ELSEVIER

Journal of Chromatography A, 910 (2001) 275–284

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Genetically engineered peptide fusions for improved protein partitioning in aqueous two-phase systems
Effect of fusion localization on endoglucanase I of *Trichoderma reesei*

Anna Collén[a], Michael Ward[b], Folke Tjerneld[a], Henrik Stålbrand[a,*]

[a]*Department of Biochemistry, Lund University, P.O. Box 124, S-221 00 Lund, Sweden*
[b]*Genencor International, 925 Page Mill Road, Palo Alto, CA 94304, USA*

Received 1 September 2000; received in revised form 27 November 2000; accepted 27 November 2000

Abstract

Genetic engineering has been used for fusion of the peptide tag, Trp–Pro–Trp–Pro, on a protein to study the effect on partitioning in aqueous two-phase systems. As target protein for the fusions the cellulase, endoglucanase I (endo-1,4-β-D-glucan-4-glucanohydrolase, EC 3.2.1.4, EGI, Cel7B) of *Trichoderma reesei* was used. For the first time a glycosylated two-domain enzyme has been utilized for addition of peptide tags to change partitioning in aqueous two-phase systems. The aim was to find an optimal fusion localization for EGI. The peptide was (1) attached to the C-terminus end of the cellulose binding domain (CBD), (2) inserted in the glycosylated linker region, (3) added after a truncated form of EGI lacking the CBD and a small part of the linker. The different constructs were expressed in the filamentous fungus *T. reesei* under the *gpdA* promoter from *Aspergillus nidulans*. The expression levels were between 60 and 100 mg/l. The partitioning behavior of the fusion proteins was studied in an aqueous two-phase model system composed of the thermoseparating ethylene oxide (EO)–propylene oxide (PO) random copolymer EO–PO (50:50) ($EO_{50}PO_{50}$) and dextran. The Trp–Pro–Trp–Pro tag was found to direct the fusion protein to the top $EO_{50}PO_{50}$ phase. The partition coefficient of a fusion protein can be predicted with an empirical correlation based on independent contributions from partitioning of unmodified protein and peptide tag in this model system. The fusion position at the end of the CBD, with the spacer Pro–Gly, was shown to be optimal with respect to partitioning and tag efficiency factor (TEF) was 0.87, where a fully exposed tag would have a TEF of 1.0. Hence, this position can further be utilized for fusion with longer tags. For the other constructs the TEF was only 0.43 and 0.10, for the tag fused to the truncated EGI and in the linker region of the full length EGI, respectively.

Keywords: Aqueous two-phase systems; *Trichoderma reesei*; Peptides; Endoglucanases; Enzymes; Proteins

1. Introduction

Aqueous two-phase systems have widespread use in biochemistry and biotechnology for the purification of biological materials [1]. An aqueous two-phase system consists of water and two-phase form-

*Corresponding author. Tel.: +46-46-2228-202; fax: +46-46-2224-534.

E-mail address: henrik.stalbrand@biokem.lu.se (H. Stålbrand).

0021-9673/01/$ – see front matter
PII: S0021-9673(00)01212-7

ing components, e.g., polyethylene glycol (PEG) together with dextran. The phase forming components separate into two phases above certain concentrations and each phase is enriched in one of the polymers [2,3]. The partitioning of proteins in aqueous two-phase systems is mainly determined by the surface properties of the protein such as hydrophobicity, net charge and size. Additionally, the properties of the system are important, such as polymer hydrophobicity, molecular mass and concentration.

The advantages of aqueous two-phase extraction are high capacity, volume reduction, rapid separation and ease to scale up which have lead to several industrial applications [4]. However, there is a need for further development of this technology in terms of increasing the selectivity in extraction of target protein. Genetically engineered peptide tags containing tryptophan (Trp) residues, have previously been shown to strongly enhance the partitioning of fusion proteins to ethylene oxide–propylene oxide random copolymer (EO–PO) and PEG phase [5–9].

In the current study the effect on partitioning of different fusion points on a two-domain protein was for the first time investigated. The target protein, endoglucanase I (EGI, also designated Cel7B) (endo-1,4-β-D-glucan-4-glucanohydrolase, EC 3.2.1.4) of *Trichoderma reesei* has an amino terminal large catalytic domain naturally fused with a small family one cellulose binding domain (CBD) [10]. Like many other fungal cellulases the two domains are linked by an *O*-glycosylated linker region [11]. Cellulases are industrially important with applications, e.g., in textile and pulp and paper industries [12,13]. Our aim was to find an optimal fusion position both in terms of partitioning in aqueous two-phase system and expression in the filamentous fungus, *T. reesei*. The conclusions from this initial study can further be utilized for construction of longer tags and selection and optimization of two-phase systems suitable for use in large-scale separation.

The two-phase system used in this study was composed of a random EO–PO copolymer of 50% ethylene oxide and 50% of propylene oxide ($EO_{50}PO_{50}$) and dextran. The EO–PO–dextran system is a suitable model system to use to study tag efficiency [5]. The partition coefficient of a fusion protein can be predicted with an empirical correlation based on independent contributions from partitioning of unmodified protein and peptide tag in this model system. EO–PO copolymers are thermoseparating, i.e., above a critical temperature the solubility of the polymer will decrease and a system composed of a water and a polymer phase is formed [14,15]. After thermoseparation the EO–PO copolymer can be separated from the target protein which is partitioned to the water phase and the EO–PO copolymer can be recycled. The system is suitable for large scale use where the dextran is replaced by a starch polymer [15].

To modify cellulases expressed in *T. reesei* by genetic engineering to improve separation is a novel approach. This is the first time fungal expression is used for expression of fusion proteins carrying a tryptophan tag. The promoter used, the *Aspergillus nidulans gpdA* promoter, has previously been utilized for heterologous expression on glucose-containing medium in *T. reesei* [16]. Under these conditions endogenous production of cellulases is suppressed [17].

2. Materials and methods

2.1. Construction of the fusion proteins

Recombinant DNA manipulations were performed according to Ref. [18]. Oligonucleotide primers were purchased from the Biomolecular Resource Facility, Lund University, Lund, Sweden. In all cloning procedures the *Escherichia coli* strain XL-1 blue was used. The vectors pLE3, pMVQ and pTOC 202 were kindly provided by Dr. Qiao Mingqiang, VTT Biotechnology and Food Research, Espoo, Finland. The plasmid pLE3 was digested with *Ksp*I and *BamH*I to release the *egl1* cDNA and ligated into the vector pMVQ, which was vector was further used for mutagenesis for the different constructs. The mutagenesis was performed by polymerase chain reaction (PCR) using the concept of Quick Change Site-Directed Mutagenesis (Stratagene, La Jolla, CA, USA). Native *Pfu* polymerase was used. The PCR reactions were optimized and run with an initial denaturing step of 95°C for 30 s, followed by 18 repetitive cycles of a denaturing step (30 s), anneal-

Table 1
Sequences of oligonucleotides used in the cloning of EGI constructs (sequence encoding fusion tag in bold)

Construct	Sequence (5′→3′)	Restriction enzyme	Annealing temperature (°C)
EGI	GCTACCCCGCTTGAGCAGCCGCGGCCATGGATCCACTTAACG	*Ksp*I	65
$EGI_{core}WPWP_{CBD}$	CGACTACACGGAGGAGC**TGGCCATGGCCG**AGCAGCCCGAGCTGC	*Nco*I	68
$EGI_{coreCBD}WPWP$[a]	CTACTCGCAATGCCTT**CCGGGCTAG**GACTTGCCTCTGGTCTG	*Bgl*I	65
$EGI_{coreCBD}WPWP$[b]	CGCAATGCCTTCCGGC**TGGCCATGGCCC**TAGGACTTGCCTCTTG	*Nco*I	68
$EGI_{core}WPWP$	CGACTACACGGAGGAGC**TGGCCATGGCCG**TAGACATTGGGC	*Nco*I	64

[a] Introduction of Pro–Gly–STOP.
[b] Introduction of Trp–Pro–Trp–Pro.

ing step using different temperatures depending on the primer pair (Table 1) and elongation at 72°C for 16 min. For PCR 100–500 ng of DNA, 125 ng of each primer, 0.6 m*M* of deoxyribonucleotide triphosphates (dNTPs) and *Pfu* polymerase buffer were used. After amplification the template was digested with *Dpn*I for at least 6 h.

Four different constructs were made (Fig. 1). The N-terminus was not used as a fusion point due to its localization close to the catalytic cleft [19] and additionally this location might affect secretion when it is situated directly after the signal peptide. The construct with Trp–Pro–Trp–Pro at the C-terminal of EGI ($EGI_{coreCBD}WPWP$) was performed in two steps. First, the linker Pro–Gly and a stop codon

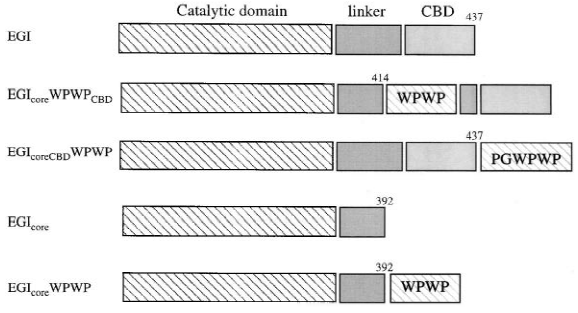

Fig. 1. Schematic picture of the different constructs: EGI is the full-length EGI, 437 amino acid residues long; $EGI_{core}WPWP_{CBD}$ is EGI with Trp–Pro–Trp–Pro (WPWP) in the linker region with the changes S415W, T416P, T417W and S418P (437 residues); $EGI_{coreCBD}WPWP$ is the EGI with the addition of Pro–Gly–Trp–Pro–Trp–Pro at the end of the CBD (443 residues long); EGI_{core} is a truncated form of EGI (392 residues) with the CBD and a part of the linker deleted; $EGI_{core}WPWP$ is EGI_{core} with the addition of Trp–Pro–Trp–Pro at the C-terminus (396 residues). The catalytic domain is between amino acid positions 1–371 [19], the linker between 372 and 399 and the CBD between 400 and 437 [37].

were introduced followed by addition of Trp–Pro–Trp–Pro between the codon for glycine and the stop codon. Addition of Trp–Pro–Trp–Pro was performed in the second step.

The construct with a stop codon introduced at amino acid position 393 (EGI_{core}) was performed by digestion of *BamH*I and *Hind*III. A 3-kilobase pair (kbp) fragment containing the *gpdA* promoter from *T. reesei* and the *egl1* cDNA was obtained after digestion and cloned into the vector pBR 322 (New England Biolabs, Beverly, CA, USA). This vector was cut with *Sac*I (internal in the *egl1* and at a position in the vector down stream *egl1*) and religated after isolation of the large 7-kbp fragment. The new ligation junction forms the stop codon. The fragment containing the truncated *egl1* cDNA was ligated back to the start vector between the *BamH*I and *Hind*III sites. A stop codon (TAG) was now constructed. This deletion construct was used for introduction of Trp–Pro–Trp–Pro in the linker region of truncated EGI ($EGI_{core}WPWP$).

The vector pAN52-1 [20], containing the *gpdA* promotor from *A. nidulans* was used for construction of the expression vector pAC1. A *Ksp*I site was introduced upstream of the start codon to facilitate cloning. All the different constructs were ligated into the expression vector between the *Ksp*I site and the *Bam*HI site. Due to the introduction of a restriction site the nucleotide sequence upstream of the start codon was changed to 5′-CCG CGG ACT GGC ATC ATG-3′.

Subsequently, the *egl1* cDNA constructs were confirmed by sequencing before transformation into *T. reesei*. Four different primers were used to be able to sequence the complete *egl1* cDNA; 5′-CCT TTC AGT TCG AGC TTT CC-3′, 5′-GGA GCT GAG

CTT CGA CG-3′, 5′-CGT GTT CAG CAT TTG GAA CG-3′ and 5′-TAT TAA GCC CAA TGT CTA CG-3′. DNA sequencing was performed at the Biomolecular Resource Facility, Lund University.

2.2. Trichoderma reesei transformation

The *T. reesei* strain QM 9414 was used as transformation host (obtained from Dr. Merja Penttilä, VTT Biotechnology and Food Research). For protoplast preparation, a spore suspension of *T. reesei* QM 9414 was spread and cultivated on plates containing PD agar (3.9% potato dextrose agar; Difco, Franklin Lakes, NJ, USA) for 4 days at 30°C. A 1-l baffled shake-flask containing 100 ml YEG medium (0.5% yeast extract and 2% glucose) was inoculated with a 1 cm^2 PD agar piece with mycelia and was incubated for 24 h at 30°C at 180 rpm. The mycelia was harvest by centrifugation at 4°C at 2000 *g* for 10 min and washed three times in 1.2 *M* $MgSO_4$, 10 m*M* sodium phosphate, pH 5.8. The pellet was resuspended in the same buffer as above with the addition of 5 mg/ml of Novozym 234 (catalogue 0412-1, Inter Spex Products, Foster City, CA, USA) and thereafter incubated at 28°C for approx. 1.5 h at 150 rpm. Transformation and selection was then carried out according to Ref. [21]. The constructs were all cotransformed into the *T. reesei* QM9414 strain together with the separate plasmid pTOC 202 carrying the *A. nidulans amdS* gene (molar rate selection vector:mutant vector 1:5). The obtained transformants were selected on the basis of acetamide utilization. The transformants were purified through single spores and were tested for EGI production by activity measurements (see Section 2.6) to identify the best producers.

2.3. Expression

The mutant strains were fermented in 1-l baffled shake-flasks in minimal medium with glucose as sole carbon source at pH 6.0 according to Ref. [22] with a few modifications, 30 g/l K_2HPO_4, 8 g/l KH_2PO_4 and 4% glucose. One-liter flasks contained 120 ml medium, which was inoculated with mycelia. The glucose concentration was monitored daily by measuring reducing sugar by dinitrosalicylic acid (DNS) reagent [23] and kept above 1%. After 4–5 days of cultivation at 30°C and 180 rpm, mycelia were removed by filtration through a glass filter (porosity 1) and buffer was changed by ultrafiltration (Minisette 10K, from Pall Gelman and Science, Ann Arbor, MI, USA) to 50 m*M* sodium acetate, pH 5.0.

2.4. Fusion proteins

The culture filtrates from the different constructs were run by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and further blotted for Western blotting. Transfer to a poly(vinylidene difluoride) (PVDF) membrane (PV4HY00010; Micron Separations, now Pall, East Hills, NY, USA) and antibody binding was performed according the manufacturer's instruction. Monoclonal anti-EGI antibodies (kindly provided by Dr. Merja Penttilä) were used at 1:5000 dilution, with alkaline phosphatase conjugated goat anti-mouse IgG (Bio-Rad, Hercules, CA, USA) as secondary antibody (1:1500 dilution). Detection was performed with the amplified alkaline phosphatase Immunoblot assay with enhanced chemifluorescence substrate (catalogue RPN5785; Amersham Pharmacia Biotech, Uppsala, Sweden) and scanned with a STORM 860 (Molecular Dynamics, Sunnyvale, CA, USA). The SDS gels were silver stained using a Novex silver staining system and were scanned with a densitometer (Personal Densitometer SI; Molecular Dynamics). The expression levels were estimated on the basis of enzyme activity measurements and by a plate assay with carboxymethylcellulose (CMC). A 100-μl volume of culture filtrate was incubated at 37°C for 1 h, stained with 0.1% (w/v) Congo red solution overnight and destained with 1 *M* sodium chloride for 2–3 h [24]. Pure EGI was used as standard.

2.5. Two phase diagram

The bottom-phase polymer dextran T500, with a molecular mass of 500 000, was obtained from Amersham Pharmacia Biotech. The top-phase polymer, Breox PAG 50A1000 ($EO_{50}PO_{50}$), a random copolymer (M_r 3900) was supplied from International Speciality Chemicals (Southampton, UK). The phase diagram for the aqueous two-phase system of

Breox–dextran–water was determined according to Ref. [5].

2.6. Two phase partitioning

Two-phase systems of a total mass of 5 g, containing 7.1% dextran and 6.8% Breox PAG 50A1000 were prepared from 25% polymer stock solution of dextran and pure Breox (liquid). All polymer concentrations are given in % (w/w). All systems were buffered to pH 5.0 in 50 m*M* sodium acetate buffer. The protein concentration in the systems was between 0.2 and 0.5 mg/ml.

The partitioning partition coefficient *K* is defined as the ratio between the enzyme concentration in the top phase and the enzyme concentration in the bottom phase. All partition coefficients are average values from at least duplicate experiments from at least two different fermentations. The systems were equilibrated at 21°C and separated into two phases by centrifugation at 1600 *g* for 3 min. The EGI activity was assayed with 3 m*M* *p*-nitrophenyl-β-D-cellobioside (pNPC) from Sigma (St. Louis, MO, USA) as substrate [25]. In order to correct for possible interference of the polymers on the assay and/or enzymatic activity, a blank system was made without protein. The blank system was used to add top and bottom phases in equal amounts in all the activity measurements in order to have the same polymer composition in all the assays.

The peptide Trp–Pro–Trp–Pro (Synpep, Dublin, CA, USA) was partitioned in the same system as above and added in the form of pre-made stock solution, giving a final concentration of 0.01 mg/g. The peptide concentration was determined from the absorbance at 280 nm corrected by subtracting the absorbance at 320 nm. The absorbance contribution from the polymers was taken into account by preparing a blank system. The peptide partition coefficient is an average value from duplicate experiments.

2.7. Computer analysis (GRASP)

The computer program Graphical Representation and Analysis of Surface Properties (GRASP) [26] was used to study protein surfaces.

3. Results and discussion

3.1. Mutagenesis and T. reesei transformations

An expression vector encoding EGI of *T. reesei* under the regulation of the *gpdA* promoter from *A. nidulans* was constructed as described in Materials and methods. The *egl1* gene was mutagenised by PCR. Three EGI mutants with a tryptophan-rich peptide fused at different localizations were constructed. A scheme of these mutants is shown in Fig. 1. A mutant consisting of the catalytic EGI_{core} domain was constructed to be used as a reference. The DNA-constructs were transformed into *T. reesei* QM 9414. 50 Amd^+ transformants of each construct were tested for expression of EGI. Between 50 and 80% of the screened colonies, depending on the different constructs transformed, released active EGI into the culture medium.

3.2. Protein expression

EGI and the mutant with the fused peptide Pro–Gly–Trp–Pro–Trp–Pro at the C-terminal end of the CBD ($EGI_{coreCBD}$WPWP) were expressed to levels of 94 and 89 mg/l, respectively as determined by activity measurements. In a silver stained SDS-polyacrylamide gel the EGI-derived proteins were seen as the most prominent bands (gel not shown). With the CMC plates an approximate expression level of 100 mg/l was estimated, which corresponds well with the activity measurements. The truncated form of EGI (EGI_{core}) (60 mg/l), the form with Trp–Pro–Trp–Pro added after the truncated EGI (EGI_{core}WPWP) (66 mg/l) and the mutant containing Trp–Pro–Trp–Pro in the linker region ($EGI_{core}WPWP_{CBD}$) (75 mg/ml) all showed lower expression levels. The *gpdA* promotor has previously only been used for heterologous *T. reesei* expression of single-chain antibodies, where the expression level was approximately 100 times lower than in the current study. For homologous expression in *T. reesei* on glucose medium other promoters, such as the cDNA1, *tef1* [27], *pgk1* [28] promoters and a modified *cbh1* promoter [29] have been used with varying success. The highest levels were obtained with the promoter cDNA1 which gave an expression of at most 53 mg/l for EGI core and CBHI [22]. The

expression level achieved in the current study compares well to this value. The Western blot analysis showed only one band estimated to be approximately of the expected size for each mutant (Fig. 2), which is an indication of intact protein.

3.3. Phase diagram

The Breox PAG 50A1000 ($EO_{50}PO_{50}$)–dextran–water phase diagram was determined (Fig. 3) and the system used for partitioning is indicated. Breox PAG 50A1000 has the same composition ($EO_{50}PO_{50}$) as the polymer Ucon 50-HB-5100 for which the phase diagram Ucon 50-HB-5100–dextran previously has been determined [14]. However, Breox PAG 50A1000 has a lower average molecular mass (M_r 3900) in comparison with Ucon 50-HB-5100 (M_r 4000), which influenced the phase diagram towards phase separation at slightly higher polymer concentrations. The critical point in the phase diagram is defined as the transition point between one phase and two phases of equal volumes. The critical point in the Breox–dextran system was determined at Breox 4.0% (w/w) and dextran 4.3% (w/w), compared with Ucon 3.3% (w/w) and dextran 4.0% (w/w) in the Ucon–dextran system.

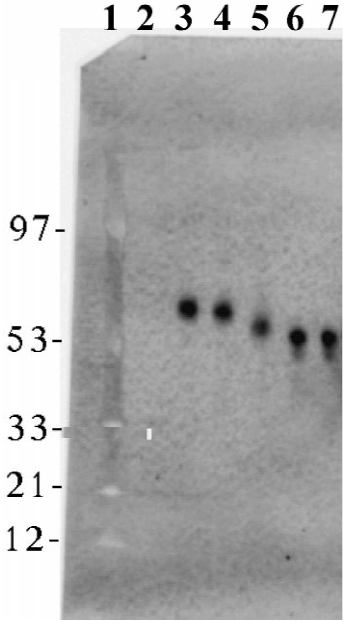

Fig. 2. Western blot of the different constructs. The estimated M_r for each construct given in parentheses. Lanes: 1 = M_r markers (12 000–97 000); 2 = culture filtrate QM 9414; 3 = full-length EGI (62 000); 4 = EGI with the addition of Pro–Gly–Trp–Pro–Trp–Pro at the end of the CBD (63 000); 5 = EGI with Trp–Pro–Trp–Pro (WPWP) in the linker region (58 000); 6 = the truncated form of EGI after amino acid residue 392 (54 000); 7 = EGI_{core}WPWP is EGI_{core} with the addition of Trp–Pro–Trp–Pro at the C-terminus (54 000).

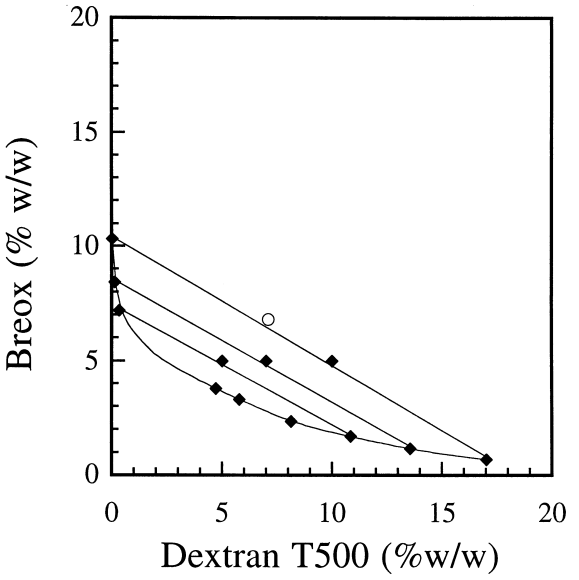

Fig. 3. Phase diagram of Breox PAG 50A1000–dextran T500–water at 21°C. For polymer concentrations above the binodial curve, the system will separate into two phases. The tie lines show the polymer concentrations in the two phases at equilibrium. The composition of the system used is indicated (○).

3.4. Peptide and unmodified protein partitioning

Tryptophan and peptides containing tryptophan residues have earlier been found to partition towards the EO–PO copolymer phase [30,31] and the PEG phase [6,32–34]. To direct the partitioning to an EO–PO or PEG phase for a specific protein, a tryptophan-rich peptide can be genetically fused to the target protein. A two-polymer system has been used to study the effect of tryptophan fusions to target proteins, an $EO_{30}PO_{70}$–dextran system [5,7].

In this study the peptide Trp–Pro–Trp–Pro was fused to EGI to increase the partitioning to the more hydrophobic EO–PO phase in an $EO_{50}PO_{50}$–dextran system. The peptide was designed to give optimal

solvent exposure of the tryptophans by avoiding secondary structure conformation. This was performed by introducing prolines which are well known to disrupt secondary structure [35,36]. Proline residues introduced between tryptophan residues have previously been shown not to affect the partitioning [7], which makes them suitable to use. Previously, a *K* value of 1.7 was obtained for the Trp–Pro–Trp–Pro peptide in the $EO_{30}PO_{70}$–dextran system [7]. The partition coefficient for the same peptide was 2.0 in the current study, with $EO_{50}PO_{50}$ as top phase polymer (Fig. 4). Increasing the propylene oxide content in the polymer increases the hydrophobicity, which thus explains the slightly higher *K* value for the tryptophan peptide compared with $EO_{50}PO_{50}$ in the system.

The *K* value for the full-length EGI was 2.9 and a yield of 81% in the EO–PO phase was obtained calculated on the basis of activity. Most of the water soluble proteins prefer the more hydrophilic phase, in this case, the dextran phase. Exposure studies were performed with the computer program GRASP based on structure determination in Refs. [19,37].

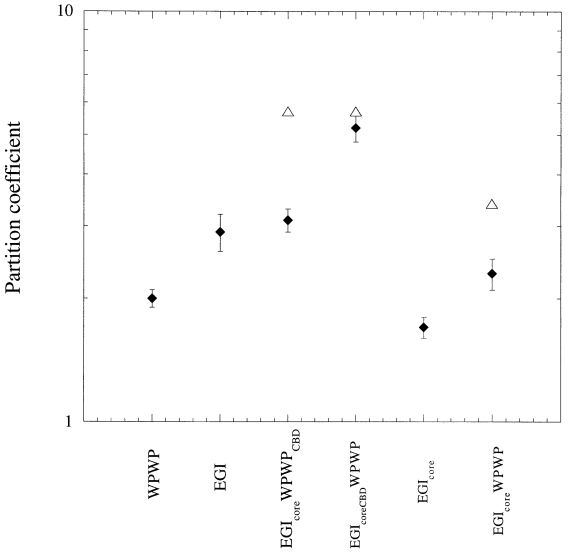

Fig. 4. Plot of the *K* values obtained from partitioning in the Breox PAG 50A1000 (6.8%) and dextran T500 (7.1%) system at 21°C, in 50 m*M* NaAc at pH 5.0. The peptide WPWP and all different constructs EGI, $EGI_{core}WPWP_{CBD}$, $EGI_{coreCBD}WPWP$, EGI_{core} and $EGI_{core}WPWP$ were partitioned and the experimental *K* values marked with ♦. The calculated *K* values by use of Eq. (1) are marked with △.

EGI contains seven tryptophan residues and five of them are partly exposed on the protein surface. Additionally, EGI has 14 tyrosine and two phenylalanine residues exposed on the surface. All these residues are also known to influence the partitioning towards the EO–PO phase in the $EO_{30}PO_{70}$–dextran system [5] and can be classified as relatively hydrophobic when partitioned in the EO–PO–water system [30].

The deletion of the cellulose binding domain and part of the linker (EGI_{core}) lowered the partition coefficient to 1.7 in comparison to the full-length EGI ($K=2.9$) (Fig. 4). The CBD has one tryptophan and three tyrosine residues exposed to the solvent [37]. Thus, deletion of the CBD will increase partitioning to the more hydrophilic dextran phase. The observed effect can be explained by reduction in exposed hydrophobic amino acid residues.

3.5. Effect on partitioning of the localization of the fused peptide

A clear difference in partitioning due to fusion point was observed (Fig. 4). The insertion of Trp–Pro–Trp–Pro into the linker region is in theory an attractive position since the fusion point is between two functional domains and thus should not interfere with the catalysis or substrate binding. Only a small effect on the partitioning was observed for the $EGI_{core}WPWP_{CBD}$ construct. The yield in the EO–PO phase was 80%. The linker region of EGI is heavily glycosylated. It is reasonable to assume that the fused peptide is not exposed to the solvent, which would explain the weak effect obtained.

An increase in partition coefficient to 5.1 was obtained when the peptide Pro–Gly–Trp–Pro–Trp–Pro was added to the C-terminus of the full-length EGI. A significant effect of the addition of only two trytophans to the comparably large, two domain protein EGI was observed. The yield was determined to 90% in the EO–PO phase on the basis of activity. This was further confirmed by estimation of the yield with a silver stained SDS–PAGE gel (Fig. 5). The purity of the target protein in the top phase was very high since no other protein bands than EGI were observed. All other proteins partitioned to the dextran phase. In this construct the amino acid residues Pro–Gly were added after the CBD as a spacer

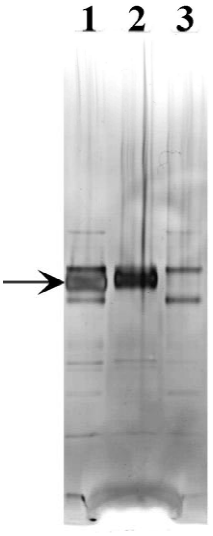

Fig. 5. SDS-gel analysis of the partitioning of EGI with Pro–Gly–Trp–Pro–Trp–Pro added at the CBD. The samples have been volume adjusted so that a direct visual comparison can be made between the lanes. Hence, if 100% of the target protein is recovered, the band of the sample should have the same intensity as in the starting material. The arrow indicates the position of $EGI_{coreCBD}$WPWP. Lanes: 1=the culture filtrate from $EGI_{coreCBD}$WPWP cultivation; 2=the proteins in the $EO_{50}PO_{50}$ phase; 3=the proteins in the dextran phase.

before the fused peptide to enhance tag exposure. The *T. reesei* cellulases CBHII (Cel6A) [38] and EGII (Cel5A) [39] carry N-terminal CBDs which are homologous to the EGI CBD. The C-termini of the CBDs are connected to the linker region with the amino acid residues Pro–Gly. Thus, this spacer is not believed to interfere with the three-dimensional structure of the CBD.

The addition of Trp–Pro–Trp–Pro to the EGI_{core} after a stretch of the linker was included in the study since the linker region might function as an efficient spacer for tag exposure. The partition coefficient for EGI_{core}WPWP increased from 1.7 to 2.3 relative to EGI_{core} (Fig. 4) and the yield in the EO–PO phase improved from 68 to 75%.

3.6. Empirical correlation for fusion protein partitioning

For the recombinant proteins ZZTn [5] and ZZ-cutinase [7] a fused peptide containing two trytophan residues was shown to independently contribute to the partition coefficient. Eq. (1) was used as an empirical correlation for the partitioning data of fusion proteins to evaluate the efficiency of the fusion [5]:

$$\log K^{\text{calc.}}_{\text{tagged protein}} = \log K_{\text{protein}} + \log K_{\text{tag}} \quad (1)$$

The experimental log K values for the tag and unmodified protein were determined. The sum log K_{wpwp} + log K_{EGI} was calculated (log $K^{\text{calc.}}_{\text{tagged protein}}$) to evaluate the effect of the added peptides in the fusion proteins.

The calculated sum for the construct with Trp–Pro–Trp–Pro after the CBD ($EGI_{coreCBD}$WPWP) corresponded to the experimental value within the standard deviations (Fig. 4). The tag efficiency factor (TEF) can be calculated from Eq. (2) [6] using the experimental log K values:

$$\text{TEF} = \frac{\log K_{\text{tagged protein}} - \log K_{\text{protein}}}{\log K_{\text{tag}}} \quad (2)$$

Thus, nearly additive effect, i.e., TEF=0.87 was obtained with the fusion of WPWP to EGI ($EGI_{coreCBD}$WPWP) (Table 2). This strongly suggest complete exposure of the tag in the fusion protein.

A discrepancy between the experimental and calculated values was obtained for the other fusion proteins. Significantly higher log $K^{\text{calc.}}_{\text{tagged protein}}$ values were obtained compared with the experimental log K values for both $EGI_{core}WPWP_{CBD}$ and EGI_{core}WPWP (Fig. 4). For the tag in the linker the TEF was only 0.1 and the tag fused to EGI_{core} gave a TEF of 0.43 (Table 2). A lower tag effect is expected if the exposure of the peptide tag to the solvent is less in the tag fused to the protein than in the free peptide. Less exposure can, however, in these cases be due to the glycosylation close to the fusion position. Thus, a fusion on the C-terminus of the CBD is more likely to ensure full exposure of the tag and optimal tag effect.

4. Conclusions

In this study it has been shown that minor changes in amino acid sequence such as the addition of the

Table 2
Experimental and calculated *K* values of the constructs with different tag localization (tag efficiency factor, TEF, was calculated from the experimental partition coefficients)

Construct	Experimental *K* value	Calculated *K* value[a]	TEF[b]	Yield (%)[c]
$EGI_{core}WPWP_{CBD}$	3.1	5.7	0.10	80
$EGI_{coreCBD}WPWP$	5.2	5.7	0.87	90
$EGI_{core}WPWP$	2.3	3.4	0.43	75

[a] Calculated using Eq. (1).
[b] Calculated using Eq. (2).
[c] Recovered in the top EO–PO phase. Yield for EGI_{core}: 68% and wild-type EGI: 81%.

tag Trp–Pro–Trp–Pro affect the partition behavior of a two-domain protein, the cellulase EGI. This target protein is larger than previously studied fusion proteins with peptide tags and it is also more complex due to its two-domain structure and its glycosylation. Three different fusion points of the tag were tested in order to find the most optimal for fusion both in respects to partitioning and expression. All the constructs were successfully expressed in the fungus *T. reesei* under the *gpdA* promoter from *A. nidulans*. The expression levels were 60–100 mg/l. The fusion point of the peptide affected the effect of the tag in respect to partitioning. The optimal position for fusion was shown to be the C-terminus of the EGI, at the end of the CBD with the spacer Pro–Gly ($EGI_{coreCBD}WPWP$), where the TEF was 0.87. Thus, the improvement in partitioning by the tag could be quantitatively calculated from peptide and protein partitioning, which indicates the exposure of the tag when positioned at the end of the CBD. Hence, this position can further be utilized for fusion with longer tags anticipating more extreme partitioning.

Acknowledgements

We are grateful to Dr. Qiao Mingqiang, VTT, Finland for providing us with the vectors pLE3, pMVQ and pTOC202. We also wish to acknowledge Dr. P. Punt, TNO, for the kind gift of the plasmid pAN52. Further, we want to thank Dr. Merja Penttilä, VTT, Finland for useful discussions and monoclonal EGI antibodies. The work was supported by grants from the EU project BIO4-CT96-0435. The Swedish Research Council for Engineering Sciences (TFR) and Carl Tryggers Research Foundation are gratefully thanked for grants to H.S.

References

[1] H. Walter, G. Johansson (Eds.), Methods Enzymol. 228 (1994).
[2] P.-Å. Albertsson, Partition of Cell Particles and Macromolecules, 3rd ed., Wiley, New York, 1986.
[3] H. Walter, D.E. Brooks, D. Fisher, Partitioning in Aqueous Two-Phase Systems, Academic Press, New York, 1985.
[4] M. Kula, K. Selber, in: M. Flicker, S. Drew (Eds.), Encyclopedia of Bioprocess Technology – Fermentation, Biocatalysis and Bioseparation, Wiley, New York, 1999, p. 2179.
[5] K. Berggren, A. Veide, P.-Å. Nygren, F. Tjerneld, Biotechnol. Bioeng. 62 (1999) 135.
[6] K. Berggren, F. Tjerneld, A. Veide, Bioseparation 9 (2000) 69.
[7] K. Berggren, A. Nilsson, G. Johansson, N. Bandman, P.-Å. Nygren, F. Tjerneld, J. Chromatogr. B 743 (2000) 295.
[8] N. Bandman, E. Collet, J. Leijen, M. Uhlén, A. Veide, P.-Å. Nygren, J. Biotechnol. 79 (2000) 161.
[9] K. Köhler, A. Veide, S.-O. Enfors, Enz. Microb. Technol. 13 (1991) 204.
[10] M. Penttilä, P. Lehtovaara, H. Nevalainen, R. Bhikhabhai, J. Knowles, Gene 45 (1986) 253.
[11] P. Tomme, H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhoven, J. Knowles, T. Teeri, M. Claeyssens, Eur. J. Biochem. 170 (1988) 575.
[12] Y.M. Galante, A. De Conti, R. Monteverd, in: G.E. Harman, C.P. Kubicek (Eds.), *Trichoderma* and *Gliocladium*, Vol. 2, Taylor and Francis, London, 1998, p. 311.
[13] J. Buchert, T. Oksanen, J. Pere, M. Siika-Aho, A. Suurnäkki, L. Viikari, in: G.E. Harman, C.P. Kubicek (Eds.), *Trichoderma* and *Gliocladium*, Vol. 2, Taylor and Francis, London, 1998, p. 343.
[14] P. Harris, G. Karlström, F. Tjerneld, Bioseparation 2 (1991) 237.
[15] J. Persson, L. Nyström, H. Agerland, F. Tjerneld, J. Chromatogr. B 711 (1998) 97.

[16] E. Nyyssönen, Ph.D. Thesis, University of Helsinki, Helsinki, 1993.
[17] M. Ilmén, A. Saloheimo, M.-J. Onnela, M. Penttilä. Appl. Environ. Microbiol. 63 (1997) 1298.
[18] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[19] G. Kleywegt, J.-Y. Zou, C. Divne, G. Davies, I. Sinning, J. Ståhlberg, T. Reinikainen, M. Srisodsuk, T. Teeri, A. Jones, J. Mol. Biol. 272 (1997) 383.
[20] P. Punt, M. Dingemanse, A. Kuyvenhoven, R. Soede, P. Pouwels, C. van den Hondel, Gene 93 (1990) 101.
[21] M. Penttilä, H. Nevalainen, M. Rättö, E. Salminen, J. Knowles, Gene 61 (1987) 155.
[22] T. Nakari-Setälä, M. Penttilä, Appl. Environ. Microbiol. 61 (1995) 3650.
[23] J. Sumner, G. Somers, in: Laboratory Experiments in Biological Chemistry, Academic Press, New York, 1949, p. 38.
[24] H. Wang, R.W. Jones, Gene 158 (1995) 125.
[25] M.V. Deshpande, L.G. Pettersson, K.-E. Eriksson, Methods Enzymol. 160 (1988) 126.
[26] A. Nicholls, K. Sharp, B. Honig, Proteins 11 (1991) 281.
[27] T. Nakari-Setälä, Ph.D. Thesis, University of Helsinki, Helsinki, 1995.
[28] S. Vanhanen, Ph.D. Thesis, VTT Publications 75, Espoo, 1991.
[29] T. Nakari, M.L. Onnela, M. Ilmén, K.H.M. Nevalainen, M. Penttilä, in: P. Suominen, T. Reinikainen (Eds.), Proceedings of the 2nd TRICEL Symposium on *Trichoderma reesei* Cellulases and Other Hydrolases, Espoo, Foundation for Biotechnical and Industrial Fermentation Research, 1993, p. 239.
[30] H.-O. Johansson, G. Karlström, B. Mattiasson, F. Tjerneld, Bioseparation 5 (1995) 269.
[31] H.-O. Johansson, G. Lundh, G. Karlström, F. Tjerneld, Biochim. Biophys. Acta 1290 (1996) 289.
[32] A. Diamond, X. Lei, J. Hsu, Biotechnol. Tech. 3 (1989) 271.
[33] M. Eiteman, J. Gainer, Biotechnol. Prog. 6 (1990) 479.
[34] C. Hassinen, K. Köhler, A. Veide, J. Chromatogr. A 668 (1994) 121.
[35] T. Creighton, in: Proteins, Structure and Molecular Properties, 2nd ed., Freeman, New York, 1993, p. 226.
[36] C.R. Cantor, P.R. Schimmel, in: Biophysical Chemistry, Part I: The Conformation of Biological Macromolecules, Freeman, New York, 1980, p. 272.
[37] M.-L. Mattinen, M. Linder, T. Drakenberg, A. Annila, Eur. J. Biochem. 256 (1998) 279.
[38] T. Teeri, P. Lehtovaara, S. Kauppinen, I. Salovuori, Gene 51 (1987) 43.
[39] M. Saloheimo, P. Lehtovaara, M. Penttilä, T. Teeri, J. Ståhlberg, G. Johanson, G. Petterson, M. Claeyssens, P. Tomme, J. Knowles, Gene 63 (1988) 11.