Co-purification from *Escherichia coli* of a plant β -glucosidase–glutathione S-transferase fusion protein and the bacterial chaperonin GroEL

Zsolt KERESZTESSY*‡, Jane HUGHES*, László KISS† and Monica A. HUGHES*§

*Department of Biochemistry and Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, U.K., and †Institute of Biochemistry, Lajos Kossuth University, Egyetem tér 1, Debrecen, H-4010 Hungary

The coding sequence of the mature cyanogenic β -D-glucosidase (β -D-glucoside glucohydrolase, EC 3.2.1.21) (linamarase) of *Manihot esculenta* Crantz (cassava) was cloned into the vector pGEX-2T and expressed in *Escherichia coli*. The bacterial chaperonin GroEL [Braig, Otwinowski, Hedge, Boisvert, Joachimiak, Horwich and Sigler (1994) *Nature* (London) **371**, 578–586] was found to be tightly associated with the fusion protein and co-purified with it. In the presence of excess MgATP, release and folding of the fusion β -glucosidase were demonstrated by a fast increase in both linamarase and p-nitrophenyl- β -D-glucopyranosidase activity at a low protein concentration. A slow endogenous folding process was also detected by activity measurements. Michaelis constants (K_m) and the ratio between the maximal velocities and efficiency constants (V_{max}/K_m) for the hydrolysis of the natural substrate, linamarin, and p-nitrophenyl β -D-glucopyranoside (PNP-Glc) by the recombinant

protein were found to be almost identical with those of the native glycosylated plant enzyme [Keresztessy, Kiss and Hughes (1994) *Arch. Biochem. Biophys.* **314**, 142–152]. Molecular dissociation constants for the free enzyme (pK^E_1 , pK^E_2) obtained with linamarin and PNP-Glc, and the enzyme substrate complexes (pK^{ES}_1 , pK^{ES}_2) were also in accordance with that of the original protein. The reactive substrate analogue N-bromoacetyl β -D-glucosylamine inactivated the fusion enzyme according to pseudo-first-order kinetics with first-order rate constant ($k_1 = 0.007 \text{ min}^{-1}$) and apparent inhibition constants ($K_i = 20 \text{ mM}$) comparable with those of the plant protein [Keresztessy, Kiss and Hughes (1994) *Arch. Biochem. Biophys.* **315**, 323–330]. In comparison with the native glycosylated plant protein, the recombinant protein was, however, found to be extremely sensitive to proteolysis and misfolding.

INTRODUCTION

The production of recombinant proteins in pro- and eu-karyotic expression systems has made a significant contribution to the progress that has been made in understanding the structure and function of these biomolecules. The use of recombinant DNA technology in protein engineering is also an important tool for the design and production of novel proteins for medical and industrial applications [1].

The overexpression of eukaryotic genes in bacteria such as *Escherichia coli* commonly gives rise to problems related to the insolubility of the produced foreign protein, leading to the formation of inclusion bodies by non-native-protein aggregates in the host cell [2]. Co-expression of protein folding factors [3], which are normally used by the host cell for the direction of folding processes of endogenous proteins, can provide a solution for this problem. These factors belong to the so-called ‘chaperonin’ molecules, which include the GroE proteins in *E. coli*. GroEL, which is a 14-mer complex of a M_r 60000 subunit and GroES, which is a 7-mer complex of an M_r 10000 subunit, assemble into a double-ring-shaped structure [4–6]. These molecules direct an ATP-dependent process which results in the protection of polypeptides from unfavourable intermolecular interactions during folding [7]. This molecular complex has also been used successfully for protein refolding *in vitro* [8].

β -Glucosidases have scientific, industrial and agricultural interest because of their involvement in important physiological processes and their potential as biotechnological tools in biomass degradation. These enzymes hydrolyse the β -glycosidic linkage

of monosaccharide glucosides or disaccharides according to a mechanism which involves an acid catalyst group for protonation of the glycosidic oxygen to help aglycone departure, and a nucleophile catalyst for the stabilization of a glycosylium cation-like transition state by ion-pair formation or as a covalent glycosyl-enzyme intermediate [9]. These catalytic side chains have been identified as carboxylates in several β -glucosidases [10–13].

The recombinant protein investigated in the present work is the cyanogenic β -glucosidase (linamarase) from the tropical crop plant *Manihot esculenta* Crantz (cassava). This enzyme is involved in a plant defence mechanism against herbivore attack, where, as a result of tissue injury, it brings about the hydrolysis of two cyanogenic glucosides, linamarin and lotaustralin, to produce glucose and hydroxynitriles (cyanohydrins). The release of HCN (cyanogenesis) results from the subsequent breakdown of the cyanohydrin by a hydroxynitrile lyase to produce a ketone and HCN [14]. Cassava is the third most important crop grown in the tropics, and cyanogenesis in the plant can cause serious health problems, owing to chronic cyanide poisoning [15,16]. Cassava linamarase is important, since it is involved in the detoxification of cassava food products during processing [17]. The commonly used lactic acid fermentation process, however, creates a demand for β -glucosidase engineered for more efficient catalytic efficacy under acidic conditions.

Cassava linamarase has been purified to homogeneity from the plant and characterized to be a 70000- M_r glycoprotein with high-mannose type oligosaccharide moieties estimated to account for 7–15% of the M_r [18]. The gene has been cloned and

Abbreviations used: PNP-Glc, p-nitrophenyl β -D-glucopyranoside; PATP-Glc, p-aminothiophenyl β -D-glucopyranoside; PTH, phenylthiohydantoin; IPTG, isopropyl 1-thio- β -D-galactopyranoside; GST, glutathione S-transferase.

‡ Present address: Institute of Biochemistry, Lajos Kossuth University, Debrecen, Egyetem tér 1, H-4010, Hungary.

§ To whom correspondence should be sent.

sequenced as cDNA [18]. Reaction kinetic studies in mixed-substrate and substrate/inhibitor systems, as well as investigation of the pH-dependence of the kinetic constants, demonstrated the involvement of two ionizable catalytic groups with $pK_{a1} = 4.6$ and $pK_{a2} = 7.1$ in the active site of the enzyme [19]. Specific chemical modification studies resulted in the identification of these groups as carboxylate side chains [19], and results of radioaffinity labeling studies using *N*-bromoacetyl β -D-[U- 14 C]-glucopyranosylamine show that the group with $pK_a = 7.1$ corresponds to Glu¹⁹⁸ of the protein, which plays an acid catalytic role in the hydrolytic mechanism of the β -glucosidase [13]. At present, together with the closely related white-clover (*Trifolium repens*) linamarase [20], the protein is under preliminary X-ray-crystallographic analysis. The accumulating information on the structure and function of this enzyme creates a solid basis for redesigning the catalytic properties using the tools of recombinant DNA technology.

In the present paper we report on the expression of the plant gene for the cyanogenic β -glucosidase as a fusion protein in *Escherichia coli*. The recombinant fusion protein co-purified with the bacterial chaperonin GroEL. The fusion β -glucosidase was characterized using reaction kinetic and pH-dependence studies, as well as inactivation kinetic measurements using *N*-bromoacetyl β -D-glucosylamine as an active-site-directed agent.

MATERIALS AND METHODS

Chemicals

p-Nitrophenyl β -D-glucopyranoside (PNP-Glc) *p*-aminophenyl 1-thio- β -D-glucopyranoside (PATP-Glc), linamarin, glutathione and ampicillin (Sigma Chemical Co.), glutardialdehyde-activated affinity-adsorbent (Boehringer), GSH-Sepharose 4B (Pharmacia Biotech), thrombin (CalBiochem). *N*-bromoacetyl β -D-glucopyranosylamine was synthesized as described by Thomas [19,21]. All other chemicals were purchased from BDH Laboratory Supplies (Poole, Dorset, U.K.).

Bacteria and plasmids

E. coli strain JM105 [22], the proteinase-deficient strains UT5600 [23] and PAM163 [24], as well as the endonuclease-deficient strain MM294 [25] were used as hosts. The glutathione S-transferase (GST) gene fusion vector pGEX-2T was supplied by Pharmacia Biotech.

Cloning and sequencing

Two oligonucleotide 27-mer primers, 5' ACT AGG CCA GCC CCG GGA ACT GAT GAT 3' and 5' TCA TAA TTG CCC GGG GAT ATT ATT TAA 3' were synthesized using an Applied Biosystems 381A DNA synthesizer. The primers correspond to bases 56–83 (sense) and bases 1653–1679 (antisense) respectively of the cassava cyanogenic β -glucosidase cDNA clone pCAS5 [18], modified to contain an *Sma*I-restriction-endonuclease recognition site to facilitate cloning in frame into the GST-fusion-protein expression vector pGEX-2T. The primers were used for PCR amplification of pCAS5 cloned into the *Not*I site of the plasmid vector Bluescript. A 40 ng portion of template DNA, 30 pmol of each primer, 0.2 mM dNTPs (Pharmacia polymerization mix), 1 $\times$ *Taq* polymerization buffer and 2 units of *Taq* polymerase were combined in a reaction volume of 100 μ l. PCR reaction conditions were: one cycle at 94 °C, 2 min; 24 cycles at 40 °C, 2 min; 72 °C, 4 min; 94 °C, 1 min; one cycle at 40 °C, 2 min; 72 °C, 5 min. The PCR product was digested with *Sma*I and cloned into the *Sma*I site of pGEX-2T. The orientation of inserts was determined by *Bam*HI and

*Eco*RI restriction digests, and the sequence of the junction between the vector and insert was confirmed with the Sequenase version 2 Kit (USB) using pGEX 5' primer (Pharmacia), the whole β -glucosidase insert was not resequenced. The clone selected for further analysis was named pEXCAS.

Expression and purification of the fusion protein

An overnight culture of *E. coli* strains containing pEXCAS was diluted 1:10 (v/v) in fresh Luria–Bertani medium supplemented with 50 μ g/ml ampicillin. Cells were grown at 37 °C for 1.5 h ($A_{600} \approx 0.4$). Expression of the fusion protein was induced by the addition of isopropyl β -D-thiogalactoside (IPTG) at a final concentration of 0.1 mM. Cells were grown for an additional 4 h at either 37 °C or 17 °C, pelleted by centrifugation at 5000 *g* for 10 min and resuspended in 2.5 % of the original volume of an ice-cold solution containing 140 mM NaCl, 2.7 mM KCl, 16 mM NaH_2PO_4 and 4 mM Na_2HPO_4 , pH 7.4 (PBS). After the addition of 1 mM (final concn.) PMSF and 1 % (v/v) Triton X-100, cells were disrupted by six to eight pulses of ultrasonication of 25 s at 8–10 μ in an MSE Soniprep 150 (MSE Scientific Instruments, Crawley, Sussex, U.K.) at 0 °C. The fusion protein was purified from the supernatant (10000 *g*, 10 min, 4 °C) by single-step affinity chromatography on GSH-Sepharose-4B. The bound fusion protein was eluted with 5 bed vol. of 10 mM GSH in 50 mM Tris/HCl, pH 8.0. The column eluate was concentrated and washed with 50 mM Tris/HCl, pH 8.0, in Filtron micro-concentrator tubes (10000 *M_r* cut-off) (Filtron Technology Co., Northborough, MA, U.S.A.). Protein concentrations were estimated using the method of Bradford [26], with BSA as a standard. Proteins were resolved on 10 %-SDS/PAGE gels and bands revealed by Coomassie Brilliant Blue R staining. A second affinity-purification step was also attempted using the competitive ligand of β -glucosidase, PATP-Glc, immobilized on a glutardialdehyde affinity adsorbent (Boehringer) according to the manufacturer's instructions. An aliquot of the fusion protein, eluted from GSH-Sepharose-4B beads, was diluted 1:10 (v/v) in 50 mM Tris/maleic acid, pH 5.6, and applied to 500 mg of the PATP-Glc affinity adsorbent. After washing with 5 bed vol. of the Tris/maleic acid buffer, the bound proteins were eluted with 1.8 M NaCl or 10 mM PATP-Glc in 0.2 M Tris/maleic acid, pH 5.6. Thrombin cleavage of the fusion protein was carried out for 2 h at 25 °C in PBS using 10 units/mg of protein.

Folding of the β -glucosidase–GST fusion protein was monitored in the absence or presence of 5 mM MgCl_2 and 5 mM ATP in 50 mM Tris/HCl, pH 8.0, at 25 °C, by measuring the PNP-Glc-ase and linamarase activity of small aliquots of the reaction mixtures, as described below. GroEL was separated from the fusion protein by GSH-Sepharose-4B affinity chromatography after 5 mM MgATP treatment.

Enzyme assays and kinetic measurements

The β -glucosidase activity of protein samples against the natural substrate, linamarin (linamarase), was estimated as described previously [19]. Enzyme reactions were carried out in a 0.25 ml final volume in 0.15 M citrate/phosphate buffer, pH 6.0, at 37 °C. Reactions were initiated by the addition of small aliquots (10–50 μ l) of the protein solutions and were stopped by cooling on ice and addition of NaOH to a final concentration of 0.1 M. The liberated HCN was quantified with the pyrazolon/bipyrazolon reagent and spectrophotometry at 620 nm [27]. PNP-Glc-ase activity was measured using the same buffer as that described above, in a 0.5 ml final reaction volume. The liberated PNP was quantified by spectrophotometry at 400 nm under basic conditions [19].

Reaction kinetic parameters ($V_{\max}$, K_m , $V_{\max}/K_m$) were determined using six different substrate concentrations ranging between 0.5 and 20 K_m . The values of the maximal velocity ($V_{\max}$) and the Michaelis constant (K_m) were calculated by using the classic Michaelis–Menten rate equation. In pH-dependence studies of the kinetic parameters, six different pH values were used between pH 4 and 8 in 0.075 M citrate/phosphate buffer. Theoretical pH profiles were non-linearly fitted to the data points using the diprotic enzyme model [19,28].

Irreversible inactivation of the GST- β -glucosidase was carried out with *N*-bromoacetyl β -D-glucopyranosylamine, using 0 and 6–50 mM as the reagent concentrations. Reactions were carried out in 0.25 ml final volumes in 0.15 M citrate/phosphate buffer, pH 6.0. Aliquots were withdrawn at intervals, and PNP-Glc-ase activity was determined as above. Kinetic parameters (k_i , K_i) were determined as described previously [13]. Calculation of kinetic constants and theoretical curves was carried out by using the program Graft [29].

Peptide sequence determination

Protein samples were boiled for 2 min with an equal volume of SDS/PAGE sample buffer containing 0.125 M Tris/HCl, pH 6.8, 4% SDS, 20% glycerol, 5% 2-mercaptoethanol and 0.002% Bromophenol Blue. The samples were run on an 10%-SDS/PAGE gel for 4.5 h at 40 mA. The separated polypeptides were electroblotted on to Hybond using a wet Bio-Rad (Richmond, CA, U.S.A.) Transblot Cell™ for 60 min at 50 mA in 10 mM 3-(cyclohexylamino)propane-1-sulphonic acid ('CAPS') buffer. The membrane was stained with 0.2% Coomassie Brilliant Blue R (Sigma) in 40% methanol/10% acetic acid at room temperature. After destaining the membrane in the same solvent for 10 min and drying, protein bands were excised and analysed on a Beckman LF 3000 protein microsequencer by Dr. J. Gray (University Facility for Molecular Biology, University of Newcastle upon Tyne, Newcastle upon Tyne, U.K.).

RESULTS AND DISCUSSION

Production of the GST- β -glucosidase fusion enzyme

In order to characterize the non-glycosylated form of the cyanogenic β -glucosidase of cassava, the coding DNA sequence of the mature enzyme was amplified from the cDNA clone pCAS5 [18] by PCR and subcloned into the *E. coli* expression vector pGEX-2T (Figure 1). The N-terminal sequence of the mature active plant enzyme is known [18] and indicates that a signal peptide is removed in the plant. The construct inserts an extra glycine residue between the N-terminus of the mature β -glucosidase and the pGEX-2T cloning site. The resulting expression plasmid (pEXCAS) produces a fusion protein with the β -glucosidase coding sequence subcloned in frame after the 3' prime end of the coding sequence of GST protein ($M_r = 27000$),

Figure 1 Fusion protein construct in pEXCAS

GST, GST coding sequence; CAS5, β -glucosidase coding sequence; restriction sites from pGEX-2T expression vector above the construct, *, cloning site; restriction sites from β -glucosidase coding sequence below the line.

Figure 2 Expression of the cassava β -glucosidase cDNA and purification of the fusion protein

E. coli JM 105 cell extracts were analysed by 10%-SDS/PAGE; gels were stained with Coomassie Brilliant Blue R. Lane 1, crude lysate from uninduced cells containing pGEX2T; lane 2, crude lysate from induced cells containing pGEX2T; lane 3, crude lysate from uninduced cells containing pEXCAS; lane 4, cell debris from induced cells containing pEXCAS; lane 5, crude lysis supernatant from induced cells containing pEXCAS; lane 6, PATP-Glc-affinity-adsorbed fraction from the GSH-Sepharose-4B purified fraction; lane 7, GSH-Sepharose-4B-purified fraction from crude lysis supernatant of induced cells containing pEXCAS (lane 5); M, M_r markers ($M_r \times 10^{-3}$: 92, 67, 45 and 29). Coomassie Blue-stained bands referred to in the text are marked as follows: M_r 87 000; M_r 60 000; M_r 27 000.

making possible single-step affinity purification using GSH-Sepharose-4B.

Induction of host *E. coli* cells harbouring pEXCAS with IPTG resulted in the appearance of a new M_r -87000 protein in the crude cell lysate pellet (Figure 2, lane 4) corresponding to the predicted M_r of the mature non-glycosylated β -glucosidase (M_r 60000) [18] plus the GST protein (M_r 27000). Although the new band of M_r 87000 was not apparent in the crude lysate supernatant of induced cells (Figure 2, lane 5), it was subsequently shown to be present by purification (Figure 2, lane 7). The induced control cells, containing pGEX-2T, over-produced the 27000- M_r GST protein (Figure 2, lane 2) and lacked the 87000- M_r polypeptide.

The amount of the soluble fusion protein produced by the *E. coli* cells was dependent on culture age and induction temperature. Analysis of the pellet and soluble fraction of crude lysates (Figure 2, lanes 4 and 5) showed that induction of cultures with an $A_{600} \approx 1.0$ at 37 °C leads to the formation of inclusion bodies containing the fusion protein. However, induction of cultures with $A_{600} \approx 0.35$ at 17 °C produced essentially no insoluble fusion protein (results not shown). It is to be supposed that the lower temperature results in slower production and proper folding of the fusion β -glucosidase, thus preventing the aggregation of unfolded species into inclusion bodies.

Analysis of crude cell lysates and soluble extracts for the activity of β -glucosidase (linamarase) showed that hydrolytic activity against the natural substrate (linamarin) of the plant enzyme was exclusively associated with induced cells harbouring the pEXCAS plasmid and producing the 87000- M_r fusion protein. Specific and total linamarase activity of lysates from induced cells containing pGEX-2T was zero, while a specific activity of 70 $\mu\text{kat} \cdot \text{kg}^{-1}$ and a total activity of 8.4 nmol/s per litre of culture of induced cells with pEXCAS plasmid were detected. This fact shows that the non-glycosylated version of

Figure 3 Endogenous proteolytic degradation and activity of the fusion protein at 4 °C

(a) Aliquots (25 μ l) of protein, purified by GSH–Sepharose-4B affinity chromatography from *E. coli* host cells JM105 containing pEXCAS and induced with 0.1 mM IPTG at 37 °C (lanes 1–6) or 17 °C (lanes 7–12) were analysed on 10%–SDS/PAGE gels and stained with Coomassie Brilliant Blue R. Aliquots (25 μ l) were withdrawn after 1 day (lanes 1 and 7), 3 days (lanes 2 and 8), 8 days (lanes 3 and 9), 14 days (lanes 4 and 10), 16 days (lanes 5 and 11) and 21 days (lanes 6 and 12) storage at 4 °C. M, M_r markers ($M_r \times 10^{-3}$: 92, 67, 45 and 29). Coomassie Blue-stained bands referred to in the text are marked as follows: $\bullet$, M_r 87 000; $\circ$, M_r 60 000; $\blacktriangle$, M_r 56 000; $\triangle$, M_r 31 000; $\circ$, M_r 27 000. (b) Linamarase activity of preparations shown in Figure 2. Samples (500 μ l) were stored at 4 °C, and aliquots were withdrawn at intervals for activity measurements (10 μ l) and protein determination (10 μ l). Specific ($\triangle$, $\blacktriangle$) and total ($\circ$, $\bullet$) enzyme activity of the samples taken from preparations obtained from *E. coli* cells induced at 37 °C (closed symbols) and 17 °C (open symbols) were determined as described in the text.

this glycosylated plant β -glucosidase is capable of folding into a conformation which retains its activity against the natural and synthetic substrates.

Purification of the fusion GST- β -glucosidase

Large-scale purification of the fusion protein by affinity chromatography on GSH–Sepharose-4B typically led to 0.2 mg and 0.1 mg fusion protein per litre of culture induced at 37 °C and 17 °C respectively, with 60 % of the total linamarase activity adhering to the GSH–Sepharose beads. When an induction temperature of 17 °C was used, the fusion protein was obtained with a higher total and specific activity against linamarin (2.4 nmol $\cdot$ s $^{-1}$, 9.4 mkat $\cdot$ kg $^{-1}$) compared with the 37 °C induction temperature (1.3 nmol $\cdot$ s $^{-1}$, 1.24 mkat $\cdot$ kg $^{-1}$). Analysis of the fractions eluted with 10 mM GSH showed, in addition to the M_r 87 000 fusion protein, a major (M_r 60 000) and several minor contaminating proteins (Figure 2, lane 7).

To investigate endogenous proteolysis occurring in the semi-pure GSH-affinity preparations, fusion protein was purified from cells induced at either 37 °C or 17 °C. Aliquots were taken during a 20-day 4 °C incubation of the protein and analysed by SDS/PAGE (Figure 3a). The 60 000- M_r protein, which co-purifies with the fusion protein, appears as a major component and is resistant to proteolysis. The sequential pattern of fragments produced during this incubation indicates that the 87 000- M_r

fusion protein is cleaved to produce a 31 000- M_r and a 56 000- M_r fragment, since a parallel accumulation of the two products can be observed after 3 days. Subsequently the 31 000- M_r fragment is degraded to a 27 000- M_r product, which corresponds to the molecular mass of GST. N-terminal peptide sequencing of the 27 000- M_r fragment resulted in the identification of an Met-Ser-Pro-Ile-Leu-Gly-(Tyr) sequence which is the GST sequence [30] between residues 1 and 6(7) (Table 1a). Interestingly both the total and the specific β -glucosidase activity against linamarin increased during the first 15 days of 4 °C incubation (Figure 3b).

Further purification of the GSH-affinity purified fraction used the competitive inhibitor of the native plant enzyme PATP-Glc immobilized on to silica beads via the primer amino group of the phenyl aglycone through a glutardialdehyde spacer. Most of the linamarase activity (95 %) present in the samples bound to this affinity adsorbent. Proteins associated with the affinity adsorbent were analysed by SDS/PAGE, and the result (Figure 2, lane 6) shows that both the fusion protein and the 60 000- M_r Coomassie band binds to the immobilised PATP-Glc. Although the active β -glucosidase could not be purified to a single Coomassie Blue-stained SDS/PAGE component, activity against the natural substrate was always associated with the presence of either the fusion protein or the 56 000- M_r polypeptide.

In order to reduce endogenous proteolysis, the proteinase-deficient and an endonuclease mutant *E. coli* strains (UT5600, PAM163 and MM294, respectively) were used as hosts for pEXCAS. GSH-affinity purified soluble protein was analysed by SDS/PAGE; however, these hosts are not deficient in all proteinases and, in comparison with *E. coli* JM105, both proteolytic degradation and linamarase activity remained the same (results not shown). In the deduced amino acid sequence of the cassava β -glucosidase [18], 11 potential thrombin cleavage sites were detected. Digestion of the fusion protein with thrombin led to the cleavage of the protein into small fragments and loss of β -glucosidase activity (results not shown).

Identification of the co-purifying M_r 60 000 protein as GroEL

The 60 000- M_r molecule co-purifying with the fusion protein was analysed by solid-phase N-terminal sequencing after trans-blotting to a poly(vinylidene difluoride) anion-exchange membrane (Table 1b). Six sequencing cycles resulted in the identification of the sequence Ala-Ala-Lys-Asp-Val-Lys, which corresponds to that of the *E. coli* GroEL subunit ($M_r \approx 60 000$) starting from residue 2 [4]. GroEL is a member of the Hsp60 class of heat-shock-induced chaperonins and plays an essential role in the ATP-dependent folding of polypeptide chains. This ring-shaped 14-mer protein complex, which is composed of $\approx 60 000$ - M_r subunits, binds partially or misfolded proteins in the cytosol of *E. coli* cells, preventing them from aggregational denaturation. Together with the modulating effect of the 7-mer GroES ($\approx 10 000$ - M_r subunits), GroEL provides a sheltered environment for proteins to fold in and subsequently releases them in their native state. The fact that this chaperonin was found to be tightly associated with the 87 000- M_r fusion β -glucosidase is probably an indication of difficulties encountered in fusion-protein folding. The higher levels of GroEL (Figure 3) and the increased aggregation of the fusion protein in cells grown at 37 °C, compared with 17 °C, may reflect either the effect of temperature *per se* or the effect of protein concentration on the fusion-protein folding process [2].

The cavity of the GroEL complex is capable of including proteins with an M_r lower than 60 000 [31]. The fact that the fusion protein co-purified with GroEL on the GSH–Sepharose affinity column indicates that the GST domain is not enclosed

Table 1 Solid phase N-terminal peptide sequencing data

Polypeptides were resolved by 10%-SDS/PAGE, stained with Coomassie Blue R and electroblotted on to Hyperbond membrane. Edman degradation was carried out on a Beckman LF 3000 peptide sequencer. Absolute yields of the phenylthiohydantoin-amino acids are given in pmol.

(a) Edman-degradation of the 27 kDa fragment

Cycle	Phenylthiohydantoin derivative	Yield (pmol)		cDNA sequence*
		Sequence I	Sequence II	
1	Met	16.59	—	Met
2	Ser + Cys/Ser	12.41	6.06	Ser
3	Pro	9.69	6.87	Pro
4	Ile	8.28	5.41	Ile
5	Leu	8.74	5.71	Leu
6	Gly	12.73	9.96	Gly
7	Tyr	—	4.43	Tyr

* Sequence of peptide derived from the cDNA sequence [30] starting at residue 1.

(b) Edman-degradation of the 60 kDa fragment

Cycle	Phenylthiohydantoin derivative	Yield (pmol)		cDNA sequence*
		Sequence I	Sequence II	
1	Ala	4.76	Ala	Ala
2	Ala	4.87	Ala	Ala
3	Lys	2.69	Lys	Lys
4	Asp	3.21	Asp	Asp
5	Val	4.70	Val	Val
6	Lys	1.92	Lys	Lys

* Sequence of peptide derived from the cDNA sequence [4] starting at residue 2.

Figure 4 Liberation and folding of the fusion β -glucosidase from GroEL in the presence of MgATP

Rate of PNP-Glc hydrolysis by a fusion β -glucosidase preparation (0.25 mg/ml) affinity-purified from cells induced at 37 °C and incubated at room temperature (25 °C) in the presence of 5 mM MgCl_2 (○) and in the presence of 5 mM MgCl_2 plus 5 mM ATP (●), determined in 0.15 M citrate/phosphate buffer, pH 6, at 37 °C.

within the chaperonin complex. This is consistent with the major primary degradation products of the fusion protein being a 31000- M_r polypeptide containing GST and a 56000- M_r polypeptide predicted to be the C-terminal 491 amino acids of the β -glucosidase. This complex, isolated from 37 °C-induced cells, can be dissociated by the addition of MgATP at room temperature, resulting in significant increase in β -glucosidase activity (Figure 4), presumably owing to the release of the recombinant protein in its native folded state. This is consistent with the endogenous

Scheme 1 Hypothetical model for the release, folding and degradation of the fusion β -glucosidase from GroEL cells

The misfolded cyanogenic β -glucosidase domain (CBGL) of the fusion protein is included in the central cavity of the GroEL complex, while the GST domain is exposed to the solvent in its native conformation indicated by its binding affinity to GSH-Sepharose-4B. Primary proteolytic cleavage can possibly occur to produce GST linked up to the N-terminal 41 amino acids of the CBGL domain, and a β -glucosidase fragment (CBGL') composed of the C-terminal 491 amino acids ($M_r \approx 56000$) hidden by GroEL. This fragment (CBGL') can be released and fold spontaneously or in a MgATP-mediated manner. Alternatively, the release and folding of the CBGL domain of the fusion protein can occur before the proteolytic cleavage (GST-CBGL). Because of the proteinase-sensitivity of the unglycosylated CBGL domain, it is sequentially degraded into small fragments by contaminating peptidases. (Domains in boxes indicate properly folded species.)

increase in β -glucosidase activity seen during a 15-day incubation at 4 °C, which could be due to a spontaneous release of the fusion protein or its active fragment(s) from the GroEL complex. Similar results were observed during refolding of a β -lactamase polypeptide [8,31]. The complexity of the patterns of the fusion-protein samples can thus be attributed to release and folding from GroEL, as well as proteolytic cleavage resulting in active and inactive species. A summary of these processes is proposed in Scheme 1.

Steady-state kinetic parameters and irreversible inactivation of the GST- β -glucosidase

Kinetic constants ($V_{\max}$, K_m and $V_{\max}/K_m$) were determined for the GSH-affinity-purified protein produced by *E. coli* cells induced at 17 °C for the hydrolysis of the synthetic substrate PNP-Glc and the natural substrate, linamarin. The values of the

Figure 5 pH-dependence of the kinetic parameters of the fusion β -glucosidase

The pH-dependence of the values of parameters $V_{\max}$, K_m and $V_{\max}/K_m$ for linamarase ($\square$) and PNP-Glc-ase ($\circ$) activities were determined in 0.075 M citrate/phosphate buffer between pH 4 and 8. The best-fit curves were calculated by non-linear fitting on the basis of the diprotic enzyme model [19,28].

Table 2 Molecular dissociation constants of the free enzyme (E) and the enzyme-substrate complexes (ES); and pH-corrected kinetic parameters of the recombinant cassava linamarase expressed in *E. coli*

The kinetic parameters ($V_{\max}$, K_m and $V_{\max}/K_m$) for the hydrolysis of linamarin and PNP-Glc were determined in 0.075 M citrate/phosphate buffer between pH 4 and 8, at 37 °C. Substrate concentrations varied between 0.5 and 20 K_m . pK_1^E , pK_2^E , pK_1^{ES} , pK_2^{ES} (the negative of the logarithm of the molecular dissociation constants of the free enzyme K_1^E , K_2^E and the ES complexes K_1^{ES} , K_2^{ES} respectively), $V_{\max}$, $V_{\max}/K_m$ were calculated by non-linear fitting of the appropriate equations of the diprotic enzyme model [19,28] to the experimental data. The S.D. of the values calculated was less than 7 %. K_m values were calculated from $V_{\max}/K_m$ and $V_{\max}$.

Parameter	Substrate ...	Linamarin	PNP-Glc
pK_1^E		4.5 ± 0.2	3.9 ± 0.1
pK_2^E		7.1 ± 0.2	7.9 ± 0.1
pK_1^{ES}		3.9 ± 0.1	4.2 ± 0.4
pK_2^{ES}		8.5 ± 0.1	—
$V_{\max}$ (mkat/kg)		2.0 ± 0.1	2.9 ± 0.1
$V_{\max}/K_m$ (kat/kg per M)		2.3 ± 0.4	9.0 ± 0.3
K_m (mM)		0.88	0.32

Michaelis constants (K_m 0.35 ± 0.04 mM for PNP-Glc and 1.14 ± 0.2 mM for linamarin) were essentially identical with those of the native plant enzyme (0.32 ± 0.3 mM for PNP-Glc and 1.2 ± 0.1 mM for linamarin) [19]. $V_{\max}$ values obtained with the partially purified recombinant protein ($V_{\max}$ 2.9 ± 0.1 mkat $\cdot$ kg $^{-1}$ for PNP-Glc, and 2.0 ± 0.1 mkat $\cdot$ kg $^{-1}$ for linamarin) were obviously lower than those of the plant protein purified to homogeneity (330 ± 10 for PNP-Glc and 420 ± 20 for linamarin) [19]. Interestingly, the maximal velocity of substrate hydrolysis was

Figure 6 Irreversible inactivation of the fusion β -glucosidase with *N*-bromoacetyl β -D-glucopyranosylamine

A semi-logarithmic plot of the residual PNP-Glc-ase activity against time is shown. The GST- β -glucosidase GSH-Sepharose-4B-affinity-purified from cells induced at 17 °C overnight was incubated at a protein concentration of 0.123 mg/ml in 0.15 M citrate/phosphate buffer, pH 6.0, in the presence of 6 mM ($\circ$), 10 mM ($\bullet$), 24 mM ($\square$), 40 mM ($\blacksquare$) and 50 mM ($\triangle$) inactivator. Inset: double-reciprocal plot of the slopes of the straight lines, fitted to the experimental points, against inactivator concentration.

higher in the case of PNP-Glc for the recombinant protein, while the opposite is true for the plant enzyme. The ratio (5.177:1) between the efficiency constants ($V_{\max}/K_m$) for PNP-Glc (9.06 kat $\cdot$ kg $^{-1}$ $\cdot$ M $^{-1}$) and linamarin cleavage (1.75 kat $\cdot$ kg $^{-1}$ $\cdot$ M $^{-1}$) was higher than that (3.4:1) for the plant enzyme. These results indicate that the hydrolysis of linamarin, which supposedly has a slightly different mechanism for the hydrolysis of the synthetic substrate PNP-Glc with its PNP (pK_a 7.18) leaving group [19], is more sensitive to conformational changes caused by the lack of the oligosaccharide side chains [18], the fusion to GST or by proteolytic truncation of the protein.

pH-dependence of $V_{\max}$ and $V_{\max}/K_m$ for PNP-Glc and linamarin hydrolysis (Figure 5) showed characteristics similar to, but not identical with, those of the glycosylated plant enzyme [19]. Molecular dissociation constants of the free enzyme (pK_1^E and pK_2^E) and the ES-complexes (pK_1^{ES} and pK_2^{ES}) obtained from the theoretical curves fitted to the experimental data using the appropriate equations of the diprotic enzyme model [28] are summarized in Table 2. Using double-reciprocal plots, the bell-shaped pH-activity profiles could be characterized by limiting slopes of +1 and -1, which indicates that the decrease in parameters was a result of the change in the protonation state of single amino acid residues in the pH range used.

Using linamarin as the substrate, the pK_a values of the two ionizable catalytic groups of the enzyme were estimated to be 4.5 ± 0.2 and 7.1 ± 0.2 , which were the same as those determined for the native plant β -glucosidase [19]. The wide pH-efficiency constant ($V_{\max}/K_m$) profile for PNP-Glc might be an indication of the presence of different active species (fragments) of the

recombinant protein having altered protic characteristics, which might have a contribution to the total PNP-Glc-ase, but not linamarase, activity. This is supported by the higher $V_{\max}$ value for PNP-Glc-ase activity and the increased ratio between $V_{\max}/K_m$ for PNP-Glc and linamarin compared with native plant enzyme [19].

The cassava cyanogenic β -glucosidase (linamarase) was demonstrated to be irreversibly inactivated by the affinity label *N*-bromoacetyl β -D-glucopyranosylamine via the formation of an ester bond with the acid catalyst Glu¹⁹⁸ (pK_a 7.1) in the active site of the enzyme [13]. Affinity-purified recombinant β -glucosidase preparations were also tested for the loss of both linamarase and PNP-Glc-ase activity during incubation with the abovementioned reactive substrate analogue (Figure 6). The observed kinetics of inactivation were in accordance with the theoretical equations provided by Kitz and Wilson [32] for active-site-specific reagents, and the apparent inhibition constant (K_i 20.5 mM) as well as the order of reaction with respect to *N*-bromoacetyl β -D-glucopyranosylamine (0.8) were essentially the same as those for the plant protein (24.5 mM and 0.9 respectively) [13]. The first-order rate constant (k_i 0.007 min⁻¹) and the efficiency constant (k_i/K_i 0.342 min⁻¹·M⁻¹) were lower than, but comparable with, those of the native plant enzyme. It may be supposed that the lower values were due to the presence of more than one recombinant β -glucosidase protein species with different contributions to the total activity and, thus, to the inactivation process. These data, together with the kinetic parameters obtained for the recombinant β -glucosidase, provides evidence for the expression in *E. coli* of the non-glycosylated version of the cassava β -glucosidase in an active form and with only slightly altered kinetic properties.

Conclusions

The results reported in the paper demonstrate the expression in *E. coli* of the coding sequence for a plant β -glucosidase using the pGEX-2T vector which results in the production of a non-glycosylated GST fusion protein. The original plant enzyme is produced in very high amounts in the latex vessels of the cassava plant. The mature protein is highly glycosylated (7–15% carbohydrate content) [18] and is very stable. The *E. coli* recombinant protein, however, is extremely sensitive to proteinases and at high concentration tends to aggregate and form inclusion bodies. Folding difficulties encountered in this process can be overcome to some extent using lower induction temperature and lower level of expression.

As a result of the presence of misfolded species, the bacterial chaperonin, GroEL, associates and co-purifies with the fusion β -glucosidase. MgATP treatment of the GST- β -glucosidase-GroEL complex triggers an increase in β -glucosidase activity which can be attributed to the release of active recombinant protein from the chaperonin.

The plant β -glucosidase was characterized with respect to its kinetic behaviour and chemical mechanism using mixed substrate and inhibitor systems, specific chemical modifications [19] and affinity-labelling studies [13], resulting in the identification of

catalytic residues. The recombinant fusion protein has less efficient folding abilities, is less stable, but has essentially unaltered kinetic and acid-base catalytic properties. The production of the unglycosylated cassava cyanogenic β -glucosidase in an active form in *E. coli* makes possible further studies of the structure and function of this enzyme.

This work was supported by The British Council, the Academy of Sciences of Hungary, the Ministry of Education of Hungary, and EC STD III Contract TS3-CT92-0108. We thank Dr. H. R. Haysom for the graphical design, and Dr. J. Gray for peptide sequencing.

REFERENCES

- Oxender, D. L. and Fox, C. F. (1989) Protein Engineering, Alan R. Liss, New York
- Marston, A. O. (1986) Biochem. J. **240** 1–12
- Golubinioff, P., Gateby, A. A. and Lorimer, G. H. (1989) Nature (London) **337** 44–47
- Braig, K., Otwinowski, Z., Hedge, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. and Sigler, P. (1994) Nature (London) **371** 578–586
- Feyton, W. A., Yechezkel, K., Furtak, K. and Horwich, A. L. (1994) Nature (London) **371** 614–619
- Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A., Clarke, A. R. and Saibil, H. R. (1994) Nature (London) **371** 261–264
- Hendrick, J. P. and Hartl, F.-U. (1993) Annu. Rev. Biochem. **62** 349–384
- Lamiet, A. A., Ziegelhoffer, T., Georgopoulos, C. and Plückthun, A. (1990) EMBO J. **9** 2315–2319
- Sinnott, M. L. (1987) in Enzyme Mechanism (Page, M. I. and Williams, A., eds.), p. 259, Royal Society of Chemistry, London
- Withers, S. G., Warren, A. J., Street, I. P., Rupitz, K., Kempton, J. B., and Aebersold, R. (1990) J. Am. Chem. Soc. **112** 5887–5889
- Tribur, D. E., Warren, R. A. J., and Withers, S. G. (1992) J. Biol. Chem. **267** 10248–10251
- Witt, E., Frank, R. and Hengstenberg, W. (1993) Protein Eng. **6** 913–920
- Keresztesy, Zs., Kiss, L. and Hughes, M. A. (1994) Arch. Biochem. Biophys. **315** 323–330
- Conn, E. E. (1981) in The Biochemistry of Plants (Stumpf, P. K. and Conn, E. E., eds.), vol. 7, pp. 479–500, Academic Press, New York
- Tylleskär, T., Banea, M., Bikangi, N., Cooke, R. D., Poulter, N. H. and Rosling, H. (1992) Lancet **339** 208–211
- Gomez, G. and Valdivieso, M. (1985) J. Sci. Food Agric. **36** 433–441
- Nambisan, B. and Sundaresan, S. (1985) J. Sci. Food Agric. **36** 1197–1203
- Hughes, M. A., Brown, K., Pancoro, A., Murray, B. S., Oxtoby, E., and Hughes, J. (1992) Arch. Biochem. Biophys. **295** 273–279
- Keresztesy, Zs., Kiss, L. and Hughes, M. A. (1994) Arch. Biochem. Biophys. **314** 142–152
- Tolley, S. P., Barrett, T. E., Suresh, C. G. and Hughes, M. A. (1993) J. Mol. Biol. **229** 791–793
- Thomas, E. V. (1970) J. Med. Chem. **13** 755–756
- Yanish-Perron, C., Uieira, J. and Messing, J. (1985) Gene **33** 103–119
- Elish, M. E., Pierce, J. R. and Earhart, C. F. (1988) J. Gen. Microbiol. **134** 1355–1367
- Johnson, B. F. (1977) Genet. Res. **30** 273–286
- Meselson, M. (1968) Nature (London) **217** 1110–1114
- Bradford, M. M. (1976) Anal. Biochem. **72** 248–254
- Mao, C.-H., Blocker, J. P., Anderson, L. and Smith, D. C. (1965) Phytochemistry **4** 297–299
- Peller, L. and Alberty, R. A. (1959) J. Am. Chem. Soc. **81** 5907–5914
- Leatherbarrow, R. J. (1990) Grafit Version 2.0, Erithacus Software Ltd., Staines
- Griffin, T. A., Wynn, R. M. and Chuang, D. T. (1990) J. Biol. Chem. **265** 12104–12100
- Creighton, T. E. (1992) Protein Folding, pp. 519–523, W. H. Freeman Co., New York
- Kitz, R. and Wilson, I. B. (1962) J. Biol. Chem. **237** 3245–3251