Pak. J. Biochem. Mol. Biol. 2010; 43(3):153-158

5′-end coding sequence of β-glucosidase gene of *Bacillus halodurans* determines the rate of expression and folding of the product in *E. coli*

*Shumaila Naz[1], Nadia Ikram[1], Saima Sadaf[1,2] and M. Waheed Akhtar[1]**
[1]*School of Biological Sciences, University of the Punjab, Lahore, Pakistan*
[2]*Institute of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan*

Abstract: A putative β-glucosidase gene (*bglA*), encoding 447 amino acids with a calculated mass of 51.6 kDa, was reported in the complete genome sequence of *Bacillus halodurans* C-125. This gene sequence was PCR amplified to produce construct pBglA-1. *E. coli* BL21 (DE3) CodonPlus cells transformed with this construct produced bglA at levels >35 % of the total *E. coli* cellular proteins, when induced with 0.6 mM IPTG. Different constructs produced with silent mutations in +2 and +3 codons of the gene generally showed similar levels of expression. Approximately 70 % of the expressed enzyme was obtained in the soluble form in *E. coli* cells transformed with the constructs pBglA-1 to -5, and pBglA-7. However, the construct pBglA-6, which had serine TCA at +3 position instead of native TCG produced this enzyme in an insoluble form. Expression analysis in the early stages after IPTG induction showed that synthesis of bglA was more rapid in the cells transformed with pBglA-6 as compared to other transformants leaving the expressed enzyme with little time to fold and assume native conformation. By lowering the cultivation temperature to 18°C the expression in soluble form, however, could be achieved. This study highlights the significance of synonymous substitutions in the 5′-end coding sequence of *bglA* gene in determining the rate of expression and folding of the expressed enzyme in *E. coli*.

Keywords: Expression and folding, β-glucosidase, *Bacillus halodurans, E. coli.*
Received: March 21, 2010 **Accepted:** July 28, 2010
***Author for Correspondence:** mwapu@brain.net.pk

INTRODUCTION

β-glucosidase (β-D-glucoside glucohydrolase, EC 3.2.1.21) constitutes a major group of enzymes amongst the glucoside hydrolases. They occur ubiquitously in all three domains (archea, eubacteria and eukaryotes) of living organisms and play pivotal roles in many biological processes, such as the biological conversion of cellulosic material[1]. Cellulase systems lacking β-glucosidase or having low levels of this enzyme display poor saccharification power and are prone to end product inhibition. Considerable work on the form, functions and kinetics of β-glucosidases has been reported[2]. The study of these enzymes has been facilitated by the use of recombinant DNA technology[3,4]. A large number of cellulase genes including several *Bacillus* β-glucosidases have been cloned and expressed in *Escherichia coli*[5-9]. But to our knowledge there has so far been no study on expression and/or characterization of β-glucosidase gene from *Bacillus halodurans*.

B. halodurans C-125 is a facultative alkalophilic bacterium, which grows well at pH 7 to 10.5. *B. halodurans* C-125 is similar to *B. subtilis* in terms of genome size, G+C content of the genomic DNA and the physiological properties used for taxonomical identification, except for the alkaliphilic phenotype property[10]. The chromosomal DNA of *B. halodurans* C-125 has been sequenced, revealing that the genome contains 4066 protein coding sequences (CDSs). In the CDSs, several glucosidase genes were found, and one of these, designated as *BH1923* is present at position 2013203 to 2014546[11]. This gene encodes 447 amino acids, and its calculated molecular mass is 51.6 kDa.

E. coli has been the choice host for the expression of many heterologous proteins of commercial significance due to its rapid growth in inexpensive cultivation media, well characterized genetics and range of commercially available compatible vectors. Different strategies have been described to achieve high-level expression of heterologous proteins in this host, which include introduction of synonymous substitutions in the 5'-end of coding sequence. These silent mutations allow the usage of preferred codons and in some cases can reduced the stability of the mRNA secondary structures to ensure high-level expression of protein under investigation[12]. Grantham *et al*. reported that variations in codon usage correlate with variations in respective tRNA abundance and that this might "modulate" gene expression[13,14]. This study reports high-level expression of β-glucosidase gene from *B. halodurans* in *E. coli.* Also the effect of codon alterations by introducing silent mutation in the 5'-end of the gene on expression and folding of the recombinant β-glucosidase in *E. coli* are described.

MATERIALS AND METHODS

Bacterial strains, culture conditions and plasmids

B. halodurans C-125 was grown in shake flask culture taking 50 ml tryptic soy broth (Sigma Aldrich, USA) in 250 ml flask and incubating in an orbital incubator shaker (Irmeco GmbH, Germany) at 150 rpm and 37°C. Chromosomal DNA was prepared from the cells harvested during the growth

phase according to the previously described method (15). pET-22b(+) was used as expression vector and *E. coli* strains DH5α and BL21 (DE3) CodonPlus (Stratagene, USA) were used as cloning and expression hosts, respectively. *E. coli* was grown in LB medium supplemented with ampicillin (100 µg/ml) in an orbital incubator shaker at 37°C and 150 rpm, unless stated otherwise.

PCR amplification and cloning

Primers were designed based on the previously determined DNA sequence of β-glucosidase (bglA) from *Bacillus halodurans* (GenBank Accession BAB05642). Genomic DNA of *Bacillus halodurans* was used as template for PCR amplification of the *bglA* using forward (5'-GGC**CATATG**GTGTCGATCATTCAATTTCCGA AAG-3') and reverse (5'-GA**GTCGAC**TTAAAGTTCAAAGAACTGATTG GCAATC-3′) primer pair. These contained *Nde*I and *Sal*I sites, respectively (as shown in bold) to facilitate directional cloning.

After 30 cycles of amplification (denaturation at 94°C for 45 seconds, annealing at 60°C for 30 seconds, extension at 72°C for 1 minute and 50 seconds, and a final extension at 72°C for 20 minutes), the product was gel purified using QIAquick gel extraction kit (QIAGEN Inc., CA, USA). The amplicon was T/A cloned in pTZ57R/T vector and then subcloned between *Nde*I and *Sal*I sites of pET-22b(+) to generate pBglA-1.

The recombinant plasmid was thereafter used to transform *E. coli* BL21 (DE3) CodonPlus. The positive transformants were selected on LB ampicillin plates and confirmed by colony PCR and restriction analysis of the purified recombinant plasmid.

Silent mutations in 5'-end of the gene

In order to study the codon effect in the early coding sequence of the *bgl*A gene, silent mutations were introduced in +2 and +3 codons using the mutagenic forward primers, as shown in Table 1. The recombinant constructs produced with these silent mutations were correspondingly named as pBglA-2 to -7. The sequence of each of the construct, thus obtained, was confirmed by sequencing with Beckman Coulter CEQ 8000 Genetic Analyzer.

Expression in E. coli

E. coli BL21 (DE3) CodonPlus was transformed with each of the construct and cultivated in 50 ml LB medium containing 100 µg/ml ampicillin at 37°C with shaking in orbital shaker (Shaking Incubator Irmeco GmbH, Germany) until OD_{600} reached ~0.8. After induction with 0.6 mM IPTG for varying periods, the cells were harvested by centrifugation at 4,500 rpm for 10 minutes.

The pellets were resuspended in 50 mM glycine-NaOH buffer (pH 8.0) by adjusting volumes to obtain the same cell density (OD_{600}=50) in each case. After sonication (UP 400s dr. hielscher GmbH, Germany) by giving 10 x 30 seconds bursts with intervals of one minute between the two pulses, the cell lysate was centrifuged at 4,500 rpm for 15 minutes to obtain soluble and insoluble protein fractions. The insoluble fraction was resuspended in glycine-NaOH buffer for further analysis.

Table 1: Forward primers used for incorporating different silent mutations (shown in bold) BglA.

Oligo Name	Sequence (5′- 3′) +1 +2 +3 +4 +5 +6 +7 +8 +9
F-01	GGC CAT ATG GTG TCG ATC ATT CAA TTT CCG AAA G
F-02	GGC CAT ATG GT**T** TCG ATC ATT CAA TTT CCG AAA G
F-03	GGC CAT ATG GT**C** TCG ATC ATT CAA TTT CCG AAA G
F-04	GGC CAT ATG GT**A** TCG ATC ATT CAA TTT CCG AAA G
F-05	GGC CAT ATG GTG TC**T** ATC ATT CAA TTT CCG AAA G
F-06	GGC CAT ATG GTG TC**A** ATC ATT CAA TTT CCG AAA G
F-07	GGC CAT ATG GTG TC**C** ATC ATT CAA TTT CCG AAA G

SDS-PAGE

The total cellular proteins of transformed *E. coli* expressing bglA, and the soluble and insoluble fractions of cell lysates were analyzed by SDS-PAGE using 12% polyacrylamide gel[16]. Following electrophoresis, the gel was Coomassie stained and percentage of bglA in the protein samples was determined by densitometric scanning of the gel (GeneTools, Syngene, UK).

Protein amounts were determined in the soluble and the insoluble fraction by taking OD_{280}. The cell pellet was resuspeded in the same volume of buffer as that of the resuspended cells. 10 µl of this suspension was mixed with 990 µl 0.5% SDS solution before taking OD_{280}.

β-glucosidase assay

β-glucosidase assay was done by incubating an aliquot of the enzyme sample, suitably diluted with 50 mM glycine-NaOH (pH 8.0), with 10 mM cellobiose in a total volume of 1 ml, and incubating the mixture at 45°C for 20 minutes.

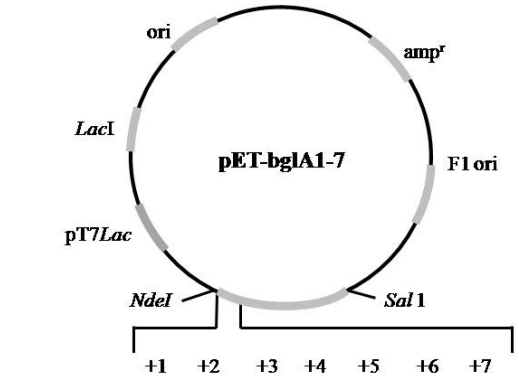
Glucose thus released was estimated by glucose oxidase method[17]. One unit of bglA activity is defined as the amount of enzyme that produces 1 nmol of glucose/per minute under the assay conditions.

RESULTS AND DISCUSSION

Construction of expression plasmids

The recombinant plasmids were produced by inserting the amplified native fragment and those

with silent mutations at +2 and +3 positions between *Nde*I and *Sal*I sites of pET-22b(+) plasmid (Figure 1). These were named as pBglA-1 to -7. Correctness of the sequence of each of the insert in the plasmid was confirmed by sequencing using DNA Genetic Analysis System (Beckman Coulter CEQ 8000 Genetic Analyzer). Free energies of any possible secondary structures predicted by the *mfold* webserver software[18] were calculated for the fragment including the bases from the ribosomal binding site to +21 bases of the coding sequence in each of the constructs, which are shown in Table 2.



	+1	+2	+3	+4	+5	+6	+7
pET-bglA −1	ATG	GTG	TCG	ATC	ATT	CAA	TTT
−2	ATG	GT**T**	TCG	ATC	ATT	CAA	TTT
−3	ATG	GT**C**	TCG	ATC	ATT	CAA	TTT
−4	ATG	GT**A**	TCG	ATC	ATT	CAA	TTT
−5	ATG	GTG	TC**T**	ATC	ATT	CAA	TTT
−6	ATG	GTG	TC**A**	ATC	ATT	CAA	TTT
−7	ATG	GTG	TC**C**	ATC	ATT	CAA	TTT

Figure 1: Constructs of pET-22b(+) and *bglA* gene with silent mutations in +2 and +3 codons. The mutations are shown in bold.

Expression of bglA gene

The recombinant plasmid pBglA-1, which harbored the *bglA* gene insert with the native sequence, was used to transform *E. coli* expression host. The culture in shake flask on induction with 0.6 mM IPTG expressed bglA rapidly reaching to the level of ~36 % of the total cell proteins within 4 hours post-induction (Figure 2). The molecular mass of the expressed enzyme, as determined from SDS-PAGE analysis, was ~51 kDa.

Effect of mutations on expression

E. coli cells were transformed with pBglA-2 to -7, which contained mutation at +2 and +3 codons. IPTG induction of all the freshly grown cultures resulted in the production of bglA in high percentages as shown in Figure 3a. The percentage of bglA in the total cell proteins of the host cell ranged from 36-38 % in various transformants (Table 2). The levels of bglA expression as percentage of the total *E. coli* cellular proteins were similar in various constructs, although the ΔG° values for the fragment between the ribosomal binding site and the first 21 bases of the coding sequence (shown in Table 2), varied significantly.

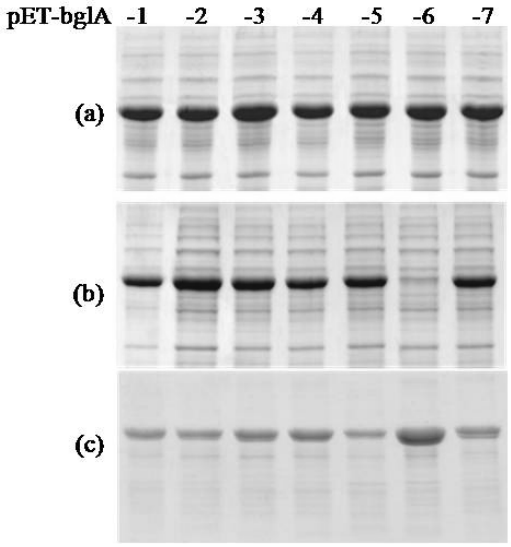


Figure 2: SDS-PAGE of total (a), soluble (b) and insoluble (c) cell protein of *E. coli* BL21 (DE3) CodonPLus cells transformed with pBglA 1-7 at 6 hrs. post -induction with 0.6 mM IPTG at 37°C.

The ΔG° value for the construct pBglA-2 was lowest i.e., -1.9, but the % age expression was almost similar to other constructs and accordingly the activity was marginally higher as compared to pBglA-1, -3 and -5. Activities in the cells transformed with pBglA-1 to -5, and -7 ranged between 135 to 165 U/ml of the cell lysate, which seemed to correspond with the proportion of the enzyme expressed in the soluble form (Table 2, Figure 2a-2c). Here, the expression does not seem to be affected, in general, by any variations in the possibility of the secondary structure formation in the early coding sequence of the mRNAs. Based on the contents of charged (Asp, Glu, Lys and Arg) and the turn forming (Asn, Gly, Pro and Ser) amino acid residues, a model for predicting the solubility of recombinant proteins in *E. coli* at 37°C has been proposed[19]. Using this model, the chance of bglA expression in the soluble form was found to be 58.9%. As shown in Table 2, the percentage expression of bglA in the soluble form, for the constructs mentioned above, was in the range of 68-75 % of the total expressed enzyme. This reflects that although high level expression of native and mutant forms of bglA in *E. coli* leads to the formation of some inclusion bodies (~ 25%) but the major proportion (upto 75%) of expressed enzyme may be recovered from the soluble fraction.

Table 2: Expression levels of β-glucosidase as percentages of total cell protein (TCP) and soluble protein (SP), in *E. coli* cells transformed with pET-bgl-1 to -7, at 6 hours after IPTG induction.

pBglA	ΔG° kcal/mole	Cell Protein (mg/ml)	Enzyme activity (U/ml)	β-glucosidase content				
				TCP		SP		
				%age	mg/ml	%age	mg/ml	%age solubility
-1	-3.6	12.3	148	36	4.42	37	3.25	73
-2	-1.9	12.7	165	36	4.57	40	3.44	75
-3	-4.2	12.4	150	38	4.72	42	3.40	72
-4	-3.1	12.6	146	36	4.48	36	3.21	71
-5	-5.6	12.0	138	36	4.20	38	3.01	72
-6	-3.1	12.3	08	38	4.42	03	0.37	08
-7	-3.6	12.4	135	36	4.46	38	3.05	68

Table 3: Distribution of total cell proteins (TCP), soluble (SP) and insoluble proteins (IP) of lysed *E. coli* cells transformed with pBglA-5 and -6 grown for 0.5-2.0 hours after induction with IPTG at 37°C.

Transformants		pBglA-5					pBglA-6				
Post-induction stage (hrs.)		0.5	1.0	1.5	2.0	4.0	0.5	1.0	1.5	2.0	4.0
Whole cell	bglA (% age of TCP)	20	24	38	40	40	18	40	38	38	38
	bglA (mg/ml/50 OD_{600})	2.6	3.1	5.0	5.2	5.2	2.3	5.2	4.9	4.9	4.9
Cell lysate supernatant	SP (mg/ml/50 OD_{600})	12.2	11.8	11.0	10.7	10.7	12.0	11.7	10.3	7.8	7.8
	bglA (% age of SP)	16	23	35	37	37	15	35	22	06	04
	bglA (mg/ml/50 OD_{600})	1.9	2.7	3.8	3.9	3.9	1.8	4.1	2.3	0.4	0.3
Cell lysate pellet	IP (mg/ml/50 OD_{600})	0.5	0.7	1.3	1.5	1.5	0.5	0.9	2.4	4.5	4.5
	bglA (% age of IP)	37	46	78	80	80	37	70	83	92	93
	bglA (mg/ml/50 OD_{600})	0.2	0.3	1.0	1.2	1.2	0.2	0.6	2.0	4.1	4.2
β-glucosidase activity (U/ml)		62	90	130	135	135	55	127	70	12	08

Table 4: Distribution of β-glucosidase activities in the total cell proteins (TCP) soluble proteins (SP) of the lysed *E. coli* cells transformed with pBglA-1,-5, and -6 grown at 25°C for 6 hours, and at 18° C for 12 hours, respectively.

Growth Temperature	pBglA	Cell protein mg/ml	Enzyme activity (U/ml)	β-glucosidase content			
				TCP		SP	
				% age	mg/ml	%age	mg/ml
25°C	-1	14.2	110	47	6.70	35	4.60
	-5	14.1	118	51	7.10	34	4.40
	-6	14.5	82	48	7.00	37	4.80
18°C	-1	12.6	101	50	6.29	47	6.20
	-5	12.6	99	50	6.30	47	6.23
	-6	12.5	91	50	6.25	44	6.10

The construct pBglA-6 having TCA at +3 position, is the only exception. When grown at 37°C, most of the enzyme expressed in *E. coli* cells harboring this construct was in the form of inclusion bodies, and accordingly the enzyme activity was very low (Figure 2).

In order to look for a possible explanation for bglA insolubilization, the expression pattern of bglA in cells transformed with pBglA-6 and pBglA-5 (having TCA at +3 position) were compared, during the early stages of IPTG induction. Within initial one hour of induction, around 80-85 % of the expressed bglA could be found in the soluble fraction of the total cell lysate, in both the casas (Table 3). Soon thereafter, with further synthesis, the enzyme was found increasingly in the insoluble fraction particularly in the case of pBglA-6, where only 47 % bglA could be obtained in soluble/active form after 1.5 hours of induction.

This percentage further dropped to <10 % with almost complete loss of activity after 2 hours. pBglA-5, in contrast, maintained the solubility and activity of bglA at 75 % and 130 U/ml, respectively, even after 4 hours of induction (Figure 3, Table 3).

Several studies have documented that individual codons translate at different rates and the limiting factor which dictates the elongation step, is infact the 'wait time' for an appropriate aminoacyl-tRNA (20 and references therein). More readily the tRNA anticodon available for the codon, more rapid would be the translation. So, the codon TCA for serine at +3 position in pBglA-6 seems to cause too rapid translation of the enzyme most probably due to the readily available tRNA. This leaves the

expressed enzyme with little time to fold and assume native conformation. The codons TCC, TCT or TCG at similar position, in contrast, display relatively slower rates of translation allowing a major proportion of the expressed product to fold properly and thus remain mainly in the soluble fraction even after longer hours of induction. In the light of these observations, it may be hypothesized that the codons in the early coding sequence of *bglA* gene a role in determining the distribution of the expressed enzyme between the soluble and insoluble fractions.

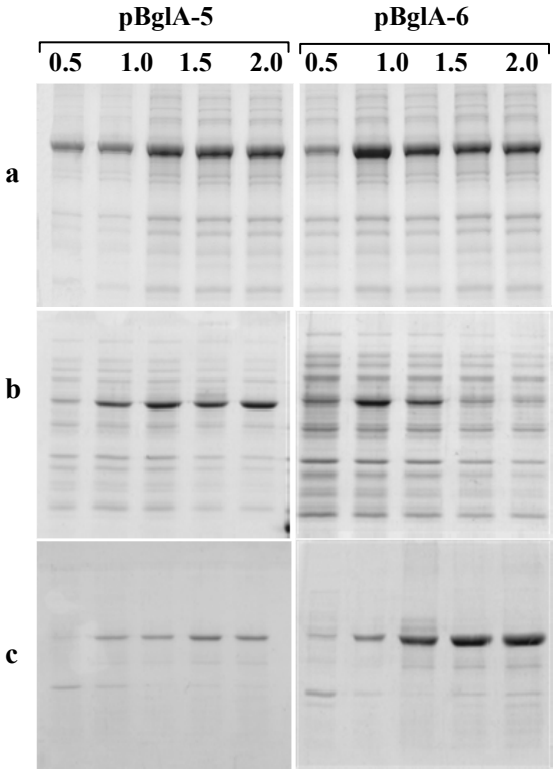


Figure 3: SDS-PAGE of total (a), soluble (b) and insoluble (c) bglA expressed in *E. coli* cells transformed with pBglA -5 and -6 at 0.5, 1.0, 1.5, 2.0 and 4.0 hours post –induction at 37°C.

Expression at lower temperature

The expression of bglA in the transformants of pBglA-1, -5, and -6 was also studied at 18^{o} and 25°C after IPTG induction, to see if the slower growth rate associated with decreased rate of protein synthesis has an effect on folding of the enzyme.

The relative amounts of bglA in the total cell proteins, and the soluble and insoluble fractions of the cell lysates of these constructs are shown in Figure 4. The proportion of bglA in the soluble form in the cells transformed with pBglA-6, showed significant increase when cultivated at 25^{o}C and this corresponded with increase in the enzyme activity. By lowering the growth temperature to 18^{o}C, nearly all the expressed enzyme could be obtained in the soluble form. Accordingly, the bglA activity increased from almost nil (when grown at 37^{o}C) to 82 and 91 U/ml of the cell lysate, when cultivated at 25^{o} and 18^{o}C, respectively (Table 4). The observation suggested that induction at 25-18°C slows down the protein synthesis which resultantly helps in preventing the aggregation of expressed protein in the form of inclusion bodies.

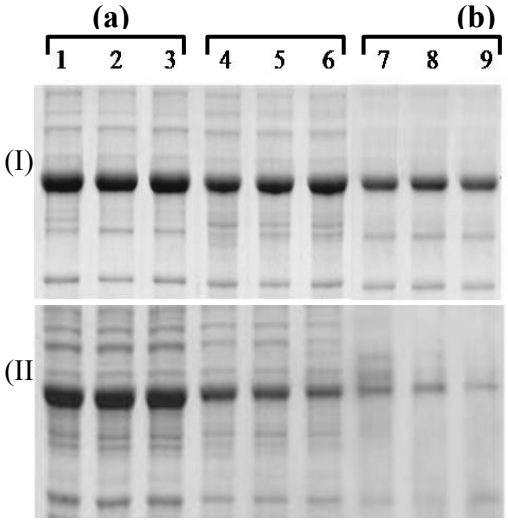


Figure 4: Total (a), soluble (b) and insoluble (c) proteins in *E. coli* cells transformed with pBglA-1 (lanes 1, 4, and 7), pBglA-5 (lanes 2, 5, and 8), pBglA-6 (lanes 3, 6, and 9), when cultivated at 25^{o} C (I) and 18^{o} C (II).

The results of this study suggest that translation efficiency is strongly influenced by codon usage in the region immediately following the start codon. Thus, choosing right codon in the sequence of interest at +3 position and changing it to that which is more often used in *E. coli* could improve expression and solubility levels. The data shows that the level of bglA expression does not correlate with the predicted free energy values as determined by *mfold* webserver. Instead of free energy values it is the sequence of nucleotides at specific points and use of preferred codons, which is important in determining the expression level. Codon like TCT and TCC for serine are the rate-limiting codons for the expression of bglA in *E. coli,* thus providing sufficient time for the expressed protein to fold into an active state. Slower rate of expression at lower temperatures would also allow the protein to fold into a native state.

ACKNOWLEDGEMENT

This work was done under a research grant from the Higher Education Commission, Government of Pakistan.

REFERENCES

1. Woodward J and Wiseman A. Fungal and other β-D-glucosidases, their properties and applications. *Enzyme Microb. Technol.*, 1982; 4: 73-93.
2. Ogunseitan OA. Biotechnology and industrial ecology: New challenges for a changing global environment. *A. J. Biotechnol.*, 2003; 2: 596-601.
3. Beguin P. Molecular biology of cellulose degradation. *Ann. Rev. Microbiol*., 1990; 44: 219-248.
4. Glick BR and Pasternak JJ. Isolation, characterization and manipulation of cellulase genes. *Biotechnol. Adv.*, 1989; 7: 361-386.
5. Love DR and Streiff MB. Molecular cloning of β-Glucosidase gene from extremely thermophilic anaerobe in *E. coli* and *B. subtilis*. *Biotechnol*., 1987; 5: 384-387.
6. Wakarchuk WW, Greenberg NM, Kilburn DG, Miller RC and Warren RAJ.: Structure and transcription analysis of the genome encoding a cellobiase from *Agrobacterium*. *J. Bacteriol*., 1988; 170: 301-307.
7. Painbeni E, Valles S, Polaina J and Flors A. Purification and characterization of a *Bacillus polymyxa* β-glucosidase expressed in *Escherichia coli*. *J. Bacteriol.*, 1992; 174: 3087-3091.
8. Paavilainen S, Hellman J and Korpela T. Purification, characterization, gene cloning, and sequencing of a new β-glucosidase from *Bacillus circulans* subsp. alkalophilus. *Appl. Environ. Microbiol*., 1993; 59: 927-932.
9. Papalazaridou A, Charitidou L and Sivropoulou A. β-Glucosidase enzymatic activity of crystal polypeptide of the *Bacillus thuringiensis* strain 1.1. *J. Endotoxin Res*., 2003; 9: 215-224.
10. Takami H and Horikoshi K. Reidentification of facultatively alkaliphilic *Bacillus* sp. C-125 to *Bacillus halodurans*. *Biosci. Biotech. Biochem.*, 1999; 63: 943-945.
11. Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S and Horikoshi K. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acid Res*., 2000; 28: 4317-4331.
12. Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev*., 1996; 60: 512-538.
13. Grantham R, Gautier C and Gouy M. Codon frequencies in 119 genes confirm consistent choices of degenerate base according to genome type. *Nucleic Acid Res*., 1980; 8: 1892-1912.
14. Grantham R, Gautier C, Gouy M, Jacobzone M and Mercier R. Codon catalogue usage is a genome strategy for genome expressivity. *Nucleic Acid Res.*, 1981; 9: 43-75.
15. Rodriguez RL and Tait RC. Recombinant DNA techniques: An introduction, Wesley Publishing Co., 1983; MA, USA.
16. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 1970; 277: 680-685.
17. Bergemeyer H, Bernt E and Bergemeyer H. Methods of enzymatic analysis, Academic Press, NY, USA.
18. Zuker M. *Mfold* web server for nucleic acid folding and hybridization prediction. *Nucleic Acid Res.*, 2003; 31: 3406-3415.
19. Wilkinson DL and Harrison RG. Predicting the solubility of recombinant proteins in *E.coli. Biotechnol*., 1991; 9: 443-448.
20. Sinclair G and Choy FYM. Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotropic yeast, *Pichia pastoris*. *Protein Exp. Purif.*, 2002; 26: 96-105.