Directed Evolution of Beta-galactosidase from *Escherichia coli* into Beta-glucuronidase

Ai-Sheng Xiong1,2, Ri-He Peng2, Jing Zhuang1,2, Jin-Ge Liu1,2, Fang Xu1,2, Bin Cai1,2, Zhao-Kui Guo3, Yu-Shan Qiao1, Jian-Min Chen4, Zhen Zhang1,* and Quan-Hong Yao2,*

1College of Horticulture, Nanjing Agricultural University, Nanjing, 210095, China
2Biotechnology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, 201106, China
3Heilongjiang Tobacco Research Institute, Mudanjiang, 157011, China
4College of Bioscience and Biotechnology, Yangzhou University, Yangzhou, 225009 China

Received 27 October 2006, Accepted 15 January 2007

*To whom correspondence should be addressed.
Tel: 86-25-84395266 (Z Zhang), 86-21-62203180 (Q-H Yao)
Fax: 86-25-84395266 (Z Zhang), 86-21-62205704 (Q-H Yao)
E-mail: yaoquanhongcn@yahoo.com.cn (Q-H Yao), zhangzhen_nj@yahoo.com.cn (Z Zhang)

In vitro directed evolution through DNA shuffling is a powerful molecular tool for creation of new biological phenotypes. *E. coli* β-galactosidase and β-glucuronidase are widely used, and their biological function, catalytic mechanism, and molecular structures are well characterized. We applied an *in vitro* directed evolution strategy through DNA shuffling and obtained five mutants named YG6764, YG6768, YG6769, YG6770 and YG6771 after two rounds of DNA shuffling and screening, which exhibited more β-glucuronidase activity than wild-type β-galactosidase. These variants had mutations at fourteen nucleic acid sites, resulting in changes in ten amino acids: S193N, T266A, Q267R, V411A, D448G, G466A, L527I, M543I, Q626R and Q951R. We expressed and purified those mutant proteins. Compared to the wild-type protein, five mutant proteins exhibited high β-glucuronidase activity. The comparison of molecular models of the mutated and wild-type enzymes revealed the relationship between protein function and structural modification.

**Keywords:** β-galactosidase, β-glucuronidase, Directed evolution, DNA shuffling, Enzyme properties, Structure-function analysis

**Introduction**

Approximately fifty marker genes used for molecular research have been assessed for efficiency, biosafety, scientific application and commercialization (Miki and McHugh, 2004). Some reporter genes can be adapted to function as selectable marker genes through the development of novel substrates. *E. coli* β-galactosidase, encoded by *lacZ*, is widely used. It is a 464 kDa homotrimer. The ~1,023 residues of each subunit is composed of five domains. The enzyme utilizes the external substrates 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) for histological localization (Jacobson et al., 1994; Huber et al., 1994; Matthews, 2005; Juers et al., 1999). The bacterial enzyme β-glucuronidase (GUS), encoded by the *E. coli gusA* gene, is another widely used reporter. The enzyme utilizes the external substrates 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) for histological localization (Jefferson et al., 1987). β-glucuronidase is the most widely used reporter gene in screening for genetically engineered plants, as indicated in the database of 2001 and 2002 field trials in the United States (Miki and McHugh, 2004). Its popularity is attributed to high stability in plant tissues and lack of toxicity even when expressed at high levels. Histochemical GUS staining protocol is a simple, rapid, highly-reliable and cost-effective method for analysis of transgenic plants (Dixon et al., 2003). Also, no specialized equipment is needed for histochemical assays for GUS activity. Moreover, the β-glucuronidase gene and its protein are considered safe for the environment and humans (Castle et al., 2004).

DNA recombination is a powerful tool for the creation of new phenotypes. Recently, methods for *in vitro* DNA shuffling have been developed and applied to the evolution of novel molecules in the laboratory. An exciting new development is the shuffling of homologous genes to create diversity for directed evolution (Stemmer, 1994). DNA shuffling has already led a significant impact on the nascent field of directed evolution; each new demonstration has expanded the method's utility, generating broad reagents for both academic and commercial interests (Chirumamilla et al., 2001; Lassner and
Many significant studies have been carried out initially on reporter genes because of their easy detection. *E. coli* β-galactosidase variants with enhanced β-β-galactosidase activity were evolved by DNA shuffling (Zhang et al., 1997) and site-saturation mutagenesis (Parikh and Matsumura, 2005). *E. coli* β-glucuronidase has also been used as a model for studying in vitro directed evolution (Rowe et al., 2003; Xiong et al., 2007). Using DNA shuffling to direct the evolution of *E. coli* β-glucuronidase, variants were selected with improved β-glucuronidase activity (Matsumura and Ellington, 2001) and thermostability (Flores and Ellington, 2002; Xiong et al., 2002). Matsumura et al. obtained a GUS variant that was significantly more resistant to both glutaraldehyde and formaldehyde than the wild-type enzyme (Matsumura et al., 1999). A GFP mutant with stronger fluorescence was also selected by directed evolution (Crameri et al., 1996; Matsumura and Ellington, 1996; Nam et al., 2003). A GFP-like protein with a shift in fluorescence to a longer wavelength region was also created by DNA shuffling (Fradkov et al., 2000). Some GUS variants were also obtained with increased p-nitrophenyl-β-D-xylopyranoside activity (Geddie and Matsumura, 2004).

In our research, we obtained some *E. coli* β-galactosidase mutants with improved β-glucuronidase activity using a high efficiency and throughput system in the in vitro directed evolution of a reporter gene.

### Materials and Methods

**Chemicals, enzymes, plasmids and strains.** All reagents were purchased from Sigma Chemical Co., Ltd (St Louis) and Sinopharm Chemical Reagent Co., Ltd (Shanghai). p-nitrophenyl-β-D-glucuronide (pNPG) and o-nitrophenyl-β-D-galactopyranoside (oNPG) were purchased from Shanghai Sangon Biological Engineering Technology & Services Co., Ltd (Shanghai). Enzymes used for molecular biology were purchased from Promega (Madison). X-Gal and Taq DNA polymerase were purchased from Takara Co., Ltd (Dalian). The DNA sequencing kit was purchased from Applied Systems (Foster City). Protein markers were purchased from Watson Bio-tech Co., Ltd (Shanghai). The QiAquick gel extraction kit was purchased from Qiagen (Stanford). X-Gluc was purchased from Clontech (Palo Alto). *E. coli* strains were purchased from Stratagene (La Jolla). The prokaryotic expression vector pYPX251 containing a promoter of the *aacC1* gene and T7T2 transcription terminator was constructed in our laboratory and deposited in GenBank (GenBank number: AY178046).

**DNA shuffling, Library construction and screening.** The *lacZ* gene was amplified from the genome of the K12 *E. coli* strain using the oligonucleotides R3742, R3743 by high fidelity PCR (Table 1). The PCR product was used as template or wild-type control. The DNA shuffling was performed as described by Xiong et al. (Xiong et al., 2007) with modifications as follows: fragments of 50-100 bp were purified from 10% PAGE by electrophoresis using a dialysis bag. After the primer-less PCR and primer PCR using oligonucleotides R3742, R3743 as primers, a collection of full length *lacZ* gene mutants were obtained and digested with BamHI and SacI enzymes and the isolated fragments were cloned into pYPX251. The mutant DNA library was transferred into the *E. coli* strain DH5α by electroporation and plated on LB plates containing 100 μg/ml ampicillin. After 16 h of growth at 37°C, colonies were absorbed onto a nitrocellulose filter and transferred colony-side-up to a Petri dish. The nitrocellulose filter papers were incubated in 0.4 mg/ml X-Gluc at 37°C with GUS buffer solution (30 mM sodium phosphate, pH 7.0; 1 mM EDTA; 5 mM 2-mercaptoethanol). Standard DNA manipulation methods were used in plasmid construction and *E. coli* transformation (Sambrook and Russell 2001).

**DNA sequencing.** The evolved mutant *lacZ* genes were sequenced using the Applied Biosystmes Big Dye protocol and the instrument 377 at our lab. The 3'-end of every evolved mutant *lacZ* genes was sequenced using the primer W73 on the vector pYPX251. The mutant *lacZ* genes were sequenced in their entirety, using the following additional primers: W73, Wxas01, Wxas02, Wxas03, Wxas04 (Table 1).

**Protein purification, enzyme characterization and kinetics.** Each protein was fused to an N-terminal six histidine-tag and all evolved enzymes were purified to homogeneity using nickel chelate affinity chromatography. The quality of protein preparations was judged by SDS-PAGE analysis and Coomassie Blue staining (Sambrook and Russell 2001). The optimal pH of all mutant proteins and wild-type enzymes were determined at 37°C at everyhalf point between 4.0 and 100. The activities were measured at 5-degree intervals between 30 and 80°C to determine the optimal temperature. Thermostabilities of the expressed enzymes at 70°C were determined by measuring the remaining activity at 37°C after heating expressed protein samples, then chilling on ice. One unit of

<table>
<thead>
<tr>
<th>Name</th>
<th>Function</th>
<th>Oligonucleotides sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>R3742</td>
<td>Shuffling</td>
<td>5-GAA,GGA,TCC,ATGATT,ACGGAT,TCAC-3'</td>
</tr>
<tr>
<td>R3743</td>
<td>Shuffling</td>
<td>5-AGGAAGCTC,TTA,TGGAC,CCA,GAC,CAA,C-3'</td>
</tr>
<tr>
<td>W73</td>
<td>Sequence</td>
<td>5-GTA,GGA,TAT,GCCTC,ACGGCA,CTGG-3'</td>
</tr>
<tr>
<td>Wxas01</td>
<td>Sequence</td>
<td>5-GCG,ATT,TTCCC,ATGCC,TCCT-3'</td>
</tr>
<tr>
<td>Wxas02</td>
<td>Sequence</td>
<td>5-GTC,GGCTG,GGGAGA,TTA,ATCC,AGG-3'</td>
</tr>
<tr>
<td>Wxas03</td>
<td>Sequence</td>
<td>5-CAG,TTT,TGT,ATG,AAAGC,CTT,CTG-3'</td>
</tr>
<tr>
<td>Wxas04</td>
<td>Sequence</td>
<td>5-GAT,AAA,GAC,ATT,GGC,GTATA,AGT,3'</td>
</tr>
</tbody>
</table>
In vitro Evolution of a Beta-galactosidase into Beta-glucuronidase

β-galactosidase will hydrolyze 1.0 µmol of oNPG per minute at a given temperature and pH. The unit of β-glucuronidase activity is the enzyme activity that increases the rate of release of 1 µM pNPG at a given temperature and pH. Each value was derived from at least three independent reactions.

Molecular modeling. We visualized the mutant enzymes by molecular modeling, using SWISS-MODEL (http://swissmodel.expasy.org/). The three-dimensional structures of mutant enzymes were aligned with the template structure of E.coli β-galactosidase, which was previously obtained by X-ray crystallography (Jacobson et al., 1994).

Results

Directed evolution and screening. Using the DNA shuffling system described by Xiong et al. (Xiong et al., 2007), we shuffled the lacZ gene and developed a highly efficient, high throughput system for in vitro directed evolution of the Lac Z reporter gene. Over 200,000 variant colonies in each round of DNA shuffling were screened using the pYPX251 vector. The colonies formed on the culture medium were fixed onto a nitrocellulose filter. The filter paper was then incubated in 0.4 mg/ml X-Gluc at 37°C for 30 min. Fifteen colonies became blue and were isolated. These 15 colonies were mixed and used as a template in the second round of DNA shuffling. In the second round, The filter paper was then incubated in 0.4 mg/ml X-Gluc at 37°C for 15 min. Five of the mutants, named YG6764, YG6768, YG6769, YG6770 and YG6771 were identified because their rate of color generation was significantly faster than the wild-type enzyme.

Sequencing. The five variants isolated in the second round of evolution were sequenced, which revealed nucleotide changes in fourteen sites (Table 2), and deduced ten amino acid substitutions (YG6764: L527I, M543I; YG6768: T266A, Q267R, Q626R; YG6769: S193N, G466A, Q951R; YG6770: Q193N, T266A, Q267R, Q626R; YG6771, S193N, V411A, D448G).

Protein purification and enzyme characterization. Wild-type β-galactosidase (LacZ-WT) and variants, which show high β-glucuronidase activity (YG6764, YG6768, YG6769, YG6770 and YG6771), were purified to homogeneity using nickel chelate affinity chromatography. Protein preparations were judged by SDS-PAGE analysis and Coomassie Blue staining. Lane M: protein marker; Lane WT: Purified wild-type β-galactosidase; Lane 1: Purified YG6764; Lane 2: Purified YG6768; Lane 3: Purified YG6769; Lane 4: Purified YG6770 and Lane 5: Purified YG6771.

| Table 2. Sequences of those mutants which exhibited higher β-glucuronidase activity |
|-----------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Position | 578 | 764 | 796 | 800 | 858 | 1232 | 1397 | 1579 | 1629 | 1877 | 1953 | 2364 | 2852 |
| WT       | G   | T   | A   | A   | T   | T   | A   | G   | C   | G   | A   | G   | A   | A   | A   | A   | A   |
| YG6764   | G   | G   | G   | G   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   |
| YG6768   | A   | C   | C   | C   | A   | G   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   |
| YG6769   | A   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   |
| YG6770   | A   | C   | C   | C   | G   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   |
| YG6771   | A   | C   | C   | C   | G   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A   |

β-galactosidase will hydrolyze 1.0 µmol of oNPG per minute at a given temperature and pH. The unit of β-glucuronidase activity is the enzyme activity that increases the rate of release of 1 µM pNPG at a given temperature and pH. Each value was derived from at least three independent reactions.

**Table 2. Sequences of those mutants which exhibited higher β-glucuronidase activity**

**Fig. 1.** Purified wild-type β-galactosidase (LacZ-WT) and variants, which show high β-glucuronidase activity (YG6764, YG6768, YG6769, YG6770 and YG6771). Each protein was fused to a N-terminal six histidine-tag and all the evolved enzymes were purified to homogeneity using nickel chelate affinity chromatography. Protein preparations were judged by SDS-PAGE analysis and Coomassie Blue staining. Lane M: protein marker; Lane WT: Purified wild-type β-galactosidase; Lane 1: Purified YG6764; Lane 2: Purified YG6768; Lane 3: Purified YG6769; Lane 4: Purified YG6770 and Lane 5: Purified YG6771.

**Protein purification and enzyme characterization.** Wild-type β-galactosidase (LacZ-WT) and variants, which show high β-glucuronidase activity (YG6764, YG6768, YG6769, YG6770 and YG6771), were purified to homogeneity using nickel chelate affinity chromatography (Fig. 1). With pNPG as a substrate, the specific activity of LacZ-WT and variants were approx. 1.11 U/mg, 2.86 U/mg, 2.63 U/mg, 3.46 U/mg, 2.95 U/mg and 3.01 U/mg at 37°C and pH 7.0. The specific activity of β-glucuronidase, variants of YG6764, YG6768, YG6769, YG6770 and YG6771 were 2.58, 2.37, 3.12, 2.67 and 2.72 times higher than that of the wild-type. With oNPG as a substrate, the specific activity of β-galactosidase, five variants were decreased 25% to 50% than that of the wild-type enzyme (Fig. 2).
was found at pH 7.0 among YG6768, YG6770, YG6771 and LacZ-WT (Fig. 4a). With ONPG as a substrate, the optimal temperature for YG6764, YG6769 and LacZ-WT tested at pH 7.0 were found to be 40°C, while that for YG6768, YG6770 and YG6771 were at 45°C (Fig. 3b). One single activity peak was found at pH 7.0 among YG6764, YG6768 and LacZ-WT, while one single activity peak was found at pH 7.5 among YG6769, YG6770 and YG6771 (Fig. 4b).

We examined thermostability of those expressed variants and wild-type enzymes using pNPG as a substrate. The activity of the wild-type LacZ-WT enzyme decreased significantly after 2 min at 70°C. However, the variants YG6771 and YG6768 retained over 25% of its activity when heated at 70°C for 2 min and over 10% when heated at 70°C for 10 min (Fig. 5a). While using oNPG as a substrate, all variants and wild-type enzyme activities decreased significantly following a heat treatment at 70°C.

**Table 3.** The specific and relative activity of β-glucuronidase and β-galactosidase among all variants and wild-type enzymes

<table>
<thead>
<tr>
<th>Sample</th>
<th>Wild-type</th>
<th>YG6764</th>
<th>YG6768</th>
<th>YG6769</th>
<th>YG6770</th>
<th>YG6771</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-glucuronidase</td>
<td>Activity (U/mg)</td>
<td>1.11</td>
<td>2.86</td>
<td>2.63</td>
<td>3.46</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>Relativity (%)</td>
<td>100</td>
<td>258</td>
<td>237</td>
<td>312</td>
<td>267</td>
</tr>
<tr>
<td>β-galactosidase</td>
<td>Activity (U/mg)</td>
<td>210</td>
<td>162.25</td>
<td>158.84</td>
<td>144.09</td>
<td>108.86</td>
</tr>
<tr>
<td></td>
<td>Relativity (%)</td>
<td>100</td>
<td>77.26</td>
<td>75.64</td>
<td>68.61</td>
<td>51.70</td>
</tr>
</tbody>
</table>

**Fig. 2.** The relative activity of β-galactosidase and β-glucuronidase of wild-type and variant enzymes.

**Fig. 3.** The activity of expressed YG6764, YG6768, YG6769, YG6770, YG6771 and LacZ-WT at various temperatures. (A) With pNPG as a substrate. (B) With ONPG as a substrate.

**Three-dimensional structural analyses.** The three-dimensional structural analyses of mutant β-galactosidase, which contained all ten changed sites from five mutant variants, were aligned with the template structure of E.coli β-galactosidase. Fig. 6 showed the positions of the ten mutated amino acid residues S193N, T266A, Q267R, V411A, D448G, G466A, L527I, M543I, Q626R and Q951R in one subunit of β-galactosidase. β-galactosidase from E.coli is a 464 kDa homotetramer. Each subunit consists of five domains, the third (red in Fig. 6) being an α/β barrel that contains most of the active site residues (Jacobson et al., 1994). Domain 3 is a so-called “TIM barrel” domain. Mutated amino acid residues V411A, D448G, G466A, L527I and M543I reside in the TIM barrel domain. Domain 5 is a β-sandwich with one structural homolog among known protein folds. The residue of Q951R lies in this domain, and residue M543I is found near this domain.

**Discussion**

Despite approximately fifty marker genes that exist for plants, only a few marker genes are used for the majority of plant research and crop development. Many of these genes have specific limitations or have not been sufficiently tested to merit their widespread use (Miki and McHugh, 2004). Its popularity is attributed to high stability in plant tissues and lack of toxicity even when expressed at high levels. Histochemical GUS staining protocol is a simple, rapid, highly-reliable and cost-effective method for analysis of transgenic plants. Also, no specialized equipment is needed.
In vitro Evolution of a Beta-galactosidase into Beta-glucuronidase

423

Some plants have background galactosidase activity, the use of beta-galactosidase in plants has also not been widely adopted (Miki and McHugh, 2004). A form of beta-galactosidase, which also exhibits beta-glucuronidase activity, maybe a good reporter gene for future transgenic research. Such a mutated enzyme may serve as a potential reporter for both enzymes, and contain activities for both beta-galactosidase and beta-glucuronidase.

We attribute the success of shifting the E. coli beta-galactosidase to beta-glucuronidase to some factors. We believed the use of the pYPX251 vector with a moderately strong promoter was the primary reason. A suitable selection strategy was another key factor for in vitro directed evolution because the possibility of obtaining a desirable mutant increases as the screening population increases. Thirdly, separation of fragments using 10% PAGE, and use of 50-100 bp fragments in shuffling the lacZ gene was also likely important (Xiong et al., 2007). Over 200,000 variant colonies in each round of DNA shuffling were screened using the pYPX251 vector in each round of DNA shuffling screening. Moreover, since colorimetric screening of a report gene is detected easily, we isolated the mutants from the library conveniently. Indeed, only two rounds of mutation and screening were required for the selection of five variant species.

E. coli beta-galactosidase, encoded by lacZ, is widely used, and its biological function, catalytic mechanism, and molecular structures are well characterized. Beta-glucuronidase is a versatile and efficient model system for studying enzyme thermostability, and for optimizing a strategy for deducing structure-function relationships of directed evolution in vitro. Accounting for this ease of use is its stability, visible phenotype, the easy of analysis of its enzymatic properties. The lacZ gene and gusA gene diverged from an ancient common ancestor (Henrissat, 1991). E. coli beta-galactosidase and beta-glucuronidase catalyze the hydrolysis of similar glycoside substrates. Matsumura and Ellington found that beta-galactosidase has no detectable activity in reactions with beta-glucuronidase. Beta-glucuronidase exhibits weak beta-galactosidase activity. Using DNA shuffling to direct the evolution of E. coli beta-galactosidase, variants have been selected with improved beta-galactosidase activity (Matsumura and Ellington, 2001). However, there were few studies regarding the shift of E. coli beta-galactosidase to beta-glucuronidase until now. The lag in research may have been due in part, to the cost of the X-Gluc substrate, the difficulty of shuffling, and the cost of sequencing because of the length of the lacZ gene (over 3.0 kb).
Both Mg$^{2+}$ and Na$^+$ were required for maximal activity of β-galactosidase. Five of the mutated amino acid residues (V411A, D448G, G466A, L527I and M543I) reside in the active domain (Jacobson et al., 1994; Matthews, 2005). Those residues that lie in the active site may be contribute to different segments of the polypeptide chain and change the metal cation binding sites. The N-terminal region of the polypeptide chains helps form one of the subunit interfaces. The substitution of residue Q951R lie in domain 5, which is a part of the dimer interface. Other attempts have confirmed that the conversion of natural “specialists” into generalists requires relatively few mutations (Matsumura and Ellington, 2001). Those mutations partially disassemble protein architectures optimized for a particular function.

In this report, we obtained some E. coli β-galactosidase mutants with improved β-glucuronidase activity using a high efficiency and high throughput system in the in vitro directed evolution of a reporter gene. Because human β-glucuronidase and E. coli β-galactosidase protein structures have been studied in great detail, this research may aid in future studies regarding the structure-function relationships of those enzymes. Five of the mutants exhibited more β-glucuronidase activity than wild-type β-galactosidase were obtained. There were ten mutated amino acid residues S193N, T266A, Q267R, V411A, D448G, G466A, L527I, M543I, Q626R and Q951R. After identifying several unique or common mutation sites among selected mutants, we will validate the functional consequences of these mutations by introducing them individually or as combinations. In order to identify key points of mutant proteins, site-directed mutagenesis property studies on these amino acid change sites will be performed (Peng et al., 2006; Xiong et al., 2004a,b; 2006).

**Acknowledgments**

This research was supported by the Shanghai Rising-Star Program (05QMX1445); Program for Science and Technology Development of China Tobacco (1102001011007); Shanghai Natural Science Foundation (04ZR14116); Shanghai Project for International Scientific and Technological Cooperation (05540708); Fund of National Natural Science Foundation of China (30471258; 30570987; 30670179); Program of Shanghai Subject Chief Scientist (06XD14017); Shanghai Key Basic Research Project (05DJ14008); Project of key laboratory of Shanghai (05dz223266; 03dz19311); National high technology research and development program of China (863 Program No. 2006AA10Z117 and 2006AA06Z358) and Development Foundation of SAAS (2006-10). We thank Prof. Chen for critical review of the manuscript.

**References**


