Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 46 (2006) 179–188
Protein Expression & Purification
www.elsevier.com/locate/yprep

Production and purification of self-assembling peptides in *Ralstonia eutropha*

David C. Reed, Gavin C. Barnard, Errik B. Anderson, Linden T. Klein, Tillman U. Gerngross *

Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755, USA

Received 16 June 2005, and in revised form 25 August 2005
Available online 22 September 2005

Abstract

Self-assembling peptides have emerged as an attractive scaffold material for tissue engineering, yet the expense associated with solid phase chemical synthesis has limited their broad use. In addition, the fidelity of chemical synthesis constrains the length of polypeptides that can be produced homogeneously by this method. Template-derived biosynthesis by recombinant DNA technology may overcome both of these problems. However, recovery of polypeptides from recombinant protein expression systems typically involves multi-step purification schemes. In this study, we report an integrated approach to recombinantly produce and purify self-assembling peptides from the recently developed expression host *Ralstonia eutropha*. The purification is based on the specific affinity of carbohydrate binding modules (CBMs) to cellulose. In a first step, we identified CBMs that express well in *R. eutropha* by assembling a fusion library of green fluorescent protein (GFP) and CBMs and determining the fluorescence of cell-free extracts. Three GFP::CBM fusions were found to express at levels similar to GFP alone, of which two CBMs were able to mediate cellulose binding of the GFP::CBM fusion. These two CBMs were then fused to multiple repeats of the self-assembling peptide RAD16-I::E (N-RADARADARADARADAE-C). The fusion protein CBM::E::(RAD16-I::E)^4 was expressed in *R. eutropha* and purified using the CBM's affinity for cellulose. Subsequent proteolytic cleavage with endoproteinase GluC liberated RAD16-I::E peptide monomers with similar properties to the chemically synthesized counterpart RAD16-I.

Keywords: Self-assembling peptides; Sapeptides; Self-assembly; Tissue engineering; Carbohydrate binding modules; Cellulose binding domains; Recombinant peptides; Multimeric fusion proteins; Multimerization; Enzymatic cleavage; High cell density; Fermentation; Protein expression

Recent reports have shown that hydrogels consisting of self-assembling peptides (sapeptides)[1] have properties that make them excellent scaffolds for tissue engineering [1–3]. Sapeptides are typically short peptides (8–16 amino acids) that consist of alternating hydrophobic and hydrophilic residues, wherein the hydrophilic amino acids alternate between positively charged and negatively charged [4–6]. For example, the sapeptide produced in this study, RAD16-I::E, has the sequence: N-RADARADARADARADAE-C (where R = arginine, A = alanine, D = aspartic acid, and E = glutamic acid) [1]. In the presence of monovalent salts, sapeptides self-assemble to form nanofibers, which on the macroscale result in a hydrogel containing >99% water [1,4–8]. The high water content, the favorable environment created by the nanofibers, and the ability to design sapeptides with specific chemistries make them an ideal cell scaffold material for tissue growth [7]. Cartilage, liver, and nerve cells have been shown to form differentiated tissue on scaffolds made from self-assembling peptides [1–3].

Sapeptides are currently produced using chemical synthesis, an expensive process in which cost increases with peptide length. In contrast, producing recombinant self-assembling peptides in a bacterial host offers a potentially

* Corresponding author. Fax: +1 603 646 2277.
E-mail address: tillman.gerngross@dartmouth.edu (T.U. Gerngross).

[1] *Abbreviations used:* Sapeptides, self-assembling peptides; CBMs, cellulose binding modules; GFP, green fluorescent protein; CHCA, α-cyano-4-hydroxycinnamic acid; MALDI, matrix-assisted laser desorption/ionization.

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.08.023

more economical solution and imposes virtually no restrictions on length due to the template-derived nature of protein biosynthesis. We have previously reported the development of a novel recombinant protein expression platform based on the Gram-negative bacterium *Ralstonia eutropha* [9–12]. Using this protein expression system, more than 10 g/L of soluble organophosphohydrolase, an enzyme prone to inclusion body formation in *Escherichia coli*, was expressed [11]. In this study we used the *R. eutropha* expression platform to produce a self-assembling peptide.

Recombinant monomeric peptides have been shown to be unstable in bacterial hosts [13–17], but fusing the peptide to an affinity tag may help avoid instability and aid in purification (see [18] for review of affinity tags). In addition, the molar ratio of peptide to affinity tag can be increased by introducing multiple repeats of the peptide, with protein cleavage sites between each repeat [19–35]. After purification of the fusion protein by affinity chromatography, the peptide can be liberated by proteolytic digest, leaving the affinity tag bound to the chromatographic matrix.

In this study, cellulose binding modules (CBMs) were chosen as the affinity tag. CBMs are ideal affinity tags because they bind strongly and with high specificity to cellulose, an inexpensive, inert chromatographic matrix that exhibits low specificity for host cell proteins (see [36,37] for review). By introducing a glutamic acid (E) after each RAD16-I repeat, the fusion protein CBM::E::(RAD16-I::E)^4 was created, allowing for the release of RAD16-I::E by treatment with endoproteinase GluC (Fig. 1) [38,39].

The first portion of this study was devoted to identifying a CBM that expresses well in *R. eutropha* and binds to cellulose. Green fluorescent protein (GFP) was fused to CBMs in a genetic library and relative fluorescence was used to compare the expression level of the various GFP::CBM fusions (Table 1 and Fig. 2). Two CBMs identified from the GFP::CBM fusion library were used to purify sapeptides. In this study, we showed that the *R. eutropha* protein expression platform can be used for the production and purification of medium length recombinant sapeptides such as RAD16-I::E.

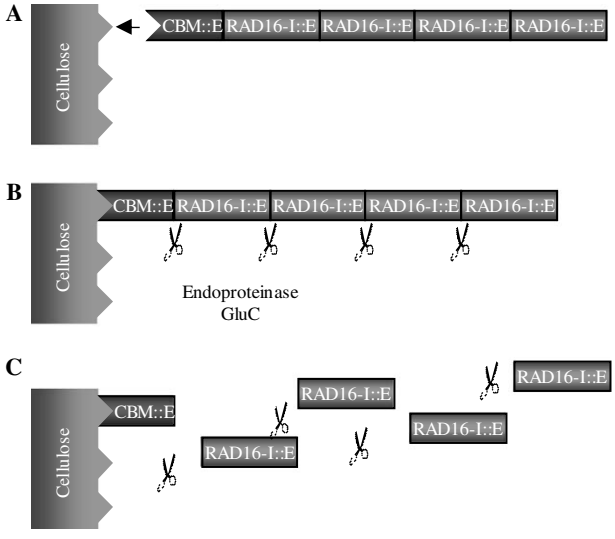

Fig. 1. Purification of fusion protein and liberation of individual RAD16-I::E monomers. (A) A cellulose binding domain is fused to tandem repeats of RAD16-I::E. The fusion protein can then be purified from soluble lysate by binding to cellulose, an inexpensive chromatography matrix. (B) Endoproteinase GluC cleaves at glutamic acids (E), (C) liberating the peptides from the cellulose-bound CBD and from each other.

Table 1
Carbohydrate binding modules used in this study

CBM	Organism	Family	Amino acids	Reference
CbhI	*Trichoderma reesei*	CBM1	34	[40]
CenD	*Cellulomonas fimi*	CBM2	101	[36]
CenB	*Cellulomonas fimi*	CBM2	101	[41]
CipB	*Clostridium thermocellum*	CBM3	83	[42]
CenC	*Cellulomonas fimi*	CBM4-9	144	[43]
LicA	*Clostridium thermocellum*	CBM4-9	144	[44]
YheB	*Escherichia coli*	CBM5-12	41	[45]
XynU	*Clostridium thermocellum*	CBM6	130	[46]
XynZ	*Clostridium thermocellum*	CBM6	130	[47]
Xyn10A	*Thermotoga maritima*	CBM9	186	[48]

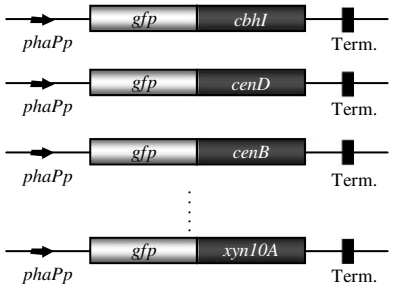

Fig. 2. Schematic representation of the GFP::CBM genetic library. Each CBM in the library was cloned between the *R. eutropha phaP* promoter (*phaPp*) and a transcriptional terminator (Term.). Next, the GFP gene *gfp-mut2* was cloned upstream of each CBM, creating a translational fusion.

Materials and methods

Growth media, antibiotics, and culture conditions

Escherichia coli strains were grown in Luria–Bertani medium (LB). *R. eutropha* strains were grown in the following media, depending on application: LB medium, PCT medium [9] (20 g/L glucose, 2.2 g/L $MgSO_4 \cdot 7H_2O$, 3 g/L K_2SO_4, 0.18 g/L Na_2SO_4, 0.18 g/L $FeSO_4$, 14 mM H_3PO_4, 2.4 ml/L trace element solution), or Lee medium [49] (20 g/L glucose, 3 g/L $Na_2HPO_4 \cdot 7H_2O$, 1 g/L KH_2PO_4, 2 g/L NH_4Cl, 0.2 g/L $MgSO_4 \cdot 7H_2O$, and 2.4 ml/L trace element solution). When appropriate, antibiotics were added in the following concentrations: chloramphenicol (50 μg/ml), kanamycin (50 μg/ml), gentamicin (10 μg/ml), and tetracycline (10 μg/ml). *R. eutropha* and *E. coli* strains were cultivated at 30 and 37 °C, respectively.

DNA preparation and manipulation

Standard procedures were used for the preparation and manipulation of DNA and for PCR [50]. All restriction

enzymes were purchased from New England Biolabs (Beverly, MA). ExTaq DNA polymerase was purchased from Takara (Otsu, Shiga, Japan). All PCR products used for cloning were confirmed by sequencing at the Molecular Biology Core Facility at Dartmouth College.

Plasmids, strains, and oligonucleotides used in the construction of the CBM library

Ten CBMs from five different species were used to create a GFP::CBM genetic library (Table 1). The methods for constructing this library are included in the supplementary information section. Briefly, *cbm*s were amplified by PCR and cloned into a pKNOCK-Cm derived plasmid between the *phaP* promoter (*phaPp*) and a transcriptional terminator. The *phaPp* is native to *R. eutropha* and is induced in the absence of phosphate [9]. Next, the GFP gene *gfpmut2* was amplified by PCR and cloned upstream of each *cbm* in the library to generate a GFP::CBM translational fusion (Fig. 2) [51]. These plasmids were integrated onto the *R. eutropha* chromosome as described previously [9].

Fluorescence measurements of soluble lysate in GFP::CBM library

R. eutropha strains were grown in 5 ml of Lee media to an OD_{600} of approximately 2.0. Four milliliters of culture were centrifuged and the supernatant was discarded. The cells were resuspended in 1 ml of PBS buffer and sonicated using a Fisher Scientific Sonic Dismembrator 550 at setting "5" with 1 pulse cycle (2 s on, 0.5 s off, 30 s total). The cell lysate was then centrifuged at 16,000*g* for 5 min and the supernatant was collected. The fluorescence of 100 μl of soluble cell lysate was measured using a SpectraMax Gemini spectrophotometer (Molecular Devices, Sunnyvale, CA). Excitation and emission wavelengths of 360 and 509 nm, respectively, were used.

Cellulose binding assay for GFP::CBM library

Of the remaining soluble cell lysate, 250 μl were then incubated with 50 mg Avicel (Fluka, St. Louis, MO) for one hour at room temperature with slow end-over-end rolling. Next, the cell lysate/cellulose slurry was centrifuged at 3800*g* for 5 min and the supernatant was removed and set aside. The cellulose pellet was washed with PBS buffer and resuspended to the original volume of 250 μl. The fluorescence of 100 μl of the supernatant and 100 μl of the resuspended cellulose pellet was then measured with the spectrophotometer.

Plasmids, strains, and oligonucleotides used in the construction of the CBM::E::(RAD16-I::E)^4 fusions

First, the Linker Duplex was inserted into the cloning vector pDR1. The oligos Linker upper and Linker lower were annealed, creating *Mun*I and *Bgl*II sticky ends at the upstream and downstream ends, respectively (Table 2). The Linker Duplex was then cloned into pDR1 digested with *Mun*I and *Bgl*II to yield pDR2.

Next, the multiple repeats of *rad16-I::e* were cloned into pDR2 (Table 2). The oligos encoding the RAD16::E repeats were designed so that their codon usage matched that of *R. eutropha* while avoiding strict repetition. Oligos RAD16 α Upper and RAD16 α Lower were annealed. Oligos RAD16 β Upper and RAD16 β Lower were also annealed. The RAD16 α and β Duplexes were ligated in a triple ligation with pDR2 cut with *Bsa*I/*Bgl*II to yield pDR3. pDR3 was in turn cut with *Bsa*I/*Bgl*II and the RAD16 α and β Duplexes were again inserted to create pDR4. This process was performed eight times, yielding up to 16 repeats.

Next, the CBM gene *xyn10A* was amplified by ExTaq PCR using primers prDR31 and prDR32 (Table 2). Plasmid pGB213 was used as a template (Supplementary Table 1). Primer prDR31 introduced a *Sac*I site at the 5′ end of the PCR product. Primer prDR32 introduced a 6His tag, as well as a *Mun*I site at the 3′ end of the PCR product. The PCR product was then isolated and cloned into pCR2.1-TOPO to create pDR52.

The gene *xyn10A* was then cloned into a pKNOCK-Cm vector upstream of four repeats of *rad16-I::e*. Plasmid pDR52 was cut with *Sac*I/*Mun*I and the 0.6 kb fragment containing *xyn10A* was isolated. This 0.6 kb fragment was then inserted into the cloning vector pDR44 cut with *Sac*I/*Mun*I, resulting in pDR54 (Table 2).

Next, the *xyn10A::e::(rad16-I::e)*^{4} fusion was cloned downstream of the T7 bacteriophage promoter. Plasmid pDR54 was cut with *Sac*I/*Apa*I and the 0.9 kb fragment was isolated. This fragment was cloned into pGH12 cut with *Sac*I/*Apa*I to yield pDR57 (Table 2).

The *cipB* gene was codon-optimized with the algorithm used by DNA 2.0 (Menlo Park, CA) (Table 3). The codon usage of *R. metallidurans* was used to optimize the *cipB* gene, resulting in $cipB_{opt}$. Plasmid pJ1:G01091 was cut with *Sac*I/*Mun*I and the 0.5 kb fragment was isolated. This fragment was cloned into pDR44 cut with *Sac*I/*Mun*I, yielding pDR70. Next, the $cipB_{opt}$::e::(*rad16-I::e*)^4 fusion was cloned in front the T7 promoter in the same fashion as the *xyn10A::e::(rad16-I::e)*^{4} fusion to yield pDR71. The gene cassettes with the T7 promoter were randomly integrated onto the *R. eutropha* chromosome using transposons as described previously [11]. Western blot using an anti-His antibody (His-probe (G-18), Santa Cruz Biotechnology, Santa Cruz, CA) was used to identify the colony that expressed the most target protein (data not shown).

Fermentation

High-cell-density fermentations were conducted as previously described [9]. During the fermentation, 50 ml aliquots were harvested and centrifuged at 3800*g* for 45 min at 4 °C. The supernatants were discarded and the pellets were frozen at −20 °C.

Table 2
Strains, plasmids, and oligonucleotides used in the sapepetides purification study

Strain, plasmid or oligo	Description	References or source
R. eutropha strains		
NCIMB 40124	Wild type (wt); gentamicin resistant	NCIMB
SS15	Wild type containing *phaPp*::T7 RNAP expression cassette; gentamycin and chloramphenicol resistant	[11]
DR57R	pDR57 introduced into *R. eutropha* SS15	This study
DR71R	pDR71 introduced into *R. eutropha* SS15	This study
E. coli strains		
S17	Host strain for plasmids derived from pKNOCK-Cm	[52–54]
Plasmid		
pCR2.1-TOPO	High copy number plasmid for cloning; confers ampicillin and kanamycin resistance	Invitrogen
pKNOCK-Cm	Suicide plasmid used for homologous recombination into *R. eutropha;*confers chloramphenicol resistance	[55]
pDR1	A cloning plasmid derived from pKNOCK-Cm with *Mun*I and *Bgl*II in its MCS	This study
pDR2	The Linker duplex in pKNOCK-Cm	This study
pDR3	$(rad16\text{-}I::e)^2$ in pKNOCK-Cm	This study
pDR4	$(rad16\text{-}I::e)^4$in pKNOCK-Cm	This study
pGB213	*xyn10A* CBM introduced into pCR2.1-TOPO	This study
pDR52	*xyn10A* PCR product in pCR2.1-TOPO	This study
pDR44	A cloning vector derived from pKNOCK-Cm with an MCS between the *phaPp* and $(rad16\text{-}I::e)^4$	This study
pDR54	*phaPp*::*xyn10A*::$(rad16\text{-}I::e)^4$ in pKNOCK-Cm	This study
pGH12	Transposon vector derived from pUTminiTn5gfp. *gfp* gene replaced with multiple cloning site; confers tetracycline resistance.	[11]
pDR57	T7p::x*yn10A*::$(rad16\text{-}I::e)^4$ in pGH12	This study
pJ1:G01091 (also pDR69)	Codon-optimized *cipB* ($cipB_{opt}$) in pJ1	DNA 2.0
pDR70	*phaPp*::$cipB_{opt}$::$(rad16\text{-}I::e)^4$ in pKNOCK-Cm	This study
pDR71	T7p::$cipB_{opt}$::$(rad16\text{-}I::e)^4$ in pGH12	This study
Oligonucleotide		
Linker upper	AATTGTGAAGAACGAGTGAGACCA	
Linker lower	GATCTGGTCTCACTCGTTCTTCAC	
RAD16 α upper	CGAGCGCGCCGACGCGCGGGCAGACGCCCGCGCGG ACGCACGGGCCGATGCGG	
RAD16 α lower	CGCTCCGCATCGGCCCGTGCGTCCGCGCGGGCGTC TGCCCGCGCGTCGGCGCG	
RAD16 β upper	AGCGCGCGGACGCACGGGCCGATGCGCGCGCAGAC GCCCGGGCGGATGCCGAGTGAGACCA	
RAD16 β lower	GATCTGGTCTCACTCGGCATCCGCCCGGGCGTCTGC GCGCGCATCGGCCCGTGCGTCCGCG	
prDR31	GAAGAGCTCATGGTAGCGACAGCAAAATACGGAACACCG	
prDR32	TTCCAATTGCGTGGTGGTGGTGGTGGTGCTTGATGAGCCT GAGGTTACCGAACTTTGAAG	

Table 3
Codon-optimized *cipB* with restriction sites and 6His tag

Gene	Sequence	Description
$cipB_{opt}$	GAGCTC**ATG**AACACGCCCGTGTCCGGCAATCTGAAAGTCGAATT CTACAACTCCAACCCGTCCGACACGACGAACTCGATTAATCCGC AATTCAAGGTCACCAACACGGGGTCCTCGGCCATCGATCTGTCC AAGCTGACCCTGCGGTATTACTACACCGTCGACGGCCAGAAGGA CCAGACGTTCTGGTGCGACCACGCCGCGATCATTGGCTCGAACG TAGCTACAACGGCATCACGTCCAATGTGAAGGGCACCTTCGTGA GATGTCGTCCAGCACGAACAATGCCGATACCTATCTCGAAATCC CTTCACCGGCGGGACCCTCGAACCGGGTGCGCATGTGCAGATCA GGGGCGCTTTGCGAAGAACGACTGGTCGAACTACACGCAGTCA CGACTACTCCTTCAAGTCCCGGTCCCAGTTCGTCGAATGGGATA GGTCACCGCCTATCTGAATGGCGTGCTGGTGTGGGGCAAGGAAC GGGCGGCTCGGTGGTCCACCACCACCATCACCACGCAATTG	Gene encoding *cipB* codon-optimized for *R. eutropha.* Contains a *Sac*I site at its 5′ end. At its 3′ end is a 6His tag and a *Mun*I site

Restriction sites are underlined. The ATG start codon is in bold.

Cellulose purification of CBM::E::(RAD16-I::E)^4

Cell pellets collected during the fermentation were thawed to room temperature and resuspended to 50 ml with CBM binding buffer (20 mM Tris–HCl, pH 7.5). Seven milliliters of this were removed for processing: 2500 U of Benzonase Nuclease (Novagen, San Diego, CA) were added and the whole cells were sonicated with a Fisher Scientific Sonic Dismembrator 550 in 20 pulse cycles (2 s on, 0.5 s off, 30 s total) with 5 min of cooling on ice between cycles. A large tip at setting "10" was used. The cell lysate was centrifuged at 16,000*g* for 20 min at room temperature and the supernatant was collected. Centrifugation was repeated two more times until the soluble cell lysate was clarified.

Fourteen milliliters of CBM binding buffer were added to the soluble lysate, which was then incubated with 3600 µl Perloza MT100 cellulose resin with a nominal particle diameter of 50–80 µm (Iontosorb, Czech Republic). The soluble lysate incubated with the cellulose resin for one hour at room temperature with slow end-over-end rolling. Next, the cell lysate/cellulose slurry was centrifuged at 3800*g* for 5 min and the supernatant was removed. The cellulose pellet was then washed with 50 ml CBM wash buffer (20 mM Tris–HCl, 4 M NaCl, pH 7.5). Finally, the cellulose pellet was washed three more times with 50 ml CBM binding buffer to remove contaminants and to remove residual NaCl.

A smaller scale experiment was performed on shake flask samples of DR57R. For these samples, Complete EDTA-Free Protease Inhibitor Cocktail (Roche Diagnostics, Mannheim, Germany) was added to the whole cell extract before sonication.

The amount of purified CipB::E::(RAD16-I::E)^4 was calculated by first eluting the protein off a sample of the cellulose with four washes of 100 mM NaOH/100 mM NaCl, pH 13.0. The amount of protein eluted off the cellulose was measured with the BCA Assay, using BSA as standards (Pierce, Rockford, IL).

Endoproteinase GluC digestion

The washed cellulose pellet from the DR71R fermentation sample was resuspended in 0.5 pellet volumes of endoproteinase GluC digestion buffer (50 mM Tris–HCl, 0.5 mM Glu-Glu, pH 8.0). Fifty micrograms of endoproteinase GluC (New England Biolabs, Beverly, MA) was added. The digest was allowed to proceed for 16 h at room temperature with slow end-over-end rolling.

rpHPLC of GluC reaction mixture

After endoproteinase GluC digestion, the cellulose was centrifuged at 16,000*g* for 1 min and the supernatant was collected and centrifuged again. The sample was loaded onto a 4.6 × 5.0 mm Jupiter 5u C5 HPLC column from Phenomenex (Torrance, CA) with a linear gradient of acetonitrile from 5 to 55% in a 0.1% TFA solution at 30 °C. The flow rate was 1.0 ml/min for 15 min. Synthetic RAD16-I standards were also loaded at concentrations ranging from 50 to 1000 µg/ml to create a standard curve with which the concentration of purified RAD16-I::E could be measured. The synthetic RAD16-I was generously provided by Dr. Shuguang Zhang at Massachusetts Institute of Technology.

Mass spectrometry

The fractions collected from rpHPLC were analyzed by mass spectrometry on a matrix-assisted laser desorption/ionization (MALDI) mass spectrometer. A Voyager DE linear MALDI-TOF (Applied Biosystems, Foster City, CA) mass spectrometer using delayed extraction was used. The matrix consisted of α-cyano-4-hydroxycinnamic acid (CHCA).

Results and discussion

Fluorescence measurements of soluble lysate in GFP::CBM library

Ralstonia eutropha strains expressing the GFP::CBM fusions, GFP, GFP::6His, and wild type controls were cultured under inducing conditions in order to compare their relative fluorescence and identify CBMs that express well in *R. eutropha*. It was assumed that GFP fluorescence was proportional to amount of soluble fusion protein.

After sonication, the fluorescence of the cell lysates were measured as described (Fig. 3, solid bars). As expected, the strain expressing GFP exhibited a high level of fluorescence compared with the wild type strain. Strains expressing GFP::CipB, GFP::XynZ, and GFP::Xyn10A had fluorescence levels comparable to that of the strain expressing GFP by itself. The GFP::LicA strain also exhibited fluorescence above that of wild type, but at a level much lower than the strain expressing GFP by itself. The remaining strains expressing GFP::CBM fusions showed no significant difference in fluorescence from that of wild type. Inter-

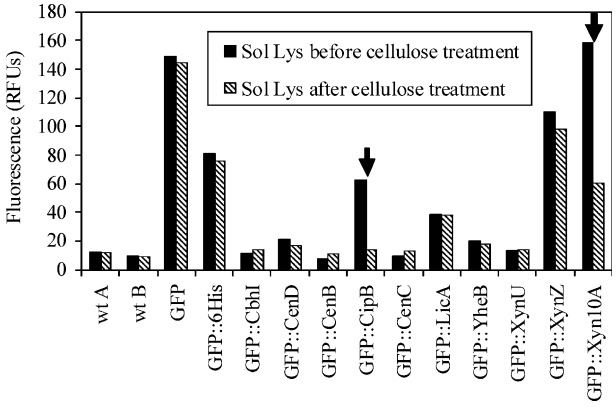

Fig. 3. Fluorescence of *R. eutropha* soluble lysates before and after treatment with crystalline cellulose. After growth in induction media, cell lysis, and clarification, the soluble lysate was assayed for fluorescence (solid bars). Crystalline cellulose was then mixed with the cell lysate for 1 h at room temperature, removed, and the fluorescence of the soluble lysate was again measured (striped bars). The CipB and Xyn10A cellulose binding domains are highlighted by arrows.

estingly, the addition of a 6His tag on the C-terminus of GFP reduced expression by about 30%, more than would be expected for such a small tag.

Cellulose binding assay for GFP::CBM library

Next, the ability of GFP::CBM fusions to bind to cellulose was tested by comparing the fluorescence of soluble lysate before and after incubation with Avicel (Fig. 3). Of the three CBMs that expressed well, only the soluble lysates of the strains expressing GFP::CipB and GFP::Xyn10A showed a significant decrease in fluorescence after Avicel treatment, which suggested binding of the fusion protein to Avicel. Although GFP::XynZ and GFP::LicA expressed, these fusion proteins did not bind to Avicel.

To confirm that a reduction in fluorescence in the soluble lysate was the result of protein binding to Avicel, the washed, resuspended Avicel pellets were then assayed for fluorescence (Fig. 4). As expected, only the GFP::Xyn10A and GFP::CipB samples showed high levels of fluorescence on Avicel pellets. These two CBMs both expressed well in *R. eutropha* and bound to cellulose, thus making them attractive candidates for purifying proteins of interest such as RAD16-I::E repeats.

R. eutropha strain generation for strains expressing CBM::E::(RAD16-I::E)^4

After CipB and Xyn10A were identified as effective affinity tags, these CBMs were then used to purify RAD16-I::E repeats. For these strains, the strong T7 bacteriophage promoter was used to control the gene of interest. This gene cassette was integrated onto the *R. eutropha* chromosome as a transposon in a strain that already expressed T7 RNA polymerase under the control of the inducible *phaPp*. Because of the random nature of transposon-mediated integration onto the *R. eutropha* chromosome, colonies of strains with the T7 promoter were screened using western blot to find the one that expressed the most target protein (data not shown).

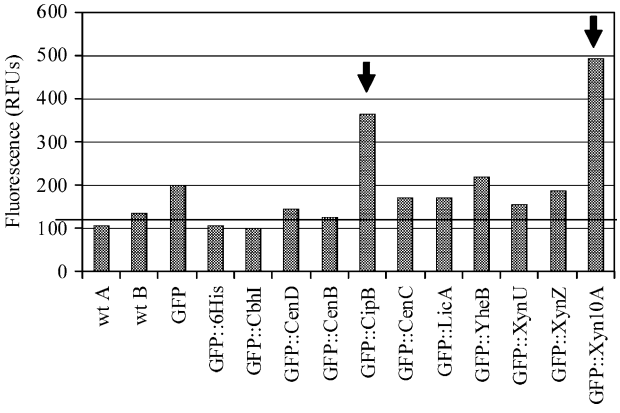

Fig. 4. Fluorescence of cellulose after interaction with soluble lysates. After the soluble lysate was treated with cellulose, the cellulose was washed with PBS buffer, resuspended to its original volume, and measured for fluorescence. The solid horizontal line marks the average of the two wild type controls. The CipB and Xyn10A cellulose binding domains are highlighted by arrows.

Cellulose purification of Xyn10A::E::(RAD16-I::E)^4

Next, purification of Xyn10A::E::(RAD16-I::E)^4 produced by DR57R was attempted with the expectation that this CBM, which performed well when fused to GFP, would also result in the expression of this fusion protein. However, when this protein was analyzed by SDS–PAGE following cellulose purification, it was found to be highly degraded (data not shown). Degradation was not prevented in the presence of a protease inhibitor cocktail during cell disruption (data not shown).

Cellulose purification of CipB::E::(RAD16-I::E)^4 from fermentation samples

After sonication of samples from high-cell-density fermentation of DR71R (*R. eutropha* expressing CipB::E::(RAD16-I::E)^4), the lysate was clarified and the fusion protein was cellulose-purified as described above. A band corresponding to the target protein is clearly visible in the whole cell fraction (Fig. 5, lane 2) and the soluble cell lysate (Fig. 5, lane 4) at approximately 35 kDa. Western blot confirmed the identity of this band as the target protein, which was not present in wild type (data not shown). According to densitometry measurements, this protein made up roughly 1.8% of total soluble protein. This band is not present in the

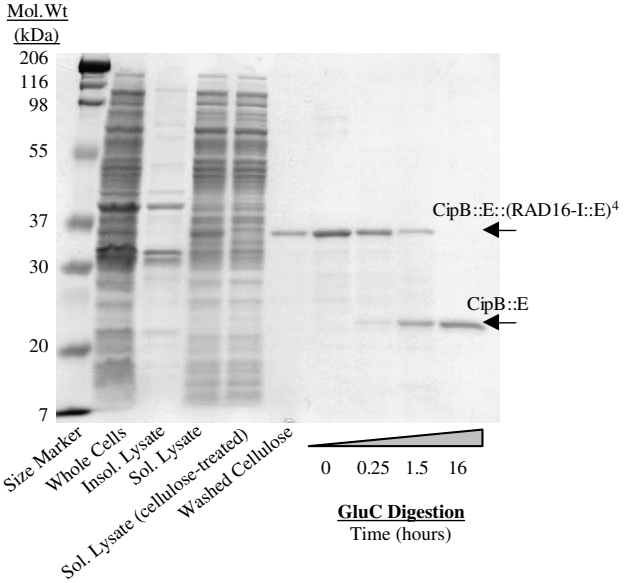

Fig. 5. Cellulose purification and endoproteinase GluC digestion of CipB::E::(RAD16-I::E)^4. Lane 1, molecular weight marker; lane 2, whole cell extract of DR71R grown in high-cell-density fed-batch culture under inducing conditions; lane 3, insoluble lysate of DR71R; lane 4, soluble lysate of DR71R; lane 5, soluble lysate of DR71R after treatment with Perloza cellulose resin; lane 6, washed cellulose pellet; lanes 7–10, endoproteinase GluC digestion of purified, cellulose-bound CipB::(RAD16-I::E)^4, $t = 0$, 0.25, 1.5, and 16 h, respectively. Proteins were stained with Simply Blue Safestain (Invitrogen, Carlsbad, CA).

soluble lysate after incubation with cellulose (Fig. 5, lane 5) and it is the only band in the washed cellulose fraction (Fig. 5, lane 6), confirming its identity as a functional CBM fusion. It was determined that 419 mg of CipB::E::(RAD16-I::E)^4 were purified per liter of *R. eutropha* culture, which corresponds to a theoretical yield of 112 mg/L of RAD16-I::E.

Endoproteinase GluC cleavage of cellulose-bound CipB::E::(RAD16-I::E)^4

To release the peptide repeats from the cellulose-bound CipB, the purified fusion protein was treated with endoproteinase GluC as described (Fig. 5). As the reaction proceeded, the size of the band decreased from 35 to 22 kDa, but it did not disappear altogether. Therefore, the four glutamic acids separating the RAD16-I peptides were assumed to be cleaved, leaving CipB intact. It is likely that the CipB portion of the fusion protein was tightly folded and the five glutamic acids within its sequence were protected from endoproteinase GluC digestion under the chosen conditions. In contrast, the multiple repeats are not expected to have a natural native conformation, giving the protease access to those glutamic acid sites.

Characterization of product after endoproteinase GluC cleavage

After endoproteinase GluC cleavage, the cellulose resin was removed and the reaction mixture was analyzed by rpHPLC as described (Fig. 6). The major peak in the chromatogram had a retention time of 9.25 min. The fraction containing the major peak was collected and analyzed using MALDI-TOF mass spectrometry (Fig. 6, inset). Mass spectrometry showed that the predominant species in the sample from the rpHPLC major peak had a molecular weight of 1801 Da, which is identical to the calculated mass of RAD16-I::E.

A standard curve was created by measuring the areas under the peaks of the rpHPLC chromatograms caused by different concentrations of RAD16-I standards. Using the RAD16-I standard curve, it was determined that 10.1 mg of RAD16-I::E were purified per liter of *R. eutropha* culture (Table 4).

The purified product was shown to self-assemble under the same conditions as synthetic RAD16-I (data not shown) [1]. Therefore, it appears that the additional C-terminal glutamic acid does not greatly affect the self-assembling properties of the peptide. Endoproteinase GluC was chosen to cleave the fusion protein in large part because the similarities between E and D made it less likely that a C-terminal E would affect the properties of the sapeptide. Furthermore, this protease is active in the absence of salt, preventing self-assembly during the cleavage reaction [39].

Conclusions

By constructing a GFP::CBM fusion library we were able to identify two CBMs that both expressed at comparable levels to GFP in *R. eutropha* and retained the ability to bind cellulose. These two CBMS were Xyn10A from *T. maritima* and CipB from *C. thermocellum*.

Although Xyn10A was identified as a functional CBM that expressed well in *R. eutropha*, the fusion protein Xyn10A::E::(RAD16-I::E)^4 was found to be degraded in vivo. In contrast, the purified CipB::E::(RAD16-I::E)^4 fusion protein was full-length and could be purified at 419 mg protein per liter of fermentation medium. However, only 10.1 mg/L of RAD16-I::E were recovered after treatment with endoproteinase GluC out of a theoretical maximum of 112 mg/L.

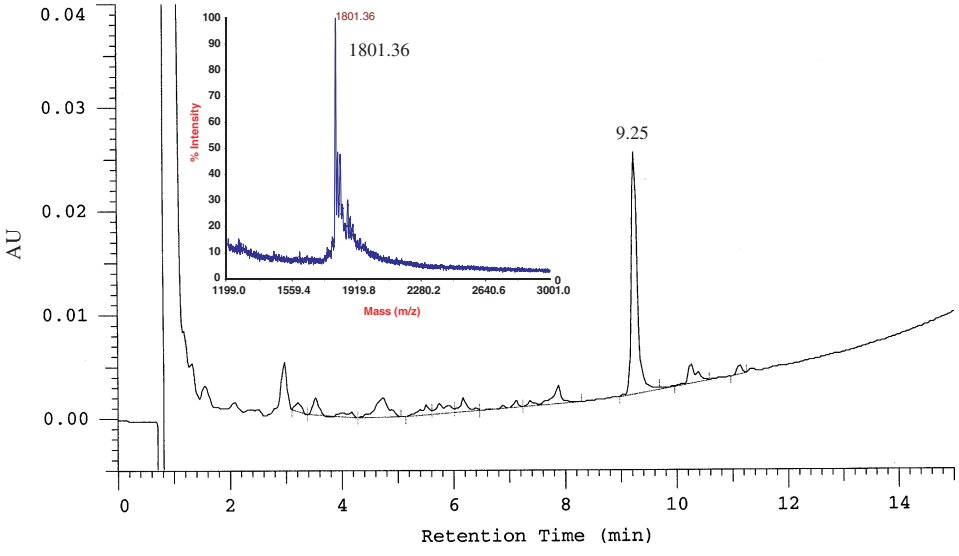

Fig. 6. rpHPLC on endoproteinase GluC reaction mixture containing purified RAD16-I::E monomers. The purified, cellulose-bound fusion protein CipB::E::(RAD16-I::E)^4 was treated with endoproteinase GluC. The cellulose was then removed and the reaction was analyzed using rpHPLC. The fraction containing the major peak was analyzed using mass spectrometry (inset).

Table 4
Purification of recombinant RAD16-I::E from *R. eutropha*

Purification step	Total protein[a] (mg)	Mass of CBD::E::(RAD16-I::E)^4 fusion protein[a] (mg)	Step yield (%)	Overall yield (%)	Purity (%)
Recovery of CipB::E::(RAD16-I::E)^4 fusion protein					
Soluble lysate	24500[b]	433[c]	na	100	1.8
Cellulose purification	419[b]	~419[b]	97	97	>95
Purification step	Total protein[a] (mg)	Mass of RAD16-I::E peptide[a] (mg)	Step yield[d] (%)	Overall yield[d] (%)	Purity (%)
Recovery of RAD16-I::E peptide					
Endoproteinase GluC digestion	14.3[e]	10.1[e]	9	9	71
rpHPLC	10.1[e]	10.1[e]	>95	~9	>95

[a] Normalized to 1 L of culture.
[b] Protein concentration was determined by BCA protein assay, using bovine serum albumin as a standard.
[c] The amount of fusion protein in the soluble cell lysate was determined by densitometry of Coomassie-stained gels.
[d] Theoretically, 1000 mg of fusion protein yields 267 mg of RAD16-I::E peptide.
[e] Protein concentration was determined by rpHPLC, using purified RAD16-I peptide as a standard.

This work further demonstrates the effectiveness of using CBMs to recover heterologous proteins at high yield and purity. Furthermore, cellulose resin is inexpensive, about 1/100 the cost of the Ni^+ resin commonly used to purify 6His-tagged proteins [48].

Now that CipB from *C. thermocellum* has been identified as a CBM that expresses well in *R. eutropha*, future work will focus on optimizing expression of the CipB::E::(RAD16-I::E)^4 fusion protein in high cell density culture. For example, the strain expressing CipB::E:: (RAD16-I::E)^4 only contains a single chromosomal copy, so increasing the gene dosage will likely increase titer of fusion protein. Also, optimizing the number of RAD16-I::E repeats may improve process yields. The most immediate work, however, is to improve the yield of the endoproteinase GluC cleavage step which currently suffers from substantial losses.

This work is important because it could decrease the cost of treatments based on self-assembling peptides and allow for the exploration of longer sapeptides that currently are difficult and expensive to synthesize. Chemical synthesis of the sapeptide RAD16-I costs about $495 per gram (95% purity, 100 g scale) [56]. In contrast, the cost of recombinant enzymes is typically on the order of $0.001 to $0.18 per gram ($1–100/lb) [57]. Clearly, recombinant production of peptides has the potential to dramatically reduce the cost of sapeptides.

Self-assembling peptides also have applications besides tissue scaffolds. For example, some sapeptides can function as a scaffold to manufacture nanowires in the computer industry [58]. Other sapeptides form nanotubes, and others create nanometer-thick surface coatings [58]. Finally, as therapeutic peptides have become more popular and manufacturing remains a major bottleneck, this work could be modified to express other peptides of interest such as mammalian peptide hormones [59].

Acknowledgments

We thank Drs. Byung-Kwon Choi and Juergen Nett for advice on planning and troubleshooting. Expertise with rpHPLC and mass spectrometry was generously given by Drs. Huijuan Li, Michael Cukan, and Bing Gong. We also thank Mary Kay Brown, M.S. for her untiring support in managing our laboratory. We thank Grant Henderson, Dr. Jesse McCool, Maaza Mehzun, and Mary Kay Brown, M.S. for critically reading the manuscript. This research was funded by NIST Grant 60NANB1D0064.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.pep.2005.08.023.

References

[1] T.C. Holmes, et al., Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds, Proc. Natl. Acad. Sci. USA 97 (2000) 6728–6733.
[2] J. Kisiday, et al., Self-assembling peptide hydrogel fosters chondrocyte extracellular matrix production and cell division: implications for cartilage tissue repair, Proc. Natl. Acad. Sci. USA 99 (2002) 9996–10001.
[3] C.E. Semino, J.R. Merok, G.G. Crane, G. Panagiotakos, S. Zhang, Functional differentiation of hepatocyte-like spheroid structures from putative liver progenitor cells in three-dimensional peptide scaffolds, Differentiation 71 (2003) 262–270.
[4] S. Zhang, T. Holmes, C. Lockshin, A. Rich, Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane, Proc. Natl. Acad. Sci. USA 90 (1993) 3334–3338.
[5] S. Zhang, et al., Self-complementary oligopeptide matrices support mammalian cell attachment, Biomaterials 16 (1995) 1385–1393.
[6] M.R. Caplan, P.N. Moore, S. Zhang, R.D. Kamm, D.A. Lauffenburger, Self-assembly of a beta-sheet protein governed by relief of electrostatic repulsion relative to van der Waals attraction, Biomacromolecules 1 (2000) 627–631.
[7] T.C. Holmes, Novel peptide-based biomaterial scaffolds for tissue engineering, Trends Biotechnol. 20 (2002) 16–21.
[8] M.R. Caplan, E.M. Schwartzfarb, S. Zhang, R.D. Kamm, D.A. Lauffenburger, Control of self-assembling oligopeptide matrix formation through systematic variation of amino acid sequence, Biomaterials 23 (2002) 219–227.
[9] S. Srinivasan, G. Barnard, T.U. Gerngross, A novel high-cell-density protein expression system based on *Ralstonia eutropha*, Appl. Environ. Microbiol. 68 (2002) 5925–5932.
[10] S. Srinivasan, G. Barnard, T.U. Gerngross, Production of recombinant proteins using multiple-copy gene integration in high-cell-den-

sity fermentations of *Ralstonia eutropha*, Biotechnol. Bioeng. 84 (2003) 114–120.
[11] G. Barnard, G.E. Henderson, S. Srinivasan, T.U. Gerngross, High level recombinant protein expression in *Ralstonia eutropha* using T7 RNA polymerase based amplification, Protein Express. Purif. 38 (2004) 264–271.
[12] G.C. Barnard, J. McCool, D.W. Wood, T.U. Gerngross, An integrated recombinant protein expression and purification platform based on *Ralstonia eutropha*, Appl. Environ. Microbiol. (2005) in press.
[13] K. Itakura, et al., Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin, Science 198 (1977) 1056–1063.
[14] D.V. Goeddel, et al., Expression in *Escherichia coli* of chemically synthesized genes for human insulin, Proc. Natl. Acad. Sci. USA 76 (1979) 106–110.
[15] R. Wetzel, et al., Production of biologically active N alpha-desacetylthymosin alpha 1 in *Escherichia coli* through expression of a chemically synthesized gene, Biochemistry 19 (1980) 6096–6104.
[16] J. Shine, I. Fettes, N.C. Lan, J.L. Roberts, J.D. Baxter, Expression of cloned beta-endorphin gene sequences by *Escherichia coli*, Nature 285 (1980) 456–463.
[17] S. Sumi, F. Nagawa, T. Hayashi, H. Amagase, M. Suzuki, Overproduction of a gastrointestinal hormone, secretin, in *Escherichia coli*cells and its chemical characterization, Gene 29 (1984) 125–134.
[18] J. Nilsson, S. Stahl, J. Lundeberg, M. Uhlen, P.A. Nygren, Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins, Protein Express. Purif. 11 (1997) 1–16.
[19] M.T. Doel, et al., The expression in *E. coli* of synthetic repeating polymeric genes coding for poly(L-aspartyl-L-phenylalanine), Nucleic Acids Res. 8 (1980) 4575–4592.
[20] S.H. Shen, Multiple joined genes prevent product degradation in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 81 (1984) 4627–4631.
[21] T. Kempe, et al., Multiple-copy genes: production and modification of monomeric peptides from large multimeric fusion proteins, Gene 39 (1985) 239–245.
[22] M.F. Schultz, G. Buell, E. Schmid, R. Movva, G. Selzer, Increased expression in *Escherichia coli* of a synthetic gene encoding human somatomedin C after gene duplication and fusion, J. Bacteriol. 169 (1987) 5385–5392.
[23] M. Lennick, J.R. Haynes, S.H. Shen, High-level expression of alpha-human atrial natriuretic peptide from multiple joined genes in *Escherichia coli*, Gene 61 (1987) 103–112.
[24] T. Murata, S. Horinouchi, T. Beppu, Production of poly(L-aspartyl-L-phenylalanine) in *Escherichia coli*, J. Biotechnol. 28 (1993) 301–312.
[25] A. Kuliopulos, C.T. Walsh, Production, purification, and cleavage of tandem repeats of recombinant peptides, J. Am. Chem. Soc. 116 (1994) 4599–4607.
[26] J. Raingeaud, et al., Production, analysis and bioactivity of recombinant vasoactive intestinal peptide analogs, Biochimie 78 (1996) 14–25.
[27] H. Ishikawa, H. Tamaoki, Production of human calcitonin in *Escherichia coli* from multimeric fusion protein, J. Ferm. Bioeng. 82 (1996) 140–144.
[28] J.H. Lee, I. Minn, C.B. Park, S.C. Kim, Acidic peptide-mediated expression of the antimicrobial peptide buforin II as tandem repeats in *Escherichia coli*, Protein Express. Purif. 12 (1998) 53–60.
[29] C.J. Park, J.H. Lee, S.S. Hong, H.S. Lee, S.C. Kim, High-level expression of the angiotensin-converting-enzyme-inhibiting peptide, YG-1, as tandem multimers in *Escherichia coli*, Appl. Microbiol. Biotechnol. 50 (1998) 71–76.
[30] P. Jonasson, et al., Integrated bioprocess for production of human proinsulin C-peptide via heat release of an intracellular heptameric fusion protein, J. Biotechnol. 76 (2000) 215–226.
[31] S.J. Lee, et al., A novel technique for the effective production of short peptide analogs from concatameric short peptide multimers, Mol. Cell 10 (2000) 236–240.
[32] S. Dilsen, et al., Fed-batch production of recombinant human calcitonin precursor fusion protein using *Staphylococcus carnosus* as an expression-secretion system, Appl. Microbiol. Biotechnol. 54 (2000) 361–369.
[33] J.H. Lee, M.S. Kim, J.H. Cho, S.C. Kim, Enhanced expression of tandem multimers of the antimicrobial peptide buforin II in *Escherichia coli* by the DEAD-box protein and trxB mutant, Appl. Microbiol. Biotechnol. 58 (2002) 790–796.
[34] K.S. Oh, D.K. Na, M.H. Kweon, H.C. Sung, Expression and purification of the anticomplementary peptide Sh-CRIT-ed1 (formerly Sh-TOR-ed1) as a tetramultimer in *Escherichia coli*, Protein Express. Purif. 27 (2003) 202–209.
[35] C. Li, M.L. Ng, Y. Zhu, B. Ho, J.L. Ding, Tandem repeats of Sushi3 peptide with enhanced LPS-binding and -neutralizing activities, Protein Eng. 16 (2003) 629–635.
[36] P. Tomme, et al., Characterization and affinity applications of cellulose-binding domains, J. Chrom. B 715 (1998) 283–296.
[37] I. Levy, O. Shoseyov, Cellulose-binding domains: biotechnological applications, Biotechnol. Adv. 20 (2002) 191–213.
[38] G.R. Drapeau, Y. Boily, J. Houmard, Purification and properties of an extracellular protease of *Staphylococcus aureus*, J. Biol. Chem. 247 (1972) 6720–6726.
[39] S.B. Sorensen, T.L. Sorensen, K. Breddam, Fragmentation of proteins by *S. aureus* strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid, FEBS Lett. 294 (1991) 195–197.
[40] G. Carrard, A. Koivula, H. Soderlund, P. Beguin, Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose, Proc. Natl. Acad. Sci. USA 97 (2000) 10342–10347.
[41] J.B. Owolabi, P. Beguin, D.G. Kilburn, R.C. Miller Jr., A.J. Warren, Expression in *Eschericia coli* of the *Cellulomonas fimi* Structural Gene for Endoglucanase B, Appl. Environ. Microbiol. 54 (1988) 518–523.
[42] Y. Berdichevsky, et al., Matrix-assisted refolding of single-chain Fv-cellulose binding domain fusion proteins, Protein Express. Purif. 17 (1999) 249–259.
[43] J.B. Coutinho, B. Moser, D.G. Kilburn, R.A. Warren, R.C. Miller Jr., Nucleotide sequence of the endoglucanase C gene (cenC) of *Cellulomonas fimi*, its high-level expression in *Escherichia coli*, and characterization of its products, Mol. Microbiol. 5 (1991) 1221–1233.
[44] K.P. Fuchs, V.V. Zverlov, G.A. Velikodvorskaya, F. Lottspeich, W.H. Schwartz, Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo-beta-1,3-glucanase bound to the outer cell surface, Microbiology 149 (2003) 1021–1031.
[45] H. Simpson, F. Barras, Functional analysis of the carbohydrate-binding domains of *Erwinia chrysanthemi* Cel5 (Endoglucanase Z) and an *Escherichia coli* putative chitinase, J. Bacteriol. 181 (1999) 4611–4616.
[46] A.C. Fernandes, et al., Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes, Biochem. J. 342 (1999) 105–110.
[47] O. Grepinet, M.C. Chebrou, P. Beguin, Nucleotide sequence and deletion analysis of the xylanase gene (xynZ) of *Clostridium thermocellum*, J. Bacteriol. 170 (1988) 4582–4588.
[48] M. Kavoosi, et al., Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritima*, J. Chrom. B 807 (2004) 87–94.
[49] J.S. Park, H.C. Park, T.L. Huh, Y.H. Lee, Production of poly-beta-hydroxybutyrate by *Alcigenes eutrophus* transformants harbouring cloned phbCAB genes, Biotechnol. Lett. 17 (1995) 735–740.
[50] J. Sambrook, D. Russel, Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY, 2001.
[51] G.M. York, B.H. Junker, J. Stubbe, A.J. Sinskey, Accumulation of the PhaP phasin of *Ralstonia eutropha* is dependent on production of polyhydroxybutyrate in cells, J. Bacteriol. 183 (2001) 4217–4226.
[52] R. Simon, U. Priefer, A. Pfihler, A broad host-range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram-negative bacteria, Biotechnology 1 (1983) 784–791.
[53] V. de Lorenzo, M. Herrero, U. Jakubzik, K.N. Timmis, Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in Gram-negative eubacteria, J. Bacteriol. 172 (1990) 6568–6572.

[54] J. Quandt, M.F. Hynes, Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria, Gene 127 (1993) 15–21.
[55] M.F. Alexeyev, The pKnock series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertions into the chromosome of Gram-negative bacteria, BioTechniques 26 (1999) 824–826.
[56] R. Skeoch, Synpep Co., personal communication, 2005.
[57] J.E. Bailey, D.F. Ollis, Biochemical Engineering Fundamentals, McGraw-Hill, New York, 1986.
[58] S. Zhang, Fabrication of novel biomaterials through molecular self-assembly, Nat. Biotechnol. 21 (2003) 1171–1178.
[59] P.W. Latham, Therapeutic peptides revisited, Nat. Biotechnol. 17 (1999) 755–757.