Molecular analysis of hyperthermophilic endoglucanase Cel12B from *Thermotoga maritima* and the properties of its functional residues

Hao Shi¹,²,³, Yu Zhang¹,², Liangliang Wang¹,², Xun Li¹,², Wengqian Li¹,²,³, Xiangqian Li³* and Fei Wang¹,²*

**Abstract**

**Background:** Although many hyperthermophilic endoglucanases have been reported from archaea and bacteria, a complete survey and classification of all sequences in these species from disparate evolutionary groups, and the relationship between their molecular structures and functions are lacking. The completion of several high-quality gene or genome sequencing projects provided us with the unique opportunity to make a complete assessment and thorough comparative analysis of the hyperthermophilic endoglucanases encoded in archaea and bacteria.

**Results:** Structure alignment of the 19 hyperthermophilic endoglucanases from archaea and bacteria which grow above 80°C revealed that Gly30, Pro63, Pro83, Trp115, Glu131, Met133, Trp135, Trp175, Gly227 and Glu229 are conserved amino acid residues. In addition, the average percentage composition of residues cysteine and histidine of 19 endoglucanases is only 0.28 and 0.74 while it is high in thermophilic or mesophilic one. It can be inferred from the nodes that there is a close relationship among the 19 protein from hyperthermophilic bacteria and archaea based on phylogenetic analysis. Among these conserved amino acid residues, as far as Cel12B concerned, two Glu residues might be the catalytic nucleophile and proton donor, Gly30, Pro63, Pro83 and Gly227 residues might be necessary to the thermostability of protein, and Trp115, Met133, Trp135, Trp175 residues is related to the binding of substrate. Site-directed mutagenesis results reveal that Pro63 and Pro83 contribute to the thermostability of Cel12B and Met133 is confirmed to have role in enhancing the binding of substrate.

**Conclusions:** The conserved acids have been shown great importance to maintain the structure, thermostability, as well as the similarity of the enzymatic properties of those proteins. We have made clear the function of these conserved amino acid residues in Cel12B protein, which is helpful in analyzing other undetailed molecular structure and transforming them with site directed mutagenesis, as well as providing the theoretical basis for degrading cellulose from woody and herbaceous plants.

**Keywords:** Cellulose, Conserved amino acid residues, Endoglucanase, Phylogenetic analysis, Thermostability

**Background**

Cellulose is the most abundant organic compound and renewable carbon resource on earth [1]. Biodegradation of cellulose, an abundant plant polysaccharide, is a complex process that requires the coordinate action of three enzymes, among which endoglucanases (EC 3.2.1.4), are able to break the internal bonds of cellulose, and disrupt its crystalline structure, exposing the individual cellulose polysaccharide chains, playing in most important role [2-4]. The degradation is mainly carried out by bacteria, fungi, and protozoa, commensals in the guts of herbivorous animals, as well as the termite *Reticulitermes speratus* [5], from which, there are variety of endoglucanases. The complex chemical nature and heterogeneity of cellulose account for the multiplicity of endoglucanases produced by microorganisms. The activity of different endoglucanases with subtle differences in substrate specificity and mode of action contributes to improvement of the degradation of plant cellulose in natural habitats. There are fourteen families of glycoside hydrolases (GHF) that are used...
for cellulose hydrolysis [6]. More and more extremophiles have been studied in recent years, especially the hyper-thermophilic enzymes. Based on amino acid sequence homologies and hydrophobic cluster analysis, hyperthermophilic endoglucanases obtained from extremophiles, which are widely distributed in terrestrial and marine hydrothermal areas, as well as in deep subsurface oil reservoirs, have been classified into GHF12 [7-14]. As described above, there are hyperthermophilic endoglucanases from archaea, most of which were chosen for sequencing on the basis of their physiology [15]. In addition, many hyperthermophilic endoglucanases gene which have been cloned were found in some heat-tolerant bacteria [16]. Those hyperthermophilic endoglucanases have a common feature that the amino acid sequences are mostly relatively short (less than 400 amino acid residues).

Although many hyperthermophilic endoglucanases of GHF12 amino acids have been reported from archaea and bacteria, a complete survey and classification of all sequences in these species from disparate evolutionary groups, and the relationship between their molecular structures and functions are lacking. The completion of several high-quality gene or genome sequencing projects provided us with the unique opportunity to make an unprecedented assessment and thorough comparative analysis of the hyperthermophilic endoglucanases encoded in archaea and bacteria. The analysis of the full set of hyperthermophilic endoglucanases genes in genomes from diverse species allows a definitive classification of hyperthermophilic endoglucanases and an assessment of their origins, evolutionary relations, patterns of differentiation, and proliferation in the various phylogenetic groups. We are interested in finding answers to the following questions: 1) What are the evolutionary relations among these hyperthermophilic endoglucanases; 2) What is the common feature between these conserved amino acid residues and 3D topological structure; 3) What the mechanism of the heat tolerance among these hyperthermophilic endoglucanases?

The broad analysis in this study provided a comprehensive classification scheme and proposed a molecular structure applicable to all hyperthermophilic endoglucanases. A clear picture of the patterns of endoglucanases classes in different species groups was provided. We identified and classified in this study a higher number of hyperthermophilic endoglucanase amino acids from the GHF12 than previously reported, allowing us to identify their relationships based on the phylogenetic clustering. We found that, similar to archaea, amino acids from hyperthermophilic bacteria are also quite different from the other sequences in GHF12. We characterized several conserved amino acid sites from these endoglucanases and predicted their functionality based on the amino acids similarity among the proteins available in databases. The resulting rich data set of hyperthermophilic endoglucanases from GHF12, comprising 19 sequences, is available downloaded from NCBI (Table 1).

Results

Protein sequences characteristics

GenBank has grown fast in recent years and offer us with much better taxonomic sampling for such BLAST-based analysis [17]. We performed similar BLAST-based analysis for the 19 thermophilic endoglucanase protein sequences (which included the T. maritima endoglucanase sequences), using the nonredundant (nr) database as a reference and recording highest ranking matches. We also searched endoglucanase sequences in several plants, bacteria, fungi and algae sequences including the sequences of the R. speratus, using the protein BLAST search engine with a variety of endoglucanase amino acid sequences as queries for most of the thermophilic endoglucanase, else using endoglucanase as a keyword for searching other amino acid sequences of endoglucanase (Table 1). In most cases, whenever significant similarity to an endoglucanase sequence was identified, the amino acid sequence was excised and homology based protein predictions were performed using the most similar query as a guide. All of these 40 protein sequences range from 252 to 438 amino acid residues in length. Of these sequences, those from archaea and bacteria showed similar lengths, especially for those 19 thermophilic endoglucanase protein sequences where the average percentage composition of the residues cysteine and histidine is only 0.28 and 0.74, which are less frequent in thermophilic proteins according to the statistics of amino acid composition based on MEGA 5 (Table 2).

Phylogenetic analysis

Phylogenetic analysis based on the Maximum-parsimony (MP) and Neighbour-joining (NJ) procedure implemented in PAUP 4.0 [18] and other approaches (see Materials and Methods), indicated that all endoglucanase proteins can be reliably grouped into 3 distinct classes except for the outgroup R. speratus, which belongs to the insect family (Figure 1). Furthermore, from the multiple sequence alignments, the hyperthermophilic endoglucanase proteins belong to the class I, and others belong to class II and III. No obvious differentiations are implied in these 19 protein sequences. It was not surprising that there was a close relationship among 19 protein sequences from bacteria and archaea supported with good bootstrap values based on Maximum-likelihood (ML) tree by using MEGA 5 (Figure 2). It was inferred that the endoglucanases of Dictyoglomus turgidum, Thermotoga napthophilica and Thermotoga maritima which are currently studied in our research group are closely related to the others, although the identity of the amino acid sequences
were shown less than 30% (Figure 1, Figure 2). Therefore, it was postulated that they may have a common origination based on protein evolution. Class II comprises of other 12 proteins from plant, fungi and bacteria, and class III comprises of 8 proteins from bacteria.

### Analysis of conserved and catalytic amino acid residues

For the further analysis of the relationship among 19 hyperthermophilic endoglucanases from bacteria and archaea, those 19 amino acid sequences were aligned again with Clustal X2 (Figure 3). We found that the conserved amino acids of hyperthermophilic endoglucanase in Cel12B (for instance) include Gly30, Pro63, Pro83, Trp115, Glu131, Met133, Trp135, Trp175, Gly227 and Glu229 which are highlighted in red (Figure 3), which is very different from the previously reported data [19,20]. Among these conserved amino acids, two glutamic acid residues might be the catalytic nucleophile and proton donor like lysozyme with acid base catalysis [21], other eight conserved amino acids might be necessary to the thermostability of protein and binding of the substrate.

### Hyperthermophilic protein homology modeling

All the hyperthermophilic protein sequences were rendered using SWISS-MODEL database for protein modeling, but only one good model, Cel12B protein model from *T. maritima*, can be used to describe conserved amino acids in which sites of secondary structure and enzymatic center of protein. As described with Cel12B protein model, Glu131, Glu229, Trp115, Trp135, Trp175 and Met133 residues, comprised the active center of the protein (Figure 4a). Cel12B protein is primarily composed of β-sheet (Figure 4a,b,c,d). Trp115, Glu131, Met133, Trp135 and Gly227 residues are in the β-sheet; Pro63 and Trp175 residues are in the turn; and Gly30, Pro83 and Glu229 residues are in the random coil (Figure 4b,d).

### Analysis of site-directed mutagenesis

Base on the homology modeling, the functional amino acid residues Glu64, Pro63, Pro83 and Met133 of Cel12B were selected to be mutated. The results showed that the P63K, P83K, M133W, E64H, E64T and E64l mutant enzymes dramatically inhibited the enzyme activity of Cel12B toward CMC-Na, while E64S mutant protein apparently increased the enzyme activity (Table 3).

### Discussion

Endoglucanases isolated from hyperthermophilic organisms are more active and stable at higher temperatures than their counterparts from mesophiles. In addition, they may be more appropriate for degradation of the cellulose. Since the enzyme activity of those hyperthermophilic endoglucanases is not high for degradation, the

| Table 1 The phylogenetic distribution of endoglucanases from glycoside hydrolase family 12 |
|-----------------------------------------------|----------------|----------------|
| **Organism**                                  | **Length** | **GenBank number** |
| Euryarchaeota                                 | 396        | ADL19785         |
| Ignisphaera aggregans                         | 360        | ADM27702         |
| Metallosphaera cuprina                        | 326        | AEB95090         |
| Pyrococcus furiosus                           | 319        | AAD54602         |
| Sulfolobus acidocaldarius                     | 311        | AAY81158         |
| Sulfolobus islandicus                         | 332        | ADX81754         |
| Sulfolobus islandicus                         | 334        | ACP37717         |
| Sulfolobus islandicus                         | 334        | ACR41545         |
| Sulfolobus islandicus                         | 334        | ADX84872         |
| Sulfolobus islandicus                         | 334        | AAK42142         |
| Thermococcus sp.                              | 319        | EE87358          |
| Thermoproteus tenax                           | 263        | CCCC80138        |
| Thermoproteus uzoniensis                      | 252        | AEA12777         |
| Vulcanisaeta distributa                       | 300        | ADN509821        |
| Bacteria                                      | 439        | ZP_07030982      |
| Bacillus licheniformis                        | 261        | AAP44491         |
| Dictyoglomus turgidum                         | 288        | YP_002352530     |
| Paenibacillus mucilaginosus                   | 266        | AEH3442          |
| Spirochaeta thermoephila                      | 438        | ADN02999         |
| Spirochaeta thermoephila                      | 433        | AEJ62362         |
| Teredinibacter turnerae                       | 278        | ACR14297         |
| Thermobispora bispora                         | 393        | ADG87082         |
| Thermotoga naphthophila                       | 274        | YP_003346783     |
| Thermotoga maritima                           | 275        | Z69341           |
| Lysobacter enzymogenes                        | 383        | ABE51435         |
| Bacillus megaterium                           | 345        | ADE69644         |
| Streptococcus dysgalactiae                    | 366        | BABH0742         |
| Streptococcus dysgalactiae                    | 366        | YP_002995956     |
| Bacillus thuringiensis                        | 349        | ZP_04030866      |
| Fungi                                         |            |                 |
| Stachybotrys ochinata                         | 237        | AF435067         |
| Aspergillus fumigatus                         | 378        | EDP05688         |
| Aspergillus fumigatus                         | 378        | XP_751495        |
| Neosartorya fischeri                          | 381        | XP_001266710     |
| Aspergillus niger                             | 396        | XP_001400178     |
| Penicillium marneffei                         | 379        | XP_002176625     |
| Takaryomyces stipitatus                       | 503        | XP_002481822     |
| Ajellomyces dermatitidis                      | 357        | XP_002621187     |
| Plantania                                     |            |                 |
| Arabidopsis thaliana                          | 484        | BAB11001         |
| Thalassiosira pseudonana                      | 499        | XP_002287341     |
| Insect                                        |            |                 |
| Reticulitermes speratus                       | 448        | AB019095         |

*All the sequences are downloaded from GenBank (http://www.ncbi.nlm.nih.gov/protein/).
hyperthermophilic modification by using genetic engineering is essential. Few structures on databases have been reported so far for transforming those enzymes. In this paper, nineteen sequences of hyperthermophilic endoglucanases were aligned and used for phylogenetic tree construction and molecular modeling to illustrate the relationship between structure and thermostability.

The features of the nature environment of ancestral organism can be inferred by reconstructing phylogenetic tree using amino acid sequences of these organisms [22]. From the alignment of the amino acid sequences, the hyperthermophilic proteins from bacteria and archaea are clustered together based on the phylogenetic tree (Figure 1). Archaea, known to be an ancient organisms on earth, grow in strictly anaerobic environment (terrestrial solfataric springs, hydrothermal areas, and deep subsurface oil reservoirs) at high temperature (generally above 80°C), and hyperthermophilic bacteria also live in the same conditions [13,23]. Therefore, it is inferred that endoglucanases from hyperthermophilic microorganisms from GHF12 could share the similar enzymatic properties and catalytic mechanism.

The stability of thermostophilic proteins depend on several amino acid residues and structural factors [24]. Specific amino acid composition plays a critical role in the thermostability of hyperthermophilic endoglucanase, with the fewest cysteine and histidine residues that are thermal stability among the whole protein sequences by using statistical comparison of the amino acid composition [25,26]. Consistent with this feature, the average content of cysteine and histidine in our research is only 0.24 and 0.72 respectively (Table 2).

Ten conserved amino acids were found by the alignment of nineteen hyperthermophilic protein sequences (Figure 3), that we hypothesize may play a significant role in proton donation, substrate binding as well as the high thermostability. Among these nineteen amino acid sequences, only the three-dimensional structure of endoglucanase from T. maritima could be obtained (Figure 4), since there is no suitable template for other proteins homologous modeling. Thus, the relationship between the ten amino acid residues of these endoglucanases and their molecular structures will be illustrated in Cel12B protein from T. maritima. The substitution of non-Gly residue with Gly residue can be used as one of the general strategies to enhance the protein stability [27,28]. In our study, residues Gly30 and Gly227 located in random and β-sheet, respectively, might contribute to the thermostability of the protein (Figure 4b,d).

It is believed that loop and turn are the weak connections among the protein secondary structure elements, but recently it was demonstrated that they played a key role in thermostability of protein, especially for the proteins that proline is located in loop or turn region [29].
Proline in the polypeptide chain possesses less conformational freedom than other amino acids, as the pyrrolidine ring of proline imposes rigid constrains on the N-C rotation and restricts the available conformational space of the preceding residue. Therefore it can bend the polypeptide chain on itself so as to prepare the backbone much more easily to form the hydrogen bonds with the polar side chains of other turns; meanwhile, the hydrophobic part of proline can interact with the adjacent hydrophobic cavity [30,31]. Compared to mesophilic proteins, thermophilic proteins contain more proline residues especially occurring at the turn, with higher frequency, as well as the shorter loop region of the glucosidase. As the consequence of the flexibility reduction of the polypeptide chain, the protein thermostability can be increased by introducing prolines at specific sites based on the facts that illustrated above [29,31,32]. Hence, residues Pro63 and Pro83, located in the turn and random coil respectively (Figure 4c,d), could provide closer packing of each region, as assumed for thermostability of protein. And then, it was finally confirmed by experimental results. Compared to other amino acids, lysine has longer side-chain groups and more vibrational degree of freedom, and it is more sensitive to the temperature. When the proline is substituted with lysine, the vibration of side-chain groups rises up at high temperature, and then the thermostability of the Cel12B decrease dramatically. Therefore, it is confirmed that residues Pro63 and Pro83 play an important role in stabilizing the Cel12B.

The crystal structure and protein molecular simulation supported that two glutamic acid residues are the catalytic nucleophile and proton donor that have been reported in many enzymes, lysozyme, xylanase as well as endoglucanase [33]. So, Glu131 (in β-sheet) and Glu229 (in random coil) residues are the proton donor and

Figure 1 The phylogenetic tree obtained using the endoglucanases and outgrouped by the protein sequence of R. speratus. The NJ (a) and MP (b) tree were generated using program PAUP 4.0 beta 10 Win on 40 aligned amino acids. All the protein sequences are from Table 1. Proteins from hypertherophilic bacteria and archaea are shown within light blue colored boxes (I). Other proteins from bacteria, fungi and plants are shown within yellow (II) and blue (III) colored boxes.
catalytic nucleophile respectively (Figure 4b,d). Although the chemical nature of the tryptophan residue in the catalytic center does not significantly affect the conformational properties of lysozyme, it exhibited a pronounced effect on the binding of substrate and the enhancement of the total enzyme activity [34]. It was reported that structural changes at the active site (W95L) of alcohol dehydrogenase from Sulfobolus solfataricus are consistent with the reduced activity on substrates and decreased coenzyme binding [35]. Therefore, we propose that three tryptophan residues (Trp115, 135 and 175, Figure 4b,c) of Cel12B protein may be essential in mediating the total cooperativity of the response of the enzyme to substrate. Met133, located in the middle of Trp135 and Glu131 in β-sheet (Figure 4b), is predicted to be related to the binding of substrate and also finally confirmed by experimental results. When it is replaced by tryptophan residue, the enzyme activity is significantly decreased. With the homology modeling result (data not shown), it is inferred that Glu64 is probably another functional acid amino located near the catalytic center. It is supposed that residue Glu64 might contribute to stabilizing the intermediate product. Maintaining the intermediate product may be caused by the interaction of side-chain group of Glu64. Polar amino acids, histidine and threonine are able to stabilize the intermediate product to some extent. However, their side-chain groups are relatively large, and possess larger steric hindrance, thus lead to decrease of the enzyme activity. Compared to glutamic acid, histidine and threonine, serine has smaller side-chain group and steric hindrance, so it can easily form hydrogen bond with product and stabilize it, and then increase the enzyme activity.

**Conclusions**

Nineteen hyperthermophilic homologous protein sequences from GHF12 were aligned and used for constructing phylogenetic tree. It was inferred from the nodes that there is a close relationship among these nineteen homologous endoglucanases from hyperthermophilic bacteria and archaea. We have made clear the function of these conserved amino acids in Cel12B protein, which is helpful in analyzing other molecular structure and transforming them with site directed mutagenesis.

**Methods**

**Extraction of sequences from databases**

Thorough BLASTP searches for several divergent endoglucanases of plants, animals, bacteria, fungi, alga and archaea were performed to retrieve endoglucanases genes through NCBI, PDB (http://www.rcsb.org/pdb/home/home.do), UniProt (http://www.uniprot.org/) database server. Hyperthermophilic endoglucanase amino acid sequence was used (GenBank No: Z69344) [16] as a BLAST query for seeking hyperthermophilic endoglucanases from bacteria and
archaea. New rounds of BLASTP searches for the nr protein and GenBank databases at NCBI restricted to plant or other organisms were carried out using representative endoglucanase from different classes of plants, bacteria, fungi and algae as queries.

Multiple sequence alignment and phylogenetic analysis
One of the most widely used bioinformatics analysis is multiple sequences alignment, and it needs several widely used software packages to analysis. In this study, the multiple sequence alignment tool Clustal X2 was

Figure 3 Alignment of 19 endoglucanases amino acids sequences using CLUSTAL X2.0. The highly conserved amino acids are colored in red.
used for sequence alignment [36]. Sequences were further edited using the MEGA 5 when necessary and aligned manually [37]. In the phylogenetic analysis, sequences were trimmed so that only the relevant conserved domains were remained in the alignment. Phylogenetic relationships were inferred using the NJ and MP methods as implemented in PAUP 4.0 [18] while the Maximum-Likelihood method as implemented in MEGA 5 [37]. The NJ, MP and ML trees, displayed using TREEVIEW 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html), were evaluated with 1000 bootstrap replicates.

**Secondary structure prediction**

For homology modeling, the crystal structure of the thermophilic endoglucanase (PDB ID: 3AAM) obtained implemented in PAUP 4.0 [18] while the Maximum-Likelihood method as implemented in MEGA 5 [37]. The NJ, MP and ML trees, displayed using TREEVIEW 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html), were evaluated with 1000 bootstrap replicates.

### Table 3 Effect of site-directed mutagenesis on enzyme activity

| Strain  | Optimum temp (°C) | Specific activity (U mg⁻¹) | Relative activity (%) |
|---------|-------------------|----------------------------|-----------------------|
| Control | 90                | 105 ± 3.4                  | 100 ± 3.2             |
| E64T    | 85                | 53 ± 1.3                   | 50 ± 1.2              |
| E64H    | 85                | 25 ± 1.0                   | 24 ± 1.0              |
| E64L    | ND                | 0                          | 0                     |
| E64S    | 90                | 133 ± 2.5                  | 127 ± 2.4             |
| P63K    | ND                | 0                          | 0                     |
| P83K    | ND                | 0                          | 0                     |
| M133W   | ND                | 0                          | 0                     |

**ND**: not determined. Values shown were the mean of triplicate experiments.

### Table 4 Nucleotide sequences of used primers

| Primers | Nucleotide sequence                        |
|---------|--------------------------------------------|
| Forward 1 | 5'-AGTAGATNNNGGATATCCCATGCACCCAGGC-3'      |
| Reverse 1 | 5'-ACGCGTACACGCTCCGTCGAAAGGGCCG-3'        |
| Forward 2 | 5'-CATGATATACAGAGATCTACTACGTTACAAG-3'     |
| Reverse 2 | 5'-CACCCAGCTGTGATGATGTAAGGAGG-3'          |
| Forward 3 | 5'-GAATTTCTTTAAGCTGAAGGAGTGAAGAACCTCC-3'  |
| Reverse 3 | 5'-AACACCGCTGTGATCCTACGCTGCGAAGG-3'       |
| Forward 4 | 5'-CGGAGATCTGGGTTGTGTTCTACAACACGTC-3'     |
| Reverse 4 | 5'-CGTCACCCGAAAGAAGAGAGGTC-3'             |
from Protein Data Bank (PDB) was used as a template. The aligned sequences were submitted to SWISS-MODEL (http://www.expasy.org/swissmod/) to obtain the 3D structure of the endoglucanases [38-40]. The model was viewed using Swiss-PDB Viewer [41], and the quality of the model was evaluated by the local model quality estimation on SWISS-MODEL. The 3D structure of the protein was further modified by PyMOL (version 1.4.1, http://www.pymol.org/).

Test of functional residues
Site-directed mutagenesis was used to analyze the related functional amino acid residues using reverse PCR. Restriction enzymes, DNA polymerase, *DpnI*, T4 polynucleotide kinase and T4 ligase were purchased from Takara (Dalian, China) and used according to the manufacturer’s instructions. The sequence of cel12B gene (GenBank Protein No. Z69341) based on the *T. maritima* genomic DNA was amplified using primers 5′-GGAATTCCATATGAGGTGGG-CAGTTCTTCTGA-3′ and 5′-CCGCTCGAGATTACTCGAGTTTCTTCTGA-3′ (primers with the added compatible restriction sites of *NdeI* and *XhoI*, respectively). PCR was performed as follows: 94°C, 5 min; 30 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 50 s; and 72°C, 10 min. The recombinant vector was constructed as follows: the amplified PCR products were purified, digested with *NdeI* and *XhoI*, and then ligated into pET-20b vector at the corresponding sites. Reverse PCR amplifications were conducted by high-fidelity *Pyrobest* DNA polymerase using recombinant pET-20b-cel12B as templates, and primers were shown in Table 4. The templates were cleaned away from the products using *DpnI*. Then, the resulting products were purified with BIOMIGA PCR Purification Kit (Shanghai, China), followed by phosphorylation resulting products were purified with BIOMIGA PCR Purification Kit (Novagen, USA) and the bounded proteins were eluted by discontinuous imidazole gradient.

Enzyme activity was determined using 5-dinitrosalicylic acid (DNS) method [42]. The reaction mixture, containing 50 mM imidazole-potassium buffer (pH 6.0), 0.5% sodium carboxymethyl cellulose (CMC-Na), and a certain amount of endoglucanase (0.1 μg) in 0.2 mL, was incubated for 10 min at 85°C. The reaction was stopped by the addition of 0.3 mL DNS. The absorbance of the mixture was measured at 520 nm. One unit of enzyme activity was defined as the amount of enzyme necessary to liberate 1 μmol of reducing sugars per min under the assay conditions. All the values of enzymatic activities shown in figures were averaged from three replicates.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
HS conceived the project, carried out phylogenetic phylogenetic analysis, LW, XL, YZ and WL carried out database searches and protein modeling, and FW and XL supervised the work. HS, XL, and FW wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This work was financially supported by the National Natural Science Foundation of China (No. 31170537), Jiangsu Provincial Government (CXZZ11_0526), Doctorate Fellowship Foundation of Nanjing Forestry University, as well as A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Author details
1. College of Chemical Engineering, Nanjing Forestry University, Nanjing 210037, China. 2. Jiangsu Key Lab of Biomass-Based Green Fuels and Chemicals, Nanjing 210037, China. 3. Department of Life Science and Chemistry, Huaiyin Institute of Technology, Huaiian 223003, China.

Received: 3 July 2013 Accepted: 13 February 2014

Published: 17 February 2014

References
1. Wang T, Liu X, Yu Q, Zhang X, Qu Y, Gao P: Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei*. Biomol Eng 2005, 22:1-9.
2. Liang C, Fioroni M, Rodriguez-Ropero F, Xue Y, Schwaneberg U, Ma Y: Directed evolution of a thermophilic endoglucanase from *Clostridium thermocellum* celulosomal endoglucanase Cel8A by a single glycine substitution. Chemcatchem 2010, 28(9):1003–1009.
3. Nakazawa H, Okada K, Onodera T, Ogasawara W, Okada H, Morikawa Y: Directed evolution of endoglucanase III (Cel12A) from trichoderma reesei. Appl Microbiol Biotechnol 2009, 83(4):649–657.
4. Anbar M, Lamed R, Bayer EA: Thermostability enhancement of *Clostridium thermocellum* cellulosomal endoglucanase Cel8A by a single glycine substitution. Chemcatchem 2010, 28(9):1003–1009.
5. Nakazawa H, Okada K, Onodera T, Ogasawara W, Okada H, Morikawa Y: Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. J Biotechnol 2011, 154(4):46–53.
6. Watanabe H, Noda H, Tokuda G, Lo N: A cellulase gene of termite origin. Nature 1998, 394(6691):330–331.
7. Davison A: Ancient origin of glycosyl hydrolase family 9 cellulase genes. Mol Biol Evol 2005, 22(5):1273–1284.
8. Mardanov AV, Svetlitchnyi VA, Beletsky A, Prokofeva M, Bonch-Osmolovskaya E, Ravin NV, Skryabin KG: The genome sequence of the crenarchaeon *Acidilobus saccharovorans* supports a new order, Acidilobales, and suggests an important ecological role in terrestrial acidic hot springs. Appl Environ Microbiol 2010, 76(16):5652–5663.
9. Reno ML, Held NL, Fields CJ, Burke PV, Whittaker RJ: Biogeography of the *Sulfobolus islandicus* pan-genome. Proc Natl Acad Sci USA 2009, 106(21):8605–8610.
10. Guo L, Brugger K, Liu C, Shah SA, Zheng H, Zhu Y, Wang S, Lillestøl RK, Chen L, Frank J, et al: Genome analyses of Icelandic strains of *Sulfobolus islandicus*, model organisms for genetic and virus-host interaction studies. J Bacteriol 2011, 193(7):1672–1680.
10. Göker M, Held B, Lapidus A, Nolan M, Spring S, Yasawong M, Lucas S, Giavina Del Rio T, Tice H, Cheng JF, et al. Complete genome sequence of *Igneococcus agaricus* type strain (AQ1151T). *Stand Genomic Sci* 2010, 3(1):165–76.

11. Angelova A, Liebl S, Ballismacher M, Boemericke M, Lehmann R, Liesegang H, Daniel R, Liebl W. Genome sequence of the polysaccharide-degrading, thermophilic anaerobe *Spironothea thermophilus* DSM 6192. *J Bacteriol* 2010, 192(24):6492–6493.

12. Mardanov AV, Gumerov VM, Beletsky AV, Prokofeva MI, Bonch-Osmolovskaya EA, Svetlitchnyi VA, Ravin NV, Skryabin KG. Complete genome sequence of the thermophilic archaeon *Thermostreptococcus uzonensis* 768–20. *J Bacteriol* 2011, 193(12):3156–3175.

13. Chen LM, Brugger K, Skovgaard E, Bonch-Osmolovskaya EA, Skryabin KG. Analysis of the increase in number of proline residues and the rise in thermostability of *Fervidobacterium nodosum* cellulase from a crenarchaeon as revealed by genome analysis. *J Bacteriol* 2010, 192(33):9109–9117.

14. Liu L-J, You X-Y, Zheng H, Wang S, Jiang C-Y, Liu S-J. Thermostabilization of *D*-xylose isomerase by introduction of glycine-15 in the bend between two alpha-helices can explain the role of the tryptophan-62 residue in the structure and function of lysozyme. *Biochimie* 2010, 92(3–4):344–349.

15. Wilgenbusch JC, Swofford D. Dissecting alpha-helices: position-specific analysis of the sequence of lysozyme. *Bioinformatics* 2003, 19(12):1427–1435.

16. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Selman V, Stamatakis A, Larkin AM, Thompson JD, et al. Clustal W and clustal X version 2.0. *Bioinformatics* 2007, 23(21):2947–2948.

17. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGAS: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28(10):2731–2739.

18. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modelling server. *Nucleic Acids Res* 2003, 31(13):3381–3385.

19. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997, 18(15):2714–2723.

20. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* 1959, 31(3):426–428.