Protein purification involving a unique auto-cleavage feature of a repeated EAAAK peptide

Yue-Jin Wu, Chia-Ying Fan, Yaw-Kuen Li

Department of Applied Chemistry, National Chiao Tung University, 1001 Ta-Hseh Rd., Hsinchu 30050, Taiwan

Article history:
Received 11 April 2009
Accepted 5 October 2009
Available online 12 October 2009

Keywords:
Protein purification
Peptide linker
Auto-cleavage
Chitin column

ABSTRACT

Protein purification generally requires many steps of column chromatography that typically involve ion-exchange, hydrophobic-interaction and gel-filtration separations. More sophisticated purification of protein might be achieved through an application of affinity binding on a functionalyzed gel such as a nickel column, glutathione-modified column, maltose-modified gel column or others. Of several drawbacks existing in these methods, fusion proteins are commonly obtained, protease digestion might be necessary to remove the fusion moiety; a costly gel is employed for affinity binding, etc. Here we report that an expression vector derived from pREST was constructed to compose the gene of the chitin-binding protein (CBP) and the nucleotide sequence of the (EAAAK)₅ peptide linker following restriction sites for target gene insertion. Fusion proteins were expressed with E. coli and purified with a chitin column. The (EAAAK)₅ linker is shown to possess a pH-dependent auto-cleavage feature. In the range pH 6–7, the target protein becomes automatically released from the fusion protein without proteolytic treatment. Although the mechanism of this auto-cleavage property of an (EAAAK)₅ linker is unclear, this feature has been successfully employed for many cases of protein purification without the tag of a fusion protein.

1. Introduction

An indispensable tool for the purification of a recombinant protein is an affinity tag, with which a target protein is obtainable from a crude cell extract via a one-step process [1]. The specificity of a target protein purification and the simplicity of its chromatographic manipulation are two attractive factors, but common drawbacks of many affinity purification systems are the costly affinity matrix and the post-proteolytic process.

Because of demands of research and industrial applications, several commercial kits have been developed using fusion proteins as tags for affinity purification, such as glutathione S-transferases (GST) (Novagen GST Bind kits), histidine tagging (Novagen His Bind Resin and Purification kit), a maltose-binding protein (pMAL system, New England BioLabs) and a chitin-binding domain (IMPACT system, New England BioLabs). The glutathione S-transferases (GST) are used to label proteins for expression and purification applications. GST-Mag agarose beads serve to bind the GST fusion protein, and gentle elution with reduced glutathione buffer. Enterokinase suffices to cleave GST and the target protein [2]. His-tag can be used in affinity chromatography with a column that has nickel bound to it; histidine is an effective ligand for nickel so that the fusion protein with His-tag binds to the nickel column, and the protein can be eluted with a solution of imidazole. In Novagen products, the Ni-NTA His-Bind Resin is used to purify the His-tag fusion protein. To cleave the fusion protein, most Novagen vectors containing the His Tag sequence on the N-terminus also encode protease cleavage sites such as that for thrombin, Factor Xa, or enterokinase [3,4]. Maltose-binding protein (MBP) has a high affinity for binding to maltose. In the pMAL system (New England BioLabs), MBP has been used as an N-terminal fusion tag to purify over-expressed proteins with affinity chromatography based on amylose resins [5,6]. Factor Xa protease is used to cleave MBP and a target protein. The expression system with chitin-binding domain fusion protein enables efficient coupling to chitin-agaroise affinity resins and purification with single-step chromatography [7]. In the IMPACT system (New England BioLabs), the chitin beads serve to purify a fusion protein; IMPACT utilizes the inducible self-cleavage activity of inteins for its ability to separate a recombinant protein from the affinity tag without the use of a protease. Nevertheless, a chitin bead is still a costly material for purification.

Our objective in this work was to develop a system for affinity purification with a cheap solid support and also possibly to eliminate the treatment of protease to retain the target protein. As chitin, a linear polymer of β-(1–4)-linked N-acetylglucosamine, is an abundant natural material produced in the biosphere [8] and exhibits substantial mechanical and chemical strength, it is a perfect candidate for our purpose. Three crystalline forms of chitin
have been described in terms of the arrangement of the chitin chains: α-chitin in which chains run anti-parallel, β-chitin in which chains run parallel, and γ-chitin with mixed parallel/anti-parallel chains. These natural polymers would provide an economical solid support for further protein purification if an appropriate binding apparatus were designed. In addition to a chitin-binding domain in chitinase, a chitin-binding protein (CBP) was found in the culture medium of *Serratia marcescens* when it was grown on chitin [9]. This protein belongs to the carbohydrate-binding module (CBM) family 33 and is known to bind to chitin and strongly promote the hydrolysis of crystalline substrate structure[11], CBP can increase the substrate accessibility and binding of CBP to chitin leads to non-hydrolytic disruption of the substrate structure[11]. Previous work showed also that CBP can bind to chitin in which chains run anti-parallel, β-chitin[9,10].

2. Material and methods

*S. marcescens* (ATCC 990) (Food Industry Research and Development Institute, Hsinchu, Taiwan), buffers (Sigma–Aldrich, St. Louis, MO, USA), Vent DNA polymerase (NEB) and Pfu DNA polymerase (Merck) were obtained from the indicated sources. Primers 1 and 2 (all sequences are summarized in Table 1) with NdeI and XhoI sites (underlined) in the sequences, respectively. PCR amplification was performed using Vent DNA polymerase (NEB) for 25 cycles with the conditions set for each cycle as 94°C, 30 s for denaturation, 58°C, 30 s for annealing, and 72°C, 4 min for extension. The PCR fragment was inserted into Ndel and Xhol sites of pRSET A to form the vector of pRSET/CBP.

2.2. Construction of the expression vector of CBP fusion protein

Several steps were involved in the construction of the expression vector of the CBP fusion protein with a linker and a genenase I recognition site. In the first step, the Ndel site in pRSET/CBP was removed by site-directed mutagenesis using primers 3 and 4 for PCR amplification; the resulting vector is designated de-Ndel-pRSET/CBP. In the second step, a genenase I cutting site and a Ndel cutting site were inserted right after the CBP gene, with which the stop codon in *cbp* was deleted simultaneously; primers 5 and 6 were designed for this purpose. Both primers were phosphorylated with T4 polynucleotide kinase before use for PCR amplification. The amplified fragments were subject to self-ligation to form pRSET/CBP-G. Various peptide linkers can be introduced in the third step. For the case of adding a (EAAAK)5 linker in pRSET/CBP-G, primers 7 and 8 were employed for backward PCR amplification. Similar to the second step, both primers were phosphorylated in advance so that the amplified DNA fragments were subject to self-ligation to give pRSET/CBP-V5G. Vectors with distinct repeated EAAAK linkers, such as (EAAAK)2, (EAAAK)3 and (EAAAK)5, were constructed with a similar strategy using primers 9 and 10, primers 9 and 11, primers 9 and 12, respectively. A vector without the genenase I proteolytic site was also constructed using primers 13 and 14; the resulting vector was named pRSET/CBP-V5. Primers 15 and 16 were employed for vector construction without CBP. For functional evaluation of this system, the gene of lamarinipentaose-producing-β,1,3-glucanase (LPHase) from *Streptomyces matensis* was inserted into Ndel/EcoRI sites; the gene of chitinase from *Bacillus cereus* [19] was inserted into Ndel/NcoI sites, and the gene of chitosanase (CNS) from *Aspergillus fumigatus* [20] was inserted into Ndel/HindIII sites of pRSET/CBP-V5G vector, separately, and expressed in *E. coli* BL21 (DE3). The three fusion proteins CBP-V5G-LPHase, CBP-V5G-chitinase and CBP-V5G-CNS were further purified on a β-chitin column (described in a following section).

2.3. β-Chitin matrix preparation

Squid pens (cartilages) were used to prepare a β-chitin matrix and column. Cartilages (50 g) were soaked in NaOH (3%, 11) and heated at 100°C for 3 h. The cartilages were then thoroughly washed with water to eliminate NaOH. The cartilages were further suspended in HCl (6M, 11) and kept near 25°C for 12 h. After washing with distilled water several times to increase the pH>6, the cartilages were homogenized and sieved to diameter
0.3–0.6 mm of a β-chitin matrix, corresponding to 30–50 mesh particle size. The β-chitin matrix was kept at 4°C for further applications.

2.4. Cultural conditions and purification of CBP and fusion proteins

E. coli BL21 (DE3) served as the host strain for protein expression. A single colony bearing pRSET/CBP was inoculated into LB medium (5 ml) containing ampicillin (0.1 mg/ml) and cultured at 37°C on a rotary shaker for 12 h. The overnight culture was then transferred into a conical flask (2 l) containing LB medium (1 l) with ampicillin (0.1 mg/ml) and IPTG at 37°C for 15 h. The culture broth was centrifuged at 4°C for 10 min at 7000 g. The cell pellet was suspended in sodium phosphate buffer (10 ml, 20 mM, pH 7.0) and then subjected to cell disruption with ultrasonication. After removal of the cell debris with centrifugation, the supernatant (∼10 ml) was mixed with β-chitin matrix (100 ml, Tris 50 mM, (NH₄)₂SO₄ 1 M, pH 8.0, 5% (v/w) 0.3–0.6 mm diameter of β-chitin) and incubated at 4°C overnight. The resulting mixture was then packed onto a 2.5 cm × 60 cm I.D. column and washed with 250 ml of the buffer solution [Tris 50 mM, (NH₄)₂SO₄ 1 M, pH 8.0] as a 10-fold volume of the solid chitin matrix to remove the non-binding protein. CBP was eluted from the β-chitin matrix using 50 ml of acetate buffer (20 mM, pH 3.6) with flow rate at 2 ml/min. Each fraction of 5 ml eluent was collected. All other CBP fusion proteins, including CBP-V5G-LPHase, CBP-V5G-chitinase and CBP-V5G-CNS, were purified based on the general protocol described above. The activity of the purified enzyme was assayed by using the corresponding polysaccharide as substrate (described in the following section) and the purity was analyzed by SDS-PAGE. The protein recovery yield and the purity enhancement were calculated based on the enzymatic activity and the amount of the protein retained. The recombinant chitosanase fusion protein without CBP (i.e. V5G-CNS) was purified according to the literature [20].

2.5. Protein determination

The protein content of the enzyme preparation was determined by the BCA method as described in the manufacturer’s protocol (BCA-1 kit for protein determination; Sigma–Aldrich, St. Louis, MO).

2.6. Enzyme activity assay

The enzymatic assay of LPHase, chitinase, and CNS were performed by using β-1,3-glucan, colloidal chitin, and chitosan as the substrate, respectively. The activities were analyzed by estimating the amount of the reducing ends of sugars using the dinitrosalicic acid method [21]. The standard assay was prepared by mixing 0.3 ml of polysaccharide substrate (1%) and 0.3 ml of enzyme (at a suitable dilution) and incubated for 4 h at 37°C; 0.6 ml of dinitrosalicic acid reagent was then added, and the resulting mixture was boiled for 15 min, chilled, and centrifuged to remove the precipitants. The resulting adducts of reducing sugars were analyzed and measured spectrophotometrically at 540 nm.

2.7. Electrospray mass-spectrometric (ESI-MS) analysis

Mass spectra were recorded with a quadrupole time-of-flight mass filter (Q-TOF, Micromass, UK). This analyzer was scanned over a range 500–3000 m/z of mass-to-charge ratio (m/z) for protein analysis, with a scan step 2 s and an interscan 0.1 s/step.

3. Results and discussion

3.1. Cloning, expression and affinity-column purification of CBP from S. marcescens

Based on the gene (cbp) of S. marcescens chitin-binding protein [9–11], two oligonucleotides were designed and used as primers for PCR cloning. With the chromosomal DNA of S. marcescens as template, a DNA fragment (0.6 kb) was amplified, and the CBP was confirmed with DNA sequencing. Sequence analysis revealed that the amplified CBP gene contained an open reading frame of 594 bp encoding 197 amino-acid residues with the first 27 amino acids as a signal peptide. The full CBP gene was further subcloned into pRSET A. The resulting plasmid was transformed into the E. coli BL21 (DE3) strain for protein expression. The recombinant enzyme, expressed as a soluble form, was further purified to high homogeneity with a β-chitin matrix. To investigate the binding affinity of CBP toward chitin, the crude recombinant protein was mixed with suitable amount of chitin matrix and further loaded onto a column and extensively washed with NaCl (50 mM) before elution with various buffers under varied pH conditions (pH 4–8), that included NaOAc (50 mM, pH 4–5), phosphate (50 mM, pH 6–7) and Tris–HCl (50 mM, pH 7–8). The eluted samples were further analyzed with SDS-PAGE. The results showed that CBP was effectively eluted from the β-chitin column when the pH was less than 7. SDS-PAGE analysis revealed that the recombinant CBP was obtained with a high homogeneity (>90%) (Fig. 1). The estimated molecular mass is ~20 kDa. ESI-MS analysis gave a more precise measurement, a molecular mass 18,785 Da (Fig. 1), consistent with a theoretical calculation of the deduced protein without the signal peptide.

**Fig. 1.** SDS-PAGE and MS analysis of CBP. (A) Lane 1: protein marker; lane 2: purified CBP; (B) mass spectrum of recombinant CBP.
3.2. Construction and evaluation of the system of CBP fusion protein with peptide linkers

The expression vector was constructed based on pRSET A. The process of plasmid construction is illustrated in Fig. 2. First, the amplified cbp was inserted in NdeI and XhoI sites. The NdeI cutting site in the resulting vector was further removed to give a plasmid designated de-NdeI-pRSET/CBP, in which the sequences of various peptide linkers and a gene nase I cutting site (PGAAHY) were inserted after cbp. For further incorporation of a target gene, the NdeI site was reinserted after the gene nase I site. The resulting plasmid was ready for various insertions of genes. To evaluate the feasibility of this expression vector, a tested target protein, such as chitosanase (CNS), was inserted. The proteolytic accessibility of the gene nase I and the chitosanase activity of the fusion protein served as criteria to evaluate the feasibility of a linker; more than seven peptide linkers, as shown in Table 2, were examined. Among the tested linkers [15], fusion proteins with (EAAAK)_n, GTGEGPCCGGPGECCGTGEGPCGG, or (GGGGS)_n linker showed significant chitosanase activity. Also, all three fusion proteins could be purified with the chitin column, and further released chitosanase (∼25 kDa) after being treated with genenase I (data not shown), indicating that the peptide linkers did not perturb the structures of both proteins and that the specific proteolytic site was accessible in the fusion protein. As the (EAAAK)_3 peptide was known to form a helical structure and gave a yield and purity better than the other two fusion proteins after chitin-column chromatography, we selected it as the linker for our further work; the vector is designated pRSET/CBP-V5G.

To verify the function of this vector, we tested another two proteins, chitinase from Bacillus cereus, and LPHase from S. matensis for over-expression in E. coli. The fusion proteins, expressed in soluble form, were further purified to high homogeneity through the

Table 2
Peptides tested as linker of the CBP fusion protein.

<table>
<thead>
<tr>
<th>Linker</th>
<th>Amino-acid sequence</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AKRGWI</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>GTGEGPCGG</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>GTGEGPCGCGPGECCGTG</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>TTGLSNLTGLK</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>(EAAAK)_n, n = 2, 3, 4, 5</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>GTGEGPCGCGPGECCGTG</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>(GGGGS)_3</td>
<td>Yes</td>
</tr>
</tbody>
</table>
The binding affinity of CBP toward protein, derived from the expression of vector pRSET/CBP-V5G-LPHase, was incubated with phosphate buffer (pH 6–7, 25 °C) for 12 h. Under these conditions, the auto-cleavage of fusion protein was incomplete. As shown in Fig. 5, the SDS-PAGE analysis revealed a systematic shift in molecular mass for all fusion proteins. The protein ladder near 20 kDa might result from a separate attachment of the linker. For instance, after partial auto-cleavage, the fusion protein CBP-V2G-CNS (lane 2 in Fig. 5) exhibited four clear bands, showing with the expected molecular masses 45, 26, 20 and 19 kDa, which presumably correspond to the intact fusion protein, the PGAAHY-CNS (CNS fusion protein with genenase I recognition site), CBP-(EAAAK)2, and CBP-EAAAK, respectively. Similarly, for the case of CBP-V3G-CNS, the five protein bands shown from large to small molecular mass were expected to be the intact fusion protein, CNS linked with PGAAHY, CBP-(EAAAK)3, CBP-(EAAAK)2, and CBP-EAAAK, respectively. Other fusion proteins with longer linkers demonstrated the same auto-cleavage pattern. We further removed either the peptide sequence after partial auto-cleavage or the CBP from the fusion protein and tested the auto-cleavage feature of the proteins. The results showed that, without the existence of the CBP portion or the recognition peptide fragment of genenase I in the fusion protein, the (EAAAK)5 linker was still susceptible to auto-cleavage at pH 6.

3.4. Analysis of auto-cleavage of fusion CNS with various repeated EAAAK linkers

To understand the detailed auto-cleavage feature, we incorporated varied repeat units, (EAAAK)ₙ, with n = 2, 3, 4, 5, in the fusion protein. Sufficient fusion protein was purified and subjected to auto-cleavage assessment. All these fusion proteins were incubated at 16 °C in phosphate buffer (100 mM, pH 7.0) containing NaCl (300 mM) for 12 h. Under these conditions, the auto-cleavage of fusion protein was incomplete. As shown in Fig. 5, the SDS-PAGE analysis revealed a systematic shift in molecular mass for all fusion proteins. The protein ladder near 20 kDa might result from a separate attachment of the linker. For instance, after partial auto-cleavage, the fusion protein CBP-V2G-CNS (lane 2 in Fig. 5) exhibited four clear bands, showing with the expected molecular masses 45, 26, 20 and 19 kDa, which presumably correspond to the intact fusion protein, the PGAAHY-CNS (CNS fusion protein with genenase I recognition site), CBP-(EAAAK)₂, and CBP-EAAAK, respectively. Similarly, for the case of CBP-V3G-CNS, the five protein bands shown from large to small molecular mass were expected to be the intact fusion protein, CNS linked with PGAAHY, CBP-(EAAAK)₃, CBP-(EAAAK)₂, and CBP-EAAAK, respectively. Other fusion proteins with longer linkers demonstrated the same auto-cleavage pattern. We further removed either the peptide sequence after partial auto-cleavage or the CBP from the fusion protein and tested the auto-cleavage feature of the proteins. The results showed that, without the existence of the CBP portion or the recognition peptide fragment of genenase I in the fusion protein, the (EAAAK)₅ linker was still susceptible to auto-cleavage at pH 6.

The suggested auto-cleavage pattern of the repeated EAAAK linkers in the fusion proteins was further confirmed with ESI/MS analyses (Fig. 6). The calculated molecular masses of CBP-(EAAAK)₂, CBP-(EAAAK)₃, CBP-(EAAAK)₄, CBP-(EAAAK)₅, CBP-(EAAAK)₆ are 20,996, 20,526, 20,056, 19,586 and 19,115 Da, respectively, whereas PGAAHY-CNS has 24,430 Da. As shown in Fig. 6, the m/z signals in each spectrum correlate well with the expected CBP-(EAAAK)ₙ portion and the PGAAHY-CNS. For the CBP/CNS fusion protein, to cleave at the EAAAK unit adjacent to CNS seems more efficient because there was no EAAAK-linked CNS observed with ESI-MS analysis throughout the reaction. The linker continued to be cleaved from the CBP-(EAAAK)ₙ portion, releasing a unit of (EAAAK) peptide (molecular mass 470 Da) at a time. Although the mechanism of the auto-cleavage is unclear, the auto-cleavage is likely chemically controlled without an involvement of protease digestion. More extensive work is necessary to solve the enigma of the intrinsic property of auto-cleavage.
Fig. 5. SDS-PAGE analyses of linkers auto-cleavage. The fusion proteins were incubated in phosphate buffer (pH 6.0 at 16°C) so that partial auto-cleavage was obtained. Lane 1: protein marker; lane 2: CBP-V2G-CNS; lane 3: CBP-V3G-CNS; lane 4: CBP-V4G-CNS; lane 5: CBP-V5G-CNS; lane 6: CBP-V5-CNS (protein expressed without the peptide sequence of the genenase I cutting site); lane 7: V5G-CNS (protein expressed without CBP).

Fig. 6. ESI-MS analyses of fusion proteins after auto-cleavage. The samples used for mass analyses were those for SDS-PAGE analyses shown in Fig. 4. (A) CBP-V2G-CNS, (B) CBP-V3G-CNS, (C) CBP-V4G-CNS, (D) CBP-V5G-CNS, (E) CBP-V5-CNS, and (F) V5G-CNS. The molecular mass (MM) of the EAAAK peptide is 470 Da.

4. Conclusion

In summary, this study demonstrated a convenient system for protein purification with a re-usable and low cost chitin matrix. The repeated EAAAK linker provides a unique auto-cleavage property that commits an effective way to obtain protein of interest from fusion protein without protease treatment. It is potential useful for cost effective protein purification in large scale and the chitin matrix can be reused for at least 6 times without losing the function.

Acknowledgment

The authors are grateful to the National Science Council of Taiwan and the MOE program of Ministry Education for financial support.
References