases to be purified (Petre et al. 1981). It is the major endoglucanase in the cellulosome as demonstrated by both early studies (Petre et al. 1981) and recent proteomic analysis (Gold and Martin 2007; Zverlov et al. 2005). It is optimally active at 75°C and is resistant to the denaturing effects of detergents and organic solvents (Schwarz et al. 1986). The catalytic domain of CelA folds into an (α/α)₆ barrel. Structural analysis has identified a number of intramolecular

Purpose of work CelA is one of the major endoglucanases in the cellulosome of thermophilic bacteria *Clostridium thermocellum* for cellulosic biomass degradation. The stabilizing effect of critical residues of CelA remain have now been studied.

Electronic supplementary material The online version of this article (doi:10.1007/s10529-010-0363-0) contains supplementary material, which is available to authorized users.

Z.-L. Yi · Z.-L. Wu (✉)
Chengdu Institute of Biology, Chinese Academy
of Sciences, P.O. Box 416, Chengdu 610041, China
e-mail: wuzhl@cib.ac.cn

Z.-L. Yi
Graduate University of the Chinese Academy
of Sciences, Beijing 100049, China

salt bridges in the catalytic core of CelA, and one occurs within the central hydrophobic region is presumed to be important for the stability of the active-site architecture (Alzari et al. 1996). So far, studies on the stabilizing effect of individual critical residues of CelA remain absent, which would otherwise provide fundamental insights into the structural basis of this remarkably stable protein.

Here, we report for the first time, the identification of critical amino acid residues responsible for the thermostability of endoglucanase CelA. Due to the well-known complicated nature of proteins, rational prediction of stability-related residues remains difficult, even though quite a few factors, intramolecular interactions in particular, such as hydrogen bondings, ion pairs, hydrophobic interactions have been identified as of general significance in the stability of proteins (Chakravarty and Varadarajan 2000, 2002; Melchionna et al. 2006; Ragone 2001). In a study on cellulase CelC from *C. thermocellum*, carefully designed mutations with additional electrostatic interactions for enhanced thermostability returned unsuccessful results, and lost of activity was also observed (Nemeth et al. 2002). Therefore, in this work, we applied a random library approach instead, which was constructed using family shuffling method with two parental enzymes, CelA and CelB from *Clostridium josui* with a temperature optimum at 60°C (Fujino and Ohmiya 1991). By analyzing the sequences of randomly picked mutants with varied thermostability, four amino acid substitutions were identified as having significant impact on the thermostability of CelA, while enzymatic activities remained unchanged. Their effects were discussed in the context of the X-ray crystal structure.

Materials and methods

Materials

Oligonucleotides were purchased from Invitrogen Life Technologies (Shanghai, China) in PAGE-purified grade. Restriction enzymes and T4 DNA ligase were from New England Biolabs (Beverly, MA). Carboxymethylcellulose sodium salt (CMC) was from Sigma-Aldrich. All other reagents were obtained from general commercial suppliers and used without further purification.

Construction of family shuffling library

The DNA sequences encoding endoglucanases CelA of *Clostridium thermocellum* (1338 bp) and CelB of *Clostridium josui* (1287 bp) (NCBI accession No. AAA83521 and P37701) were obtained via PCR-based gene synthesis with codons automatically adjusted to fit the bias of *Escherichia coli* using DNAWorks program (<http://helixweb.nih.gov/dnawork>) (Hoover and Lubkowski 2002) (see Supplementary Data for sequence informations). Both sequences were constructed into pET-28a(+) vector (Novagen, Madison, WI) and amplified using *Taq* polymerase with homologous primers 5'-GTGAGCGGATAACAATTCCC-3' and 5'-CCTCAAGACCCGTTTATAGAGG-3'. After treatment with DNase I (Stemmer 1994), fragments in the range of 50–150 bp were excised, purified using the QIAEX II Gel Extraction kit (Qiagen, Germany). The reassembly protocol was carried out as previously described (Abecassis et al. 2000) using the *Taq* polymerase to achieve low-fidelity assembly based on its relatively low replication fidelity lacking a 3'- to 5'-exonuclease proof-reading activity (Zhao and Arnold 1997). Full-length reassembly products were amplified with sense primer 5'-CAGCAAATGGGTCGCGGATCCG-3' and antisense primer 5'-GAGTGC GGCCGCAAGCTTCTAA-3'. The resulting hybrid genes were purified using the TIANgel Mini Purification kit (TianGeng, China), digested with *Bam*H I and *Hind* III and inserted into the corresponding site of pET-28a(+). *E. coli* DH10B electro-competent cells (Invitrogen, Carlsbad, CA) were transformed with the ligation mixture using an Electroporator 2510 (Bio-Rad Laboratories, Veenendaal, The Netherlands) and plated on to Luria-Bertani (LB) agar containing 50 µg kanamycin/ml.

High-throughput enzymatic assays for variants from family shuffling library

The library plasmids in *E. coli* DH10B were purified using the TIANprep Mini Plasmid Kit (TianGeng, China), transformed into *E. coli* BL21(DE3) (Novagen, Madison, WI) competent cells and selected on LB plates fortified with 50 µg kanamycin/ml. After 13 h at 37°C, bacterial colonies were overlaid with top-agarose medium, which was composed of 0.7% agarose, 0.3% CMC and 1 mM IPTG, then incubated at 37°C for another 2 h to allow the hydrolysis of

CMC. The plates were then flooded with 0.1% (w/v) Congo Red (BIO BASIC INC.). After 20 min, the Congo red solution was poured off and the plates were washed twice with 1 M NaCl for 10 min.

Colonies with endoglucanase activity were identified by yellowish halos (Teather and Wood 1982), and transferred into sterile 96-well microplates containing LB medium fortified with 50 µg kanamycin/ml and 1 mM IPTG, and incubated at 37°C with shaking at 230 rpm for 30 h. The resulting culture in each well was centrifuged, cells were resuspended with 200 µl 100 mM succinate buffer (pH 5.8) and split into two duplicated microplates. One microplate was pre-incubated at 85°C for 2 h to cause a large loss of activity of wild type CelA (~40% residual activity). The enzymatic assays in both microplates were then carried out at 45°C for 3.5 h in the presence of 1% (w/v) CMC as substrate. The reactions were quenched by the addition of 3,5-dinitrosalicylic acid reagent (Aiba et al. 1983) and heated at 105°C for 30 min to determine the amount of reducing sugar.

Construction of point mutations

Plasmid pET-28a(+) containing the DNA fragment encoding CelA was used as the template for all site-directed mutagenesis reactions. PCRs were performed with mutagenesis primers (Supplementary Table S3) following the instruction of Quick Change Mutagenesis Kit (Stratagene, La Jolla, CA), and the products were treated with *DpnI* to digest the parental DNA templates. The nicked vectors with desired mutations were then transformed into *E. coli* DH5α cells. All constructs were validated by DNA sequencing.

Expression of wild-type endoglucanases and variants

The constructed plasmids were introduced into *E. coli* BL21 (DE3). Single colonies were cultivated in 20 ml LB medium fortified with 50 µg kanamycin/ml and 1 mM IPTG at 37°C for 30 h. Cells were collected by centrifugation, resuspended in 10 ml 100 mM succinate buffer (pH 5.8) and homogenized by sonication. The homogenates were treated at 65°C for 10 min to remove majority of the proteins from *E. coli*, and centrifuged at 13,000×*g* for 20 min. The

supernatants were stored at –20°C for enzymatic assays.

Measurement of enzyme activities

CMC was used as the substrate to assay endoglucanase activity according to established method (Schwarz et al. 1986). All measurements were performed in triplicates. Enzymatic reactions were performed at 60°C for 15 min in a 1% (w/v) CMC in 100 mM succinate buffer (pH 5.8), and reducing sugars were determined with DNS method (Aiba et al. 1983). One unit of endoglucanase was defined as the quantity of endoglucanase capable of releasing 1 µmol glucose equivalent per min.

To measure irreversible thermal inactivation, samples of crude enzymes were incubated at 86°C for different time and immediately placed in ice. Then the remaining activities were assayed as described above. The thermal inactivation half-life ($t_{1/2}$) of each enzyme was determined by plotting the natural logarithms of remaining activity values versus incubation time and deduced from linear regression.

The thermostabilities of the enzymes were analyzed by incubating samples of crude enzymes at various temperatures ranging from 62 to 94°C for 10 min. The samples were then placed on ice and the remaining activities were assayed under standard conditions.

Results and discussion

To generate functionally diverse variants of CelA, we attempted to employ family shuffling method with CelA and a less stable endoglucanase as parental enzymes. BLAST search revealed CelB from *Clostridium josui* as the best candidate with moderate degree of similarity to facilitate library construction. It shares 56% sequence identity with CelA, and both belong to family 8 of the inverting glycosidases. CelA and CelB have temperature optimum at 75 and 60°C, respectively (Fujino and Ohmiya 1991), which is expected to generate chimeric endoglucanases with significantly altered thermal stability.

A low-fidelity assembly strategy was employed to introduce more point mutations and increase the diversity of the library. The constructed family shuffling library contained around 20,000 colonies.

Sequence analysis of seven randomly picked variants revealed that the gene fragments were randomly recombined and point mutations were universally introduced (Supplementary Fig. S3). Around 10000 colonies were rapidly evaluated on agar plates using Congo red staining, and a total number of 90 active variants were then cultivated in 96-well plates in duplicates for quantitative assays. Both activities with or without heat challenge were tested for each mutant with CelA on the same plates as the control.

Around 50% colonies showed ~40% residual activity, similar to CelA, and 30% colonies showed <20% residual activity. From either of the group, four colonies with similar enzymatic activities as the wild-type were randomly selected and subjected to sequencing. Four of the mutants, G12, 10G12, 13D7 and 10H12, shared similar sequence patterns (Fig. 1), but differed sharply in thermostability, which made them suitable for comparative analysis. Detailed analysis on their thermal inactivation pattern revealed that the thermal stability of variants 13D7 and 10H12 were close to CelA, while variants G12 and 10G12 displayed significantly lower stability than CelA (Fig. 2).

For all four variants, the C-terminal dockerin domain of CelA was exchanged to that of CelB. This domain is known to anchor CelA to a scaffolding protein in cellulosome (Gold and Martin 2007) and is devoid of catalytic activity (Alzari et al. 1996). The exchange of this domain appeared have no significant effect on the thermostability of CelA.

Ten amino acid substitutions were observed to be present only in variants G12 and 10G12, but not in 10H12 or 13D7 (Fig. 1), which suggested that some of them might be responsible for the thermostability of CelA. The individual effect of these substitutions on the thermostability of CelA was investigated by constructing the corresponding point mutations. All

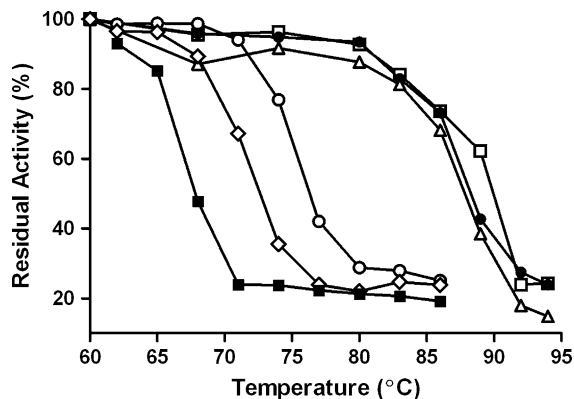


Fig. 2 Thermostability of endoglucanase CelA (closed circle), CelB (closed square) and variants at different temperatures. Shown are mutants 13D7 (open square), 10H12 (open triangle), 10G12 (open circle) and G12 (open diamond). Residual activities were measured after heat treatment at various temperatures for 10 min. Each point presented is the mean of triplicate assays

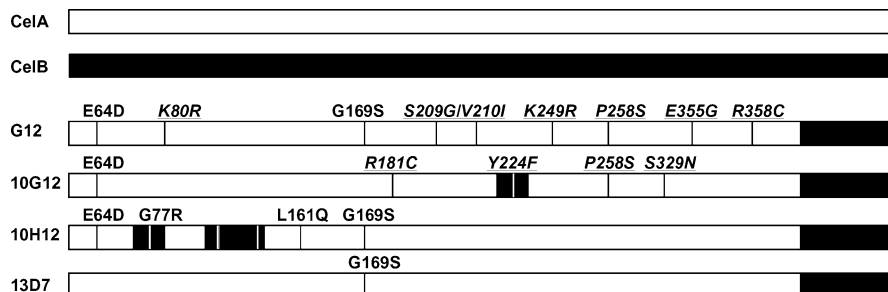
Table 1 Half-lives of thermal inactivation at 86°C for CelA and mutants

Enzymes	$t_{1/2}$ ^a
WT-CelA	20 min
G12, 10G12	<3 min
K80R, R181C, F224Y, R358C	Normal
S209G, V210I	Slightly diminished
K249R, P258S, S329N, E355G	<3 min

^a Mutant enzymes with $t_{1/2}$ s of 18–22 min and 12–14 min are grouped as “normal” and “slightly diminished”, respectively

mutants showed no significant change of activity (1.3–2.2 U/mg) compared with the wild-type CelA (1.6 U/mg), but their half-lives of thermal inactivation ($t_{1/2}$) at 86°C varied markedly (Table 1) with similar grouping patterns as Fig. 3, which showed the residual activities after heat treatment at 73°C for 1 h. The

Fig. 1 Diagrammatic sketch of variants with varied thermal resistance. Ten amino acid substitutions subjected to further analysis were in *italics* and *underlined*



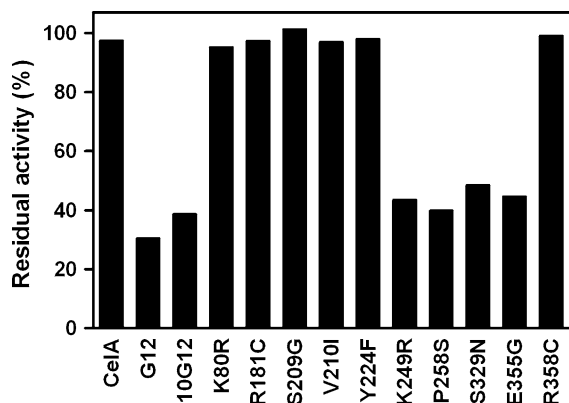


Fig. 3 Residual activities of CelA and variants. Heat treatment was performed at 73°C for 1 h. Each column presented is the mean of triplicate assays

Table 2 Half inactivation temperatures of wild-type CelA and mutants

	T_{50} (°C)	ΔT_{50} (°C) ^a
Wild-type CelA	88	–
G12	73	–15
10G12	77	–12
K249R	84	–4
P258S	84	–4
S329 N	85	–3
E355G	85	–3

^a ΔT_{50} (°C) is the change in T_{50} value for the mutant enzymes relative to the wild type

results indicated that the decreased thermostability of variants G12 and 10G12 were originated from the combination of a few negative mutations: K249R, P258S and E355G for variant G12, and P258S and S329N for variant 10G12. These four mutations also caused significant decrease in half inactivation temperatures (T_{50}) as defined by the temperature at which the enzyme loses 50% activity (Table 2). Their negative effect on the thermostability of CelA appeared additive or synergistic as judged by further decreased ΔT_{50} values of variants G12 and 10G12.

Based on the X-ray crystal structure of CelA in complex with cellopentaose (PDB code: 1kwf), catalytically critical amino acid residues have been identified previously (Guerin et al. 2002; Yao et al. 2007). However, those responsible for the thermostability of CelA remain difficult to predict (Alzari et al. 1996; Guerin et al. 2002). No previous analysis

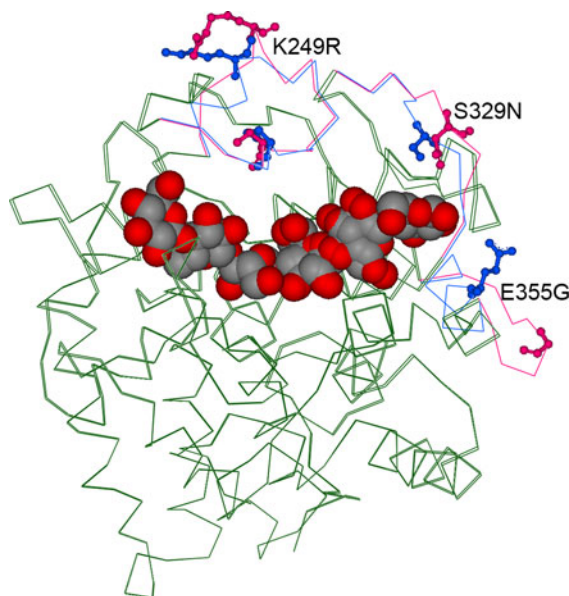


Fig. 4 Overlay of wild-type (PDB code: 1kwf) and mutant CelAs. Four mutations are labeled in black and showed in stick styles, and nine amino acid residues adjacent to positions 249, 258, 329, 355 are shown in blue for the wild type and pink for the mutant. Substrate cellopentaose is shown in spheres style, complementary to activity cleft. Amino acid substitutions were constructed using SWISS-MODEL (<http://swissmodel.expasy.org/>)

has indicated the importance of residues Lys249, Pro258, Ser329 and Glu355, except that a salt bridge between Lys249 and Glu246 has been mentioned along with other 28 salt bridges (Alzari et al. 1996). None of the four residues are located within the D-glucosyl-binding cavity (Fig. 4), which is consistent with our results that substitutions K249R, P258S, S329N and E355G have little effect to the catalytic activity of CelA.

In the crystal structure, the catalytic domain of CelA folds into an $(\alpha/\alpha)_6$ barrel, consisting of six internal, mutually parallel α -helices interconnected by six external helices (Alzari et al. 1996). Lys249 is located near an external α -helix, the H9 helix (Supplementary Fig. S4) on the surface of the protein and may form an intramolecular salt bridge with Glu246, which situates on the H9 helix (Alzari et al. 1996). The universal occurrence of salt bridges in thermophilic proteins is well known to have significant contribution to their stabilities (Chakravarty and Varadarajan 2000). The replacement of this residue by arginine apparently breaks this salt bridge, and

results in deleterious effect to the stability of CelA probably through changing the conformation of H9 α -helix (Fig. 4). Pro258 is located at a type VIII β -turn and buried inside. Proline is frequently found in turns and sequence statistics indicate that proline is preferred to serine at the *i* position of type VIII β -turn (Trevino et al. 2007; Guruprasad and Rajkumar 2000). Pro258 also forms hydrophobic interactions with Val247, Val257, Tyr273 and Ile323 within 5Å, which is considered important in stabilizing thermophilic proteins (Chakravarty and Varadarajan 2002; Melchionna et al. 2006). The P258S substitution reduces the surrounding hydrophobic property and may impair the conformational stability of the β -turn. Ser329 is located on a surface loop near the activity cleft and its side chain might form a hydrogen bond with Asp319. The substitution of S329N results in the lost of this hydrogen bond. Glu355 is located on the surface at the H15 helix, an external α -helix (Supplementary Fig. S4). It forms hydrogen bonds with residues Asn351, Phe352 and Glu359 within the helix, which is believed to play a major role in stabilizing proteins at elevated temperatures (Ragone 2001). The substitution of glycine results in the loss of these intramolecular interactions and might have severely destroyed the conformation of H15 α -helix as shown in Fig. 4.

Stabilization of secondary structure is known to be important for the thermostability of proteins (Chakravarty and Varadarajan 2002; Petukhov et al. 1997). Structural analysis has revealed the destabilization effect of substitutions K249R and E355G towards α -helices H9 and H15 respectively, and P258S to a β -turn. Moreover, impaired helices H9 and H15 result in weakened helix-helix interactions with helices H6, H8, H11 and H12, and helices H1, H12, H14 and H17, respectively (showed on EMBL-EBI Webserver), which might also contribute to the decreased thermostability. On the contrary, the S329N substitution leads to no secondary structure changes, and no significant factor to stabilize the structure motif is observed for Ser329. We speculate that this position would be more tolerant to amino acid changes, and could act as a target site for further modifications to improve the thermostability of CelA. Work on saturation mutagenesis of this position is under way in our laboratory.

Acknowledgments This work was supported by the Program of 100 Distinguished Young Scientists of the Chinese

Academy of Sciences, the Province Science Foundation of Sichuan, China (No. 08ZQ026-023 & 2010SZ0128) and the Knowledge Innovation Program of the Chinese Academy of Sciences (No. KSCX1-YW-11B2).

References

- Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res* 28:e88
- Aiba S, Kitai K, Imanaka T (1983) Cloning and expression of thermostable α -amylase gene from *Bacillus stearothermophilus* in *Bacillus stearothermophilus* and *Bacillus subtilis*. *Appl Environ Microbiol* 46:1059–1065
- Alzari PM, Souchon H, Dominguez R (1996) The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* 4: 265–275
- Chakravarty S, Varadarajan R (2000) Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Lett* 470:65–69
- Chakravarty S, Varadarajan R (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry* 41: 8152–8161
- Fujino T, Ohmiya K (1991) Cloning of the CelB gene encoding endo-1,4-beta-glucanase-2 from *Clostridium josui* in *Escherichia coli* and the properties of the translated product. *J Ferment Bioeng* 72:422–425
- Gold ND, Martin VJ (2007) Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J Bacteriol* 189:6787–6795
- Guerin DM, Lascombe MB, Costabel M, Souchon H, Lamzin V, Beguin P, Alzari PM (2002) Atomic (0.94 Å) resolution structure of an inverting glycosidase in complex with substrate. *J Mol Biol* 316:1061–1069
- Guruprasad K, Rajkumar S (2000) Beta- and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J Biosci (Bangalore)* 25:143–156
- Hoover DM, Lubkowski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* 30:e43
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66:506–577
- Melchionna S, Sinibaldi R, Briganti G (2006) Explanation of the stability of thermophilic proteins based on unique micromorphology. *Biophys J* 90:4204–4212
- Nemeth A, Kamondi S, Szilagyí A, Magyar C, Kovari Z, Zavodszky P (2002) Increasing the thermal stability of cellulase C using rules learned from thermophilic proteins: a pilot study. *Biophys Chem* 96:229–241
- Petre J, Longin R, Millet J (1981) Purification and properties of an endo-beta-1,4-glucanase from *Clostridium thermocellum*. *Biochimie* 63:629–639

- Petukhov M, Kil Y, Kuramitsu S, Lanzov V (1997) Insights into thermal resistance of proteins from the intrinsic stability of their alpha-helices. *Proteins* 29:309–320
- Ragone R (2001) Hydrogen-bonding classes in proteins and their contribution to the unfolding reaction. *Protein Sci* 10:2075–2082
- Schwarz WH, Grabnitz F, Staudenbauer WL (1986) Properties of a *Clostridium thermocellum* endoglucanase produced in *Escherichia coli*. *Appl Environ Microbiol* 51:1293–1299
- Stemmer WPC (1994) Rapid evolution of a protein in-vitro by DNA shuffling. *Nature* 370:389–391
- Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* 43:777–780
- Trevino SR, Schaefer S, Scholtz JM, Pace CN (2007) Increasing protein conformational stability by optimizing β -turn sequence. *J Mol Biol* 373:211–218
- Yao Q, Sun T, Chen G, Liu W (2007) Heterologous expression and site-directed mutagenesis of endoglucanase CelA from *Clostridium thermocellum*. *Biotechnol Lett* 29:1243–1247
- Zhao HM, Arnold FH (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res* 25:1307–1308
- Zverlov VV, Kellermann J, Schwarz WH (2005) Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: Identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics* 5:3646–3653