Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 336 (2005) 299–308
BBRC
www.elsevier.com/locate/ybbrc

Cloning, sequence analysis, and characterization of a novel β-glucosidase-like activity from *Pichia etchellsii*

Pranita Roy, Saroj Mishra *, Tapan K. Chaudhuri

Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, Delhi, Hauz-Khas, New Delhi 110016, India

Received 1 July 2005
Available online 19 August 2005

Abstract

Genomic DNA fragment encoding a novel β-glucosidase-like activity of the yeast *Pichia etchellsii* was cloned and expressed in *Escherichia coli*. An open-reading frame of 1515 bp, termed *mug*A, coding for a protein of predicted molecular mass of approximately 54 kDa was confirmed for this activity. The sequence of the deduced protein did not show homology with the generic β-glucosidases but a high degree of identity was seen with several Ser-Asp (SD)-rich cell-surface-associated proteins. The secondary structure prediction program 3D-PSSM indicated the protein to be composed of largely helical and coiled structures, which was confirmed by circular dichroism spectroscopy. The encoded protein, MUGA, was purified by about 53-fold and characterized as a monomer of 52.1 kDa by SDS–PAGE and MALDI-TOF. The protein displayed high hydrolytic activity on methylumbelliferyl β-D-glucoside but relatively very little hydrolysis of *p*-nitrophenyl β-D-glucoside and gentiobiose, characteristic substrates for β-glucosidases. The binding experiments performed between *P. etchellsii* cells and the purified *E. coli* expressed MUGA indicated binding with the cell surface, which was monitored by fluorescence microscopy. In competition experiments with the SD dipeptide, less protein was shown to bind to the cell surface, in a concentration-dependent manner.

Keywords: β-Glucosidase-like protein; MUG hydrolysis; CD spectroscopy; Cell-surface-associated protein; Fluorescence microscopy

β-Glucosidases (β-D-glucoside glucohydrolase, EC 3.2.1.21) are able to cleave the β-glucosidic linkages in di- and oligoglucosaccharides, and several other glycoconjugates. These enzymes are widely distributed in the living world and play pivotal roles in many biological processes. The physiological roles associated with this enzyme are diverse and depend on the location of the enzyme and the biological system in which these occur. In cellulolytic microorganisms, β-glucosidase is involved in cellulase induction and cellulose hydrolysis [1,2]. In plants, this enzyme is involved in β-glucan synthesis during cell wall development, pigment metabolism, fruit ripening, and defense mechanisms [3,4]. In humans and other mammals, the enzyme is involved in the hydrolysis of glucosyl ceramides [5] and the deficiency of the enzyme leads to Gaucher's disease. Like many hydrolases, these enzymes can be used for synthesizing a variety of glycoconjugates such as alkyl glucosides, aminoglycosides, and special disaccharide fragments of phytoalexin—elicitor oligosaccharides which are involved in plant and other microbial defense mechanisms (For review, see [6]).

A number of bacterial, yeast, and fungal β-glucosidase genes have been cloned. At the sequence level, they are placed in family 1 and family 3 of glycoside hydrolase families [7], which are 99 (last update, May 19, 2005) in number. The sequence based classification is continuously updated (URL: http://afmb.cnrs-mrs.fr/~pedro/CAZY) and is useful in characterizing the enzymes from the structural point of view but the substrate specificity with respect to the aglycone moiety still serves a primary, or, in some cases, the only lead in isolating and characterizing unknown or structurally undefined glucosidases. In general,

* Corresponding author. Fax: +91 11 2658 2282.
E-mail addresses: saroj@dbeb.iitd.ernet.in, saroj98@hotmail.com (S. Mishra).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.08.067

β-glucosidases hydrolyze *p*-nitrophenyl β-D-glucoside (*p*NPG) and gentiobiose, the former being the substrate of choice for routine measurement of enzyme activity. It has also been used as a substrate in the chromogenic screening of recombinant colonies expressing this enzyme. While many β-glucosidases are reported to be bound to the cell wall or involved in cell-surface-associated phenomena, the sequenced genes have not indicated any structural features normally associated with the cell-surface-bound proteins. A number of enzymes have been crystallized, particularly from family 1, and the structures are typical $(\beta/\alpha)_8$ barrel structures. The structure of a barley β-D-glucan exohydrolase (Exo I) of *Hordeum vulgare* [8], an enzyme related to family 3 enzymes based on sequence similarity, is known and also contains characteristic $(\beta/\alpha)_8$ barrel in one of the two domains. The other domain is an α/β sandwich structure.

In our search for enzymes that are thermo-tolerant and have glycosyltransferase activities that can be used for the enzymatic synthesis of various glycoconjugates, we have reported on the cloning and expression of two β-glucosidases, *Bgl*I [9] and *Bgl*II [10], from *Pichia etchellsii* into *Escherichia coli*. Both the enzymes have been well characterized [11,12] and *Bgl*II has been extensively used for synthesis of oligosaccharides [13], alkyl and terpene glucosides [14], and asparagine mimetics [15]. Two additional enzymes, BGLI and BGLII [16], have been purified from the cell wall of the yeast and their detailed properties described. All the four enzymes efficiently hydrolyzed methylumbelliferyl β-D-glucoside (MUG) and *p*NPG. In this paper, we describe, for the first time, isolation and properties of a novel enzymatic activity that resembled β-glucosidases in terms of its ability to hydrolyze MUG but displayed relatively very little hydrolytic activity on *p*NPG and had no sequence similarity with the generic β-glucosidases. Instead, sequence identity was seen with several members of the Ser-Asp repeat protein family which are cell-surface-associated. The purified protein was predicted to contain helical/coiled regions, which was confirmed by circular dichroism spectroscopy. Binding experiments were performed which indicated strong binding between the yeast cells and the purified protein. Investigations of this protein might further explain the role of β-glucosidases in cell-surface-associated phenomena.

Materials and methods

Strains, plasmids, and culture conditions. Pichia etchellsii (Deutsche Sammelung für Mikroorganismen, Germany) was used as the donor of β-glucosidase gene. *E. coli* DH5α cells were used as the host and pUC19 as cloning vector. *P. etchellsii* was maintained on YPD (0.5% yeast extract, 1% bacto-peptone, and 2% D-glucose) medium. *E. coli* was grown in LB at 37 °C according to standard protocols [17]. For *E. coli* transformants, ampicillin (amp) was added at 50 μg/ml to LB medium.

Cloning of β-glucosidase gene. Chromosomal DNA from *P. etchellsii* was prepared using zymolyase according to Pandey and Mishra [9]. The pUC19 plasmid was made by standard alkaline lysis method [17]. Partially *Sau*3AI digested chromosomal DNA of 5–20 kbp size was ligated to dephosphorylated pUC19 at the *Bam*HI site. The recombinant plasmids were transformed into competent *E. coli* DH5α cells and the cells were plated on LB containing amp, X-Gal, and IPTG. The white colonies were picked and further plated on LB + amp plates. Screening for β-glucosidase activity was done by overlaying the overnight grown colonies with 1 mM MUG (Sigma) and incubating the plates at 37 °C. Hydrolysis of MUG by the cloned β-glucosidase resulted in the release of 4-methylumbelliferrone (or methylumbelliferone) making the positive clones fluorescent under UV. MUG positive clones were tested further on *p*NPG plates. For this, the positive clones were streaked on LB + amp containing filter-sterilized *p*NPG at 4 mM. The colonies were incubated at 37 °C for 16–18 h and checked for the appearance of a yellow zone (*p*-nitrophenol liberation) around the clones. A positive and a negative control were included in all plate assays.

Sequence analysis and delineation of the correct ORF. Sequencing reactions on the plasmid pMG8 were carried out on Applied Biosystems automatic sequencing unit at Microsynth, Switzerland. The second strand sequencing was carried out at the DNA sequencing facility of the Department of Biochemistry, University of Delhi, South Campus. ORF search and sequence analysis were done using DNASTAR program. The correct reading frame was delineated by sub-cloning experiments. Based on identification of an overly expressed new protein of approximate 50 kDa molecular mass in the pMG8:DH5α clone, the possible initiating codon was identified, which was flanked by a unique *Pst*I site (shown in bold in Fig. 1) at the 5′end. The plasmid pMG8 was digested with *Pst*I and the ~2.85 kbp DNA fragment (containing the putative reading frame), of the three fragments generated by *Pst*I digest, was gel-eluted and ligated to *Pst*I digested/CIP treated pUC19 vector. The ligation mixture was used for transforming competent DH5α cells following the standard protocol [17]. The transformants were screened for MUG hydrolyzing activity as described above. The reading frame was further confirmed by digesting the plasmid pMG8 with *Eco*RI, which cut the putative ORF intragenically at a single site. The fragment containing the entire plasmid backbone and the 3′end (~50%) of the gene was self-ligated and transformed into competent DH5α cells. The plasmids were analyzed for the correct DNA changes and screened for MUG hydrolyzing activity.

Confirmation of origin of mugA by the PCR method. The origin of *mugA* insert in pMG8:DH5α was confirmed by PCR experiment. The chromosomal DNA of *P. etchellsii* was used as the template (1 μg) in PCR (100 μl) and the two, forward (5′-CCT CGG AAT CAG ACA CGG-3′) and reverse (5′-CCG TTT GGG CAT TTC CGC-3′) primers (underlined by arrows in Fig. 1), were used. The amplification conditions were as follows: initial denaturation at 94 °C/2 min, followed by 28 cycles of denaturation at 94 °C/30 s, annealing at 45 °C /30 s, and extension at 72 °C/3 min. A step of extension at 72 °C/10 min was carried out at the end. About 10 μl of the sample was run on 0.7% agarose gel along with the λ (*Eco*RI–*Hin*dIII cut) markers.

Sequence analysis and prediction of secondary structure. The putative sequence of the delineated protein, termed MUGA, was subjected to homology search using the National Centre for Biotechnological Information (NCBI) on-line program BLAST against protein (BlastX) and nucleotide (BlastN) sequences stored in GenBank. β-Glucosidase sequences of family 1 and family 3 were downloaded (http://afmb.cnrs-mrs.fr/~pedro/CAZY) and a sequence comparison was performed using pairwise and multiple sequence alignment program CLUSTAL W [18]. The sequence was analyzed for secondary structural elements by the 3D-PSSM program [19] available online. The sequence was also analyzed by the 'DAS' TM-segment prediction program for prediction of trans-membrane proteins (URL: http://www.sbc.su.se/~miklos/DAS/).

Purification of MUGA protein. For purification of intracellularly localized MUGA, recombinant E. coli cells were grown overnight at 37 °C in 1-L LB containing 50 μg/ml amp with vigorous shaking. Harvested cells were suspended in 1/10th of the original culture volume in Tris–Cl buffer, pH 8.3, and subjected to sonication (Soniprep) for 30 min at maximum output. The clear cell-free extract obtained after centrifugation (10,000*g*, 20 min) was subjected to differential ammonium sulfate precipitation between 30% and 80%, and the precipitate was dissolved in minimal volume of 20 mM Tris–Cl buffer, pH 8.3. Dialyzed and concentrated enzyme was loaded on a column (1 × 14 cm) containing DEAE–Sepharose equilibrated

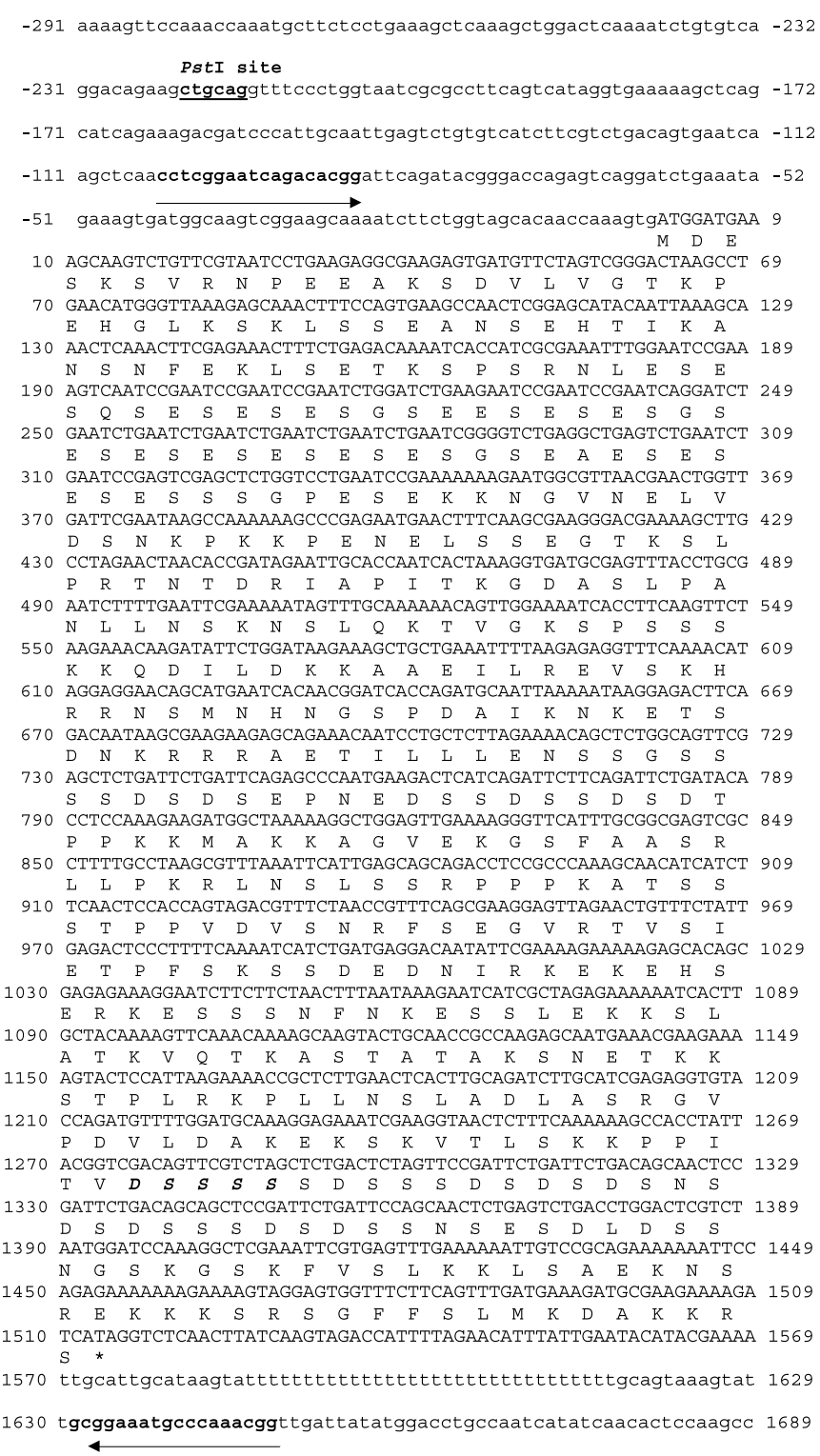

Fig. 1. Nucleotide sequence of *mug*A gene from *P. etchellsii*. Predicted amino acid sequences from the first ATG are given using the one-letter code. The oligonucleotides used for PCRs with the chromosomal DNA are indicated by arrows. *Pst*I site used in sub-cloning experiment is underlined.

with 20 mM Tris–Cl buffer, pH 8.3. The enzyme was eluted using a linear gradient of 0.1–0.5 mM KCl in the same buffer at a flow rate of 0.5 ml/min, and 1-ml fractions were collected. The fractions were analyzed for MUG hydrolyzing activity and protein concentration (at 280 nm). The active fractions were pooled and loaded on hydroxyapatite column (1 × 14 cm) equilibrated with 50 mM sodium phosphate buffer, pH 6.8.The enzyme was eluted using a linear gradient of 0.05–0.5 M sodium phosphate buffer, pH 6.8, at a flow rate of 0.3 ml/min and 1-ml fractions were collected. The active fractions were pooled and the enzyme was further purified to homogeneity by passage through a final Sephadex G-25 column (5-ml syringe). The homogeneity of the preparation was confirmed by gel electrophoresis [20].

Molecular mass determination. The molecular mass of the purified protein was determined by MALDI-TOF (Shimazdu) spectrometer (Department of Biophysics, All India Institute of Medical Sciences, New Delhi). The subunit molecular mass was determined by SDS–PAGE on 10% separating gel using standard molecular mass markers (medium range 14–97 kDa).The gels were stained with Coomassie blue R-250 and destained according to standard procedures [21].

Determination of pH, temperature optimum, and substrate specificity. The routine assay of activity was performed using 2 mM MUG as described [22]. Absorbance was read at 347 nm after terminating the reaction with 1 M Na_2CO_3. One milliunit corresponded to release of 1 nmol of methylumbelliferrone/min. The pH optimum was determined using MUG (2 mM) as the substrate in the pH range of 2–9. The buffers used were glycine HCl (2.0–3.0), citrate buffer (4.0–6.0), and Tris-Cl buffer (7.0–9.0). The temperature optimum was determined on MUG at temperatures ranging from 30 to 60 °C. Activity on methylumbelliferyl β-D-cellobioside was determined in a similar manner. Aryl substrates, as listed in the relevant table, were taken at 4 mM level as described [16]. One international unit (IU) of enzyme activity corresponded to the release of 1 μmol *p*-nitrophenol/min. Activity on disaccharides, salicin and amygdalin, was determined by measuring the release of glucose using glucose oxidase-peroxidase kit as described previously [16]. One IU corresponded to release of 1 μmol glucose/min under these conditions. For comparison, activity on these substrates was determined with almond β-glucosidase (Sigma). The protein concentration of the almond enzyme was adjusted to give similar activity on MUG as obtained with the MUGA protein.

Protein concentration was determined by the method of Bradford [23] using bovine serum albumin as the standard.

Equilibrium CD spectral analysis. Equilibrium CD spectrum of purified MUGA (0.46 mg/ml) was measured in 10 mM sodium cacodylate buffer, pH 7.0 (the enzyme was extensively dialyzed at 4 °C against this buffer for 48 h with several buffer changes), on a Jasco J-810 spectropolarimeter using an optical cuvette with a path length of 1 cm for measurements in the peptide region. The baseline of the spectrum was corrected with the same buffer. The instrumental parameters selected for the measurement were: sensitivity 100 mdeg, bandwidth 1 nm, response 1 s, scanning speed 50 nm/min, and data pitch 0.2 nm as described previously [24]. Data obtained in millidegrees were converted into molecular ellipticity using the software manager supplied by Jasco. Secondary structure analysis was made according to Yang et al. [25].

Binding of MUGA with yeast cells. The binding experiments were performed as follows: *P. etchellsii* was grown to late logarithmic phase in phosphate-succinate (PS) medium as described previously [16]. The cells were washed once with sodium phosphate buffer and re-suspended in the same buffer at an approximate cell number of 2×10^5 cells/ml. About 300 μl of the cell suspension was taken in an Eppendorf tube and purified protein (50 nmol) was added. The tube was vortexed and incubated at 4 °C for 30–45 min after which the cells were separated by centrifugation at 10,000*g* (10 min, 4 °C). The cells were washed twice with sodium phosphate buffer, pH 7.0, and re-suspended in the same buffer at the original cell concentration. A loopful of cells was removed and mixed with a few drops of MUG placed on a pre-heated (37 °C) glass slide. The slide was incubated further at 37 °C and observed under a fluorescence microscope (Nikon Optiphot, Episcopic-Fluorescence). Photography was performed with an automatic recorder. The yeast cells without any added MUGA were processed in a similar manner and served as the control.

Competition experiments between MUGA and Ser-Asp (SD) dipeptide. Competition experiments were performed under non-saturating concentrations between MUGA and increasing concentrations (0.05–0.5 nM) of the SD dipeptide for binding to the yeast cell surface. The peptide was synthesized at 0.10 mM level and purified by HPLC. The structure was confirmed by MS/MS nanospray and the sequence by the sequencer. The synthesis and analysis of the dipeptide were performed at Protein Chemistry Core Laboratory at Baylor College of Medicine, Houston, Texas, USA. The yeast cells, grown in PS medium, were washed and suspended in the reaction buffer. About 400 μl (containing approx 2×10^5 cells) of cell suspension was taken in an Eppendorf tube and purified MUGA (0.5 nM) was added to the cells. In parallel experiments, different concentrations of the SD dipeptide (0.05, 0.1, and 0.5 nM) were added along with the purified MUGA. The tubes were vortexed and then incubated at 4 °C for 4 h to allow binding to occur. The cells were separated by centrifugation at 10,000*g* (10 min, 4 °C), washed at least three times with the buffer, and, then suspended in 350 μl of reaction buffer. One hundred and fifty microliters of MUG (2 mM), pre-incubated for 10 min at 37 °C, was added in all Eppendorf tubes and the tubes were further incubated for 45 min. The cells were separated by centrifugation and the supernatant was collected for measuring OD_{347} to quantify the amount of methylumbelliferone released. The values were converted to mU/ml of activity.

Results and discussion

In an attempt to study the biological role of β-glucosidases in lower eukaryotes and to exploit these activities for synthesis of glycoconjugates, we have studied a number of these enzymes from the native yeast *P. etchellsii* or after isolation and expression of the genes in *E. coli.* Four such enzymes (*Bgl*I, *Bgl*II, BGLI, and BGLII) have been described so far differing from each other in terms of physical properties and substrate specificities. The enzymes (*Bgl*I, *Bgl*II, and BGLI) have also been used for synthesis of a number of compounds. All the four enzymes hydrolyze both *p*NPG and MUG in liquid assay conditions and on solid media. In this paper, we report on the identification and some properties of a novel MUG hydrolyzing protein that has different properties.

Isolation and characterization of the mugA gene

The gene encoding the novel β-glucosidase-like activity was isolated by screening a genomic DNA library on MUG containing plates. Positive clones were identified by the presence of fluorescence under UV. Two fluorescent clones, namely pMG8:DH5α and pMG16:DH5α, were identified. Based on restriction mapping and PCR amplification of same sized fragments with a given set of primers, these were concluded to have the same insert. The sequencing of the 6.35 kbp insert in the pMG8 plasmid showed multiple ATGs terminating on a single TAG. The subcloning experiment localized the coding region on the *Pst*I segment that contained a single ATG. The putative coding sequence was digested with *Eco*RI to make an intragenic cut and this was shown to result in loss of MUG hydrolytic activity in the corresponding clone. The gene sequence and the predicted amino acid sequence are shown in Fig. 1 and comprise 1515 bp with regularly spaced poly (A) stretches around the 3′end of the gene. Interestingly, a long tail of poly(T) was observed just downstream of the coding sequence. The ORF codes for a polypeptide of 504 amino acids with a predicted molecular mass of approximately 54.1 kDa. The molecular mass determined by MALDI-TOF was 52.1 kDa, which indicated some processing of the protein in *E. coli* cytoplasm. Based on SDS–PAGE, it was concluded to be monomeric.

The origin of the *mug*A insert in pMG8: DH5α from *P. etchellsii* was confirmed by PCR instead of the routine

Southern hybridization. A single fragment of about 1.5 kbp was amplified (Fig. 2) as expected, confirming that it was of yeast origin. We found this method to be very convenient and accurate as duplication, if any, in the gene sequence contained on the plasmid, was ruled out.

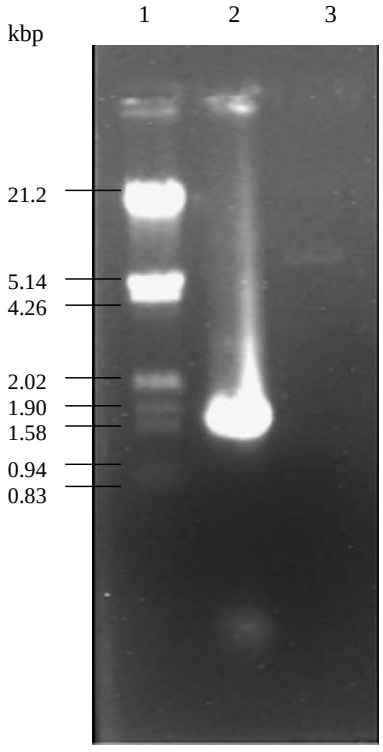

Fig. 2. Gel showing ~1.5kbp amplified fragment by genomic PCR. Lane 1, λ-*Eco*RI/*Hin*dIII digest; lane 2, amplified fragment using *P. etchellsii* genomic DNA as template; and lane 3, control tube with the template DNA and primers but without *Taq* polymerase.

Sequence analysis of the MUGA protein

The sequence analysis of the encoded MUGA protein by DNASTAR showed the presence of 89 basic amino acids (K, R), 89 acidic amino acids (D, E), 95 hydrophobic amino acids (A, I, L, F, and V), and 182 polar amino acids (N, C, Q, S, T, and Y). The predicted isoelectric point (p*I*) of the protein was 5.983, and the A + T% and G + C% was 58.75 and 41.25, respectively. The ratio of hydrophilic/hydrophobic amino acids was much higher than the value of 1.2 observed for most globular proteins [26]. This also suggested that the protein is likely to be exposed more to aqueous environment, expected for a cell-surface protein. Long stretches of Ser-Glu (SE) and Ser-Asp (SD) were observed in the gene sequence. The first stretch (SE) was localized near the N-terminus of the protein spanning from 62 to 106 position, interrupted only by a single Gln at 65 and Gly at 73, 82, and 96 positions. The second stretch (SD) was localized towards the C terminus from 431 to 454 position. A similar but short stretch of SD was also observed from 243 to 263 region, albeit a bit irregular. The comparison of the MUGA amino acid sequence with family 1 and family 3 β-glucosidases revealed little sequence similarity (less than 15%) but in the BLASTX search, similarity with a number of Ser-Asp dipeptides containing the protein families of *Staphylococcus aureus* (SdrC, SdrD, and SdrE) and *Staphylococcus epidermidis* (SdrF, SdrG, and SdrH) was observed (Table 1). The SD repeats were first reported in *S. aureus* in fibrinogen-binding clumping factors ClfA [27] and ClfB [28], and were located between a variable ligand binding region and a conserved C-terminal sequence associated with the attachment of these proteins to the cell wall. The SD repeat region is predicted to span the cell wall and extend the ligand binding region from the surface of the bacteria. The proteins SdrC, SdrD, and SdrE [29] contain an additional region, B, located between the ligand

Table 1
BlastX result of the comparison of MUGA deduced protein to the sequence databases

Sequence length	Protein ID	Name	% Identity	*E*-value
1253	DSPP_HUMAN	Dentin sialophosphoprotein (precursor)	24.8	1.24E−36
1733	Q9KI14	Putative cell-surface adhesin SdrF	26.1	1.56E−29
970	Q8VBY1	Phosphophoryn (precursor)	27.2	4.38E−29
3360	Q88XB6	Cell-surface SD repeat protein	25.5	4.38E−29
1633	Q8CMP4	Ser-Asp-rich fibrinogen-binding, bone sialoprotein-binding protein	25.5	8.71E−29
934	DSPP_MOUSE	Dentin sialophosphoprotein (precursor)	25.5	6.82E−28
1189	YJH6_YEAST	Hypothetical 128.5 kDa	48.7	4.10E−26
2310	Q8CMU7	Streptococcal hemagglutinin protein	28.4	1.60E−25
877	Q99R07	Clumping factor B	31.7	3.37E−24
907	Q8NUL0	Clumping factor B	30.5	6.61E−24
944	Q8NXJ1	Fibrinogen-binding protein	29.3	3.56E−23
935	Q932C5	Fibrinogen-binding protein	33.6	4.99E−23
989	Q99VJ4	Fibrinogen-binding protein A, clumping factor	33.6	4.99E−23
933	Q53653	Clumping factor	36.9	4.99E−23
485	Q8CNM7	Ser-Asp-rich fibrinogen binding protein	35.3	9.77E−23
487	Q9KI12	SdrH	35.3	9.77E−23
1056	Q8CQ72	Ser-Asp-rich fibrinogen binding protein	31.5	2.67E−22
1171	Q9KWX6	Bone sialoprotein-binding protein	32.1	1.02E−21

binding and the SD repeat region. Similar proteins (Sdr F, Sdr G, and Sdr H) have been found in *S. edidermidis* [30]. By analogy to these proteins, the MUGA protein contained shorter (24 aa) SD regions interspersed with a DSSSS motif (shown by bold italicized font in Fig. 1), suggesting similarity to the cation binding motif (DXSXS), which is reported to facilitate binding of ClfA with the internal site in the γ-chain of fibrinogen. In comparison to the SD repeat protein families mentioned, the MUGA protein also contained a long SE repeat near the N-terminus.

Sequence comparison between MUGA and the members of the SD repeat protein family was performed using CLUSTAL W and regions of homology were observed only in the SD repeat regions. Some (50%) identity was also observed in the C-terminus region. While *P. etchellsii* is a non-pathogenic yeast, the role of MUGA in regulating other surface related activities cannot be ruled out and needs to be further analyzed. We have conducted a secondary structural investigation and catalytic hydrolytic profile of the protein and the results are presented below.

```
Conf: 98754566832212352430463222022202345772364105751331113576455 6
Pred: CCCCCCCCCHHHCCCCEEECCCCCCCHHHHHCCCCCCCEEEECCCCHHHHHCCCCCCCC
  AA: MDESKSVRNPEEAKSDVLVGTKPEHGLKSKLSSEANSEHTIKANSNFEKLSETKSPSRNL
             10        20        30        40        50        60

Conf: 53344555555687544455687665555555556765554455566787300104414
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHCCCH
  AA: ESESQSESESESGSEESESESGSESESESESESESGSEAESESESESSSGPESEKKNGVN
             70        80        90       100       110       120

Conf: 432057888861100333322267667330000045651134554213102467516698
Pred: HHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCHHHHHHHCCCC
  AA: ELVDSNKPKKPENELSSEGTKSLPRTNTDRIAPITKGDASLPANLLNSKNSLQKTVGKSP
             130       140       150       160       170       180

Conf: 741024444556899999888741123577888201035556642101020789985478
Pred: CCCHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCEEEEEECCC
  AA: SSSKKQDILDKKAAEILREVSKHRRNSMNHNGSPDAIKNKETSDNKRRRAETILLLENSS
             190       200       210       220       230       240

Conf: 887877777888777788767777166577536654430343200666442113788766
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHCCCCCCHHHHHHHHHHHHHHHCCCCCCC
  AA: GSSSSDSDSEPNEDSSDSSDSDTPPKKMAKKAGVEKGSFAASRLLPKRLNSLSSRPPPKA
             250       260       270       280       290       300

Conf: 677889602000001360577742376667651110101110000133456762111002
Pred: CCCCCCCCCCHHHHCCCCEEEEECCCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCCHH
  AA: TSSSTPPVDVSNRFSEGVRTVSIETPFSKSSDEDNIRKEKEHSERKESSSNFNKESSLEK
             310       320       330       340       350       360

Conf: 212111001232212346654444652145687788886317830003221120220478
Pred: HHHHHEEECCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCCCCCCCCCEEECCC
  AA: KSLATKVQTKASTATAKSNETKKSTPLRKPLLNSLADLASRGVPDVLDAKEKSKVTLSKK
             370       380       390       400       410       420

Conf: 967640567777777777677777777777777753445555787763245432101 2
Pred: CCEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCCC
  AA: PPITVDSSSSSDSSSDSDSDSNSDSDSSSDSDSSNSESDLDSSNGSKGSKFVSLKKLSAE
             430       440       450       460       470       480

Conf: 442011001001367775431159
Pred: CCCHHHHCCCCHHHHHHHHHCCC
  AA: KNSREKKKSRSGFFSLMKDAKKRS
            490       500
```

Fig. 3. Secondary structure prediction by 3D-PSSM program. Helices are indicated by H, β-strands by E, and random coil by C. The numbers indicate the confidence levels assigned by the program for each residue; higher the number, greater the confidence level of the residue to be present in that conformation.

Secondary structure prediction

The putative reading frame was analyzed to predict the possible secondary structural elements and the results are shown in Fig. 3. Predominantly coiled and helical structures were predicted with very little β-sheet region, a structural feature characteristic of many cell surface/membrane bound proteins. Further analysis by 'DAS' TM-segment prediction program did not confirm the membrane nature of the protein, as the predicted helices were below the threshold set for the membrane proteins. According to the recent structural classification of proteins database (SCOP, Andreeva et al. [31]), the membrane proteins are placed in a separate category and provide useful information on the structural features of these proteins. The CD spectrum of the protein (see below) supported predicted secondary structure.

Purification and characterization of MUGA

The cellular localization studies indicated that about 40% of the total MUG hydrolytic activity was found in the extracellular medium. However, due to considerable proteolytic degradation in the extracellular medium, it was purified from the cell-free extract. The summary of purification is given in Table 2. Overall yield of 2.4% and specific activity of 805.5 U/g protein were obtained, which resulted in 52.8-fold purification of the protein. The electrophoretic analysis showed a homogeneous protein band (Fig. 4A). Analysis of the purified protein by MALDI-TOF indicated a single major *m*/*z* peak corresponding to a molecular mass of 52.1 kDa. Most of the β-glucosidases studied from yeasts are reported to be multimeric [32] and of high molecular mass, hence, this protein appears different. The dependence of hydrolytic activity on pH showed it to be active in the pH range of 7–9. The protein retained over 50% of its hydrolytic activity in the pH range of 6–11. The temperature optimum was 45 °C and the half-life of the protein was 26.6, 14.1, and 6.0 min at 40, 45, and 50 °C, respectively. At 55 and 60 °C, rapid inactivation of the protein occurred with half-lives of 2.5 and 1.0 min, respectively. The activity of the protein was determined on various aryl glucosides and other β-linked disaccharides. Maximum hydrolytic activity (168.5 mU/ml) was obtained on MUG followed by MUC (24.3 mU/ml). Relatively low activity was observed on *p*NPG, a characteristic substrate for β-glucosidase (Table 3). Almost a similarly low activity was also observed on *p*NPα Glu. Almond β-glucosidase was taken as a reference enzyme at a concentration that gave similar activity on MUG and, as seen, this

Table 2
Summary of purification of MUGA from pMG8:DH5α recombinant *E. coli*

Step	Total protein (mg)	Total activity (mU)	Specific activity (mU/mg protein)	Yield (%)	Fold-purification
Cell-free extract (40 ml)	400	6100	15.25	100	1
Ammonium sulfate precipitation (10 ml)	57	4060	71.23	66.5	4.7
DEAE-Sepharose chromatography (4 ml)	2.29	730	318.7	11.96	20.9
Hydroxyapatite chromatography (3 ml)	0.69	480	695.65	7.86	45.6
Sephadex G-25 (1 ml)	0.18	145	805.5	2.4	52.8

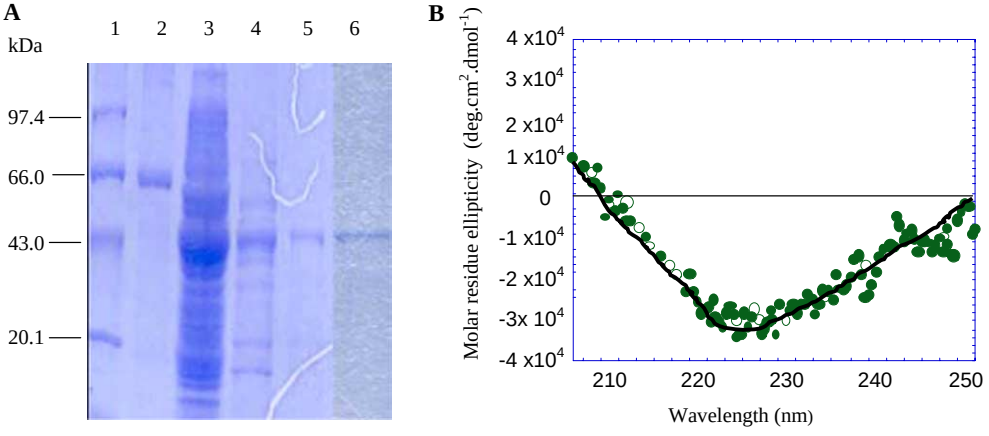

Fig. 4. (A) Migration of purified β-glucosidase in the SDS–PAGE: lane 1, molecular mass markers; lane 2, BSA protein; lane 3, 30–80% ammonium sulfate precipitated fraction; lane 4, DEAE–Sepharose active fractions (pooled); lane 5, hydroxyapatite active fractions (pooled); lane 6, Sephadex G-25 active fractions (pooled, 3 in number). (B) Far-UV CD spectrum of purified MUGA measured in 10 mM sodium cacodylate buffer, pH 7.0, at 25 °C. Protein concentration was 0.46 mg/ml. Spectrum was measured using a quartz cuvette of 1 cm path length and the value of ellipticity was converted to molar ellipticity as described in the text.

Table 3
Comparative hydrolytic profile of purified MUGA and almond β-glucosidase

Substrates	MUGA[a] (mU/ml)	Almond[b] (mU/ml)
Aryl glucosides		
p-Nitrophenyl β-D-glucopyranoside (*p*NPG)	0.56	2.72
p-Nitrophenyl β-D-xylopyranoside (*p*NPβ X)	0.09	0.12
p-Nitrophenyl β-D-galactopyranoside (*p*NPβ Gal)	0.07	0.35
p-Nitrophenyl α-D-glucopyranoside (*p*NPα Glu)	0.33	0.02
Disaccharides		
Cellobiose (β-1,4-linked glucose units)	0.10	0.27
Gentiobiose (β-1,6-linked glucose units)	0.05	0.25
Others		
Methylumbelliferyl β-D-glucoside (MUG)	168.5	178.0
Methylumbelliferyl β-D-cellobioside (MUC)	24.3	5.3
Salicin (2-hydroxyphenylmethyl β-D-glucoside)	0.03	0.44
Amygdalin (D-mandelonitrile 6-*O*-β-D-glucosido-β-D-glucoside)	No activity	0.92

[a] The protein concentration of MUGA was held constant in all reactions at 0.38 mg/ml.
[b] The almond enzyme concentration was kept at 10^{-3} mg/ml.

enzyme exhibited a higher relative activity on *p*NPG. Higher activity was also shown on gentiobiose, another characteristic substrate for these enzymes. No activity was observed of MUGA on amygdalin, as this is a substrate for plant β-glucosidases. It was also clear that the recognition and hydrolysis were specific for the aglycone moiety.

Equilibrium CD spectrum studies

Far-UV CD spectrum of active MUGA protein in sodium cacodylate buffer is shown in Fig. 4B. The extent of secondary structure in the native state of the protein was calculated to be 52% (α-helical) with no substantial β-sheet structure. Randomly coiled regions constituted 46% of the structure. These values were in good agreement with the predicted structure from the 3D-PSSM program. The nature of the CD spectrum appeared to be unique. Usually for all α, all β, or α + β containing proteins there are two negative troughs at about 220 and 210 nm [33] with different intensities corresponding to the extent of secondary structural elements. The CD spectrum of the unordered proteins has a negative trough around 200 nm. Since the MUGA protein does not contain substantial amount of β-sheet structure, the large negative trough around 220 nm is predominantly attributed to the large helical regions in the protein. The absence of any negative trough at 200 nm indicated the protein to have an organized structure. The data strongly indicated the protein to have a very different structure from those of resolved β-glucosidases.

Binding of MUGA protein to yeast cells

The SD repeat containing proteins are reported to be predominantly cell-surface-associated. While the role of these proteins is well understood in bacteria, where they mediate cell binding or clumping leading to pathogenicity, their role is not well defined in eukaryotes. In our BLASTX search results, an SD-rich hypothetical protein was also listed from the sequenced yeast genome (Table 1). Before a possible functional role can be assigned to these proteins, it must be demonstrated that they do indeed bind to the cell surface. For the MUGA protein, this binding was observed through fluorescence microscopy and the results are shown in Fig. 5. The PS grown yeast cells (Fig. 5A) when reacted with MUG did not show any fluorescence (Fig. 5B), indicating the absence of any MUG hydrolyzing activity by the cells. Incubating these cells (Fig. 5C) with the MUGA allowed binding to occur, which remained fairly strong in spite of repeatedly washing the cells with the buffer. The presence of the protein on the cell surface was monitored by bringing the cells in contact with the substrate on a solid plate. The hydrolytic activity of the protein resulted in the formation of the fluorescent product methylumbelliferone, which, during the initial period, remained associated with the cells making them fluorescent (Fig. 5D). As the time of incubation increased, the product started to diffuse away from the cells. Under longer conditions of incubation on the slide, the fluorescence increased to the extent that the cells appeared beautiful pale-violet when viewed under normal light (Fig. 5E).

Competition experiments for binding were performed between the purified MUGA and the SD dipeptide, and the results are shown in Fig. 6. The rationale of the designed experiment was based on the fact that if binding is mediated between the SD repeats in the purified protein, addition of the SD dipeptide should result in binding of lesser molecules of MUGA to the cell surface. This can be monitored by performing the MUG assay with the whole cells. Lesser the protein bound to the cells, lesser the released molecules of methylumbelliferone. The results in Fig. 6 indicate that more protein was bound to the cell surface in the absence of the dipeptide. As the concentration of the dipeptide increased in the experiment, it displaced more protein from the cell surface resulting in lower activity.

In conclusion, identification and sequence analysis of a novel β-glucosidase-like activity that exhibited hydrolysis

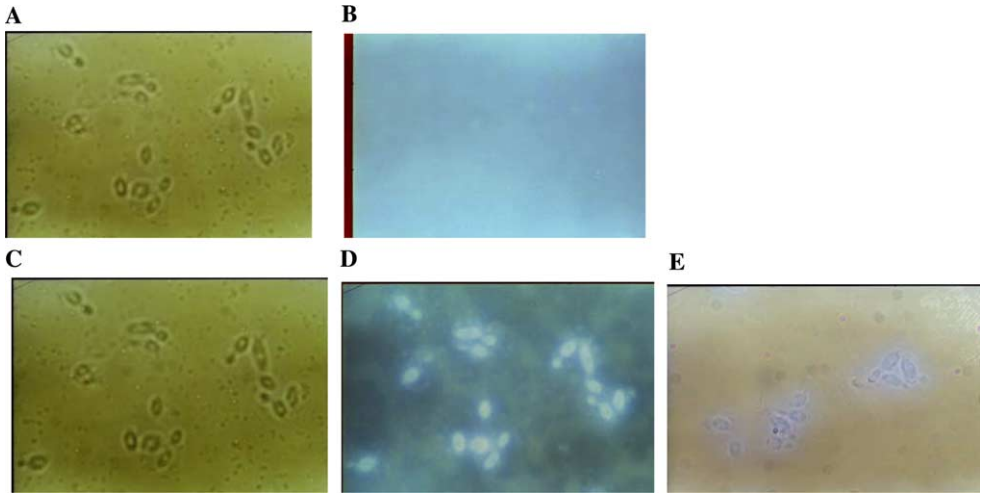

Fig. 5. Binding of purified MUGA to cell surface of *P. etchellsii*. Phosphate-succinate (PS) grown cells were mixed with the non-fluorescent substrate MUG on a pre-heated (37 °C) glass slide. The cells were incubated for 30 min at 37 °C and visualized under ordinary light (A) and fluorescent light (B). In a separate experiment, the PS grown cells (C) were incubated with the purified MUGA protein, washed as described in the text, and then allowed to come in contact with MUG. The cells were incubated at 37 °C and visualized under fluorescent light after 30 min (D) or under ordinary light (E) after 2.5 h of incubation.

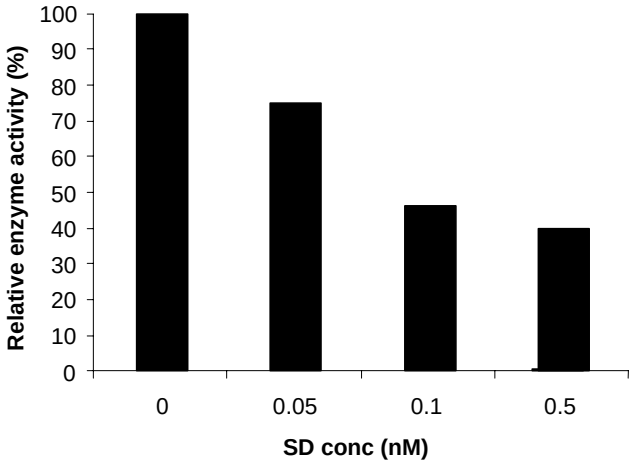

Fig. 6. Competition experiments between the purified MUGA and SD dipeptide. Enzyme activity in the supernatant, in the absence of any dipeptide, was taken as 100% and activities obtained in increasing concentrations of the dipeptide were expressed in terms of percentage of this number.

of MUG is reported in this paper. The amino acid sequence of the protein showed little similarity with the generic β-glucosidases. By virtue of regions containing long repeats of Ser-Asp (SD), the protein appeared to resemble the SD repeat containing proteins of *S. aureus* and *S. epidermidis*. The predicted secondary structure was predominantly helical/coil which was strongly supported by the CD spectral analysis. The binding of this protein with the yeast cell surface was confirmed through fluorescence microscopy. Although the role of this protein is not understood at the moment, it may function as a mediator of external signals based on the likely presence of a cation binding motif.

Acknowledgments

We thankfully acknowledge the help given by Dr. Sujata in the laboratory of Prof. T.P. Singh (Department of Biophysics, All India Institute of Medical Sciences, New Delhi), for conducting the MALDI-TOF analysis on the protein. Technical assistance of Mr. V.K. Ghose for fluorescence microscopy of the cells is gratefully acknowledged.

References

[1] V.S. Bisaria, S. Mishra, Cellulase biosynthesis and regulation, CRC Crit. Rev. Biotechnol. 5 (1989) 61–103.
[2] P. Tomme, R.A.J. Warren, N.R. Gilkes, Cellulose hydrolysis by bacteria and fungi, Adv. Microbiol. Physiol. 37 (1995) 1–81.
[3] A. Esen, β-Glucosidases: Biochemistry and Molecular Biology, American Chemical Society, Washington, DC, 1993.
[4] B. Brozobohaty, I. Moore, P. Kristofferson, L. Bako, N. Campos, J. Schell, K. Palme, Release of active cytokinin by a β-glucosidase localized to the maize root meristem, Science 262 (1993) 1051–1054.
[5] N.W. Barton, F.S. Furbish, G.T. Murray, M. Garfield, R.O. Brady, Therapeutic response to intravenous infusions of glucocerebrosidase in patients with Gaucher's disease, Proc. Natl. Acad. Sci. USA 87 (1990) 1913–1916.
[6] Y. Bhatia, S. Mishra, V.S. Bisaria, Microbial β-glucosidases: cloning, properties and applications, CRC Crit. Rev. Biotechnol. 22 (2002) 375–407.
[7] P.M. Coutinhu, B. Henrissat, Carbohydrate active enzymes server. Available from: <http://afmb.cnrs-mrs.fr/~pedro/CAZY> (2004).
[8] J.N. Varghese, M. Hrmova, G.B. Fincher, Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase, Structure 7 (1999) 179–190.
[9] M. Pandey, S. Mishra, Cloning and expression of β-glucosidase gene from the yeast *Pichia etchellsii*, J. Ferm. Bioeng. 80 (1995) 446–453.
[10] B. Sethi, M. Jain, M. Chowdhary, Y. Soni, Y. Bhatia, V. Sahai, S. Mishra, Cloning, characterization of *Pichia etchellsii* β-glucosidase II and effect of media composition and feeding strategy on its production in a bioreactor, Biotechnol. Bioprocess Eng. 7 (2002) 43–51.
[11] M. Pandey, S. Mishra, Expression and characterization of *Pichia etchellsii* β-glucosidase in *Escherichia coli*, Gene 190 (1997) 45–51.

[12] Y. Bhatia, S. Mishra, V.S. Bisaria, Purification and characterization of recombinant *Escherichia coli*-expressed *Pichia etchellsii* β-glucosidase II with high hydrolytic activity on sophorose, Appl. Microbiol. Biotechnol. 66 (2005) 527–535.
[13] Y. Bhatia, S. Mishra, V.S. Bisaria, Biosynthetic activity of recombinant *Escherichia coli* expressed *Pichia etchellsii* β-glucosidase II, Appl. Biochem. Biotechnol. 102/103 (2002) 367–379.
[14] P. Bacchawat, S. Mishra, Y. Bhatia, V.S. Bisaria, Enzymatic synthesis of oligosaccharides, alkyl and terpene glucosides by recombinant *Escherichia coli* expressed *Pichia etchellsii* β-glucosidase II, Appl. Biochem. Biotechnol. 118 (2004) 269–282.
[15] T. Kannan, D. Loganathan, Y. Bhatia, S. Mishra, V.S. Bisaria, Transglycosylation catalyzed by almonds β-glucosidase and cloned *Pichia etchellsii* β-glucosidase II using glycosylasparagine mimetics as novel acceptors, Biocatal. Biotransform. 22 (2004) 1–7.
[16] A. Wallecha, S. Mishra, Purification and characterization of two β-glucosidases from a thermo-tolerant yeast *Pichia etchellsii*, Biochim. Biophys. Acta 1649 (2003) 74–84.
[17] T. Maniatis, E.F. Fritsch, J. Sambrook, Molecular Cloning, A Laboratory Manual, second ed., Cold Spring Harbor Laboratory, New York, 1989.
[18] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W, improving the sensitivity of progressive multiple sequence alignments through sequence weighting, positions-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[19] L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, Enhanced genome annotation using structural profiles in the program 3D-PSSM, J. Mol. Biol. 299 (2000) 499–520.
[20] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head bacteriophage T4, Nature 227 (1970) 680–685.
[21] S. Gallagher, S.E. Winston, S.A. Fuller, J.G.R. Huriell, in: F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl (Eds.), Current Protocols in Molecular Biology, Analysis of Proteins, Publishing Associates and Wiley Interscience, New York, Unit 10.8, 1993.
[22] H. Van Tilbeurgh, F.G. Loontiens, C.K. De Bruyne, M. Claeyssens, Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases, Methods Enzymol. 160 (1988) 45–59.
[23] M.M. Bradford, A rapid and sensitive method for the quantitation of the microgram quantities of the protein utilizing the principle of protein dye binding, Anal. Biochem. 72 (1976) 248–254.
[24] T.K. Chaudhuri, K. Horii, T. Yoda, M. Arai, S. Nagata, T.P. Terada, H. Uchiyama, T. Ikura, K. Tsumoto, H. Kataoka, M. Matsushima, K. Kuwajima, I. Kumagai, Effect of extra N-terminal methionine residue on stability and folding of recombinant α-lactalbumin expressed in *Escherichia coli*, J. Mol. Biol. 285 (1999) 1179–1194.
[25] J.T. Yang, C.S. Wu, H.M. Martinez, Calculation of protein conformation from circular dichroism, Methods Enzymol. 130 (1986) 208–269.
[26] D. Voet, J.G. Voet, Biochemistry, second ed., Wiley, New York, 1995.
[27] D. McDevitt, P. Francois, P. Vandaux, T.J. Foster, Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*, Mol. Microbiol. 11 (1994) 237–248.
[28] D. NiEidhin, S. Perkins, P. Francois, P. Vandaux, M. Hook, T.J. Foster, Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*, Mol. Microbiol 29 (1998) 245–257.
[29] E. Josefsson, K.W. McCrea, D.N. Eidhin, D.O. Connell, J. Cox, M. Hook, T.J. Foster, Three new members of the serine-aspartate repeat protein multigene family of *Staphylococcus aureus*, Microbiology 144 (1998) 3387–3395.
[30] K.W. McCrea, O. Hartford, S. Davis, D.N. Eidhin, G. Lina, P. Speziale, T.J. Foster, M. Hook, The serine-asparatate repeat (Sdr) protein family in *Staphylococcus epidermidis*, Microbiology 146 (2000) 1535–1546.
[31] A. Andreeva, D. Howorth, S.E. Brenner, T.J.P. Hubbard, C. Chothia, A.G. Murzin, SCOP database in 2004: refinements integrate structure and sequence family data, Nucleic Acids Res. 32 (2004) D226–D229.
[32] A. Wallecha, Purification and characterization of two β-glucosidases from a thermo-tolerant yeast *Pichia etchellsii*, Ph.D. thesis, Indian Institute of Technology Delhi, 2002.
[33] G.D. Fasman, Circular Dichroism and the Conformational Analysis of Biomolecules, Plenum press, New York, 1996.