Simple protein purification through affinity adsorption on regenerated amorphous cellulose followed by intein self-cleavage

Jiong Hong, Yiran Wang, Xinhao Ye, Y.-H. Percival Zhang

Introduction

Developing simple, low-cost, and environmentally friendly methods for recombinant protein purification on a large scale remains challenging [1,2]. Affinity chromatography by using various affinity tags on various resins is popular in laboratories and biotechnology companies [1,3,4], but it cannot be applied to low-selling-price non-therapeutic proteins, such as industrial enzymes [5]. A low-cost and scalable method for large-scale protein purification is prerequisite for commercialization of industrial scale biocommodity production mediated by enzymes or synthetic enzymatic pathways [6–8].

A cellulose-binding module (CBM) tag has been used for recombinant protein purification on commercial cellulose matrix or powder (Avicel, microcrystalline cellulose or SigmaCell) [9–14], because of (i) highly specific binding for the CBM-tag protein, (ii) low non-specific binding for other proteins, (iii) low-cost affinity matrix (cellulose), (iv) enhanced protein folding [13], and (v) increased protein yields [14]. But commercial crystalline cellulose is a low binding capacity porous matrix, most of whose binding surface is internal [15]. The bound protein cannot be removed efficiently due to enzyme entrapment effect [16], resulting in lower protein recovery yields. In addition, protease could not efficiently work on entrapped protein to release cleaved target proteins [15].

Regenerated amorphous cellulose (RAC), which is made from Avicel through phosphoric acid dissolution followed by regeneration [17], has a greater than 20-fold surface area of Avicel [15]. In addition, the entire binding surface of RAC is externally accessible to the target protein, and protein binding on RAC is faster than on Avicel.

Inteins – protein introns – can excise themselves and/or rejoin two fragments together [18,19]. In order to avoid using costly peptide-specific protease and simplify purification process, self-cleavage intein can be used to replace costly peptide-specific proteases through the changes in pH or thiol reagent concentration [18–20].

Here we developed a generic, low-cost, scalable, protein purification method based on affinity adsorption on a low-cost, biodegradable, high adsorption capacity adsorbent, RAC. The high-
purity target proteins can be separated through self-cleavage by intein that linked CBM and target protein.

2. Materials and methods

2.1. Chemicals and strains

All chemicals were reagent grade, purchased from Sigma–Aldrich (St. Louis, MO, USA) and Fisher Scientific (Pittsburgh, PA, USA), unless otherwise noted. Microcrystalline cellulose – Avicel PH105 (20 μm) – was purchased from FMC (Philadelphia, PA, USA). Escherichia coli DH5α was used as a host cell for DNA manipulation. E. coli BL21 Star (DE3) (Invitrogen, Carlsbad, CA, USA) containing a protein expression plasmid was used for producing the recombinant protein. Luria–Bertani (LB) medium was used for E. coli growth and protein expression with 100 μg/mL ampicillin. Clostridium thermocellum genomic DNA was gifted from Dr. Mielenz at Oak Ridge National Laboratory (Oak Ridge, TN, USA). The oligonucleotides were synthesized by Integrated DNA Technologies (Coraville, IA, USA) (Table 1).

2.2. Regenerated amorphous cellulose preparation [17]

Approximately 0.2 g of microcrystalline cellulose (FMC PH-105) was added to a 50-mL centrifuge tube, and 0.6 mL distilled water was added to wet the cellulose powder to form a cellulose-suspended slurry. Ten millilitres of ice-cold 86% H3PO4 (i.e., commercial 85% grade) was slowly added to the slurry with vigorous stirring so that the final phosphoric acid concentration was approximately 83.2%. Before the last 2 mL of phosphoric acid was added, the cellulose suspension solution was evenly mixed. The cellulose mixture turned transparent after addition of last 2 mL concentrated phosphoric acid within several minutes, and stood for ca. an hour on ice with occasional stirring. Approximately 40 mL of ice-cold water was added at a rate of approximately 10 mL per addition with vigorous stirring between additions, resulting in a white cloudy precipitate. The precipitated cellulose was centrifuged at ∼10,000 × g and 4 °C for 20 min. The pellet was suspended by ice-cold water, followed by centrifugation to remove the supernant containing phosphoric acid four times. Approximately 0.5 mL of 2 M Na3CO3, and 45 mL of ice-cold distilled water were used to suspend the cellulose pellet. After centrifugation, the pellet was suspended and centrifuged by distilled waters twice or until pH 5–7. The carbohydrate concentration of RAC was calibrated by the phenol–H2SO4 method [21]. No detective amount of cellulose (<1 wt.%) was lost during the treatment most times. Sigmacell or other cellulose powders can be used to replace FMC PH105. The RAC slurry can be stored during the treatment most times. Sigmacell or other cellulose powders can be used to replace FMC PH105. The RAC slurry can be stored

2.3. Recombinant protein expression plasmids

The pCIG plasmid encoding the CBM-intein-GFP (CIG) fusion protein was constructed based on the New England Biolabs plasmid pTWIN1 (Ipswich, MA USA). Three cbm, SnpDnaB intein, and gfp DNA fragments amplified by PCR amplification by the primers of CBM-F1/CBM-R2, InTein-F3/InTein-R3, and GFP-F2/GFP-R1, followed by double digestion by NdeI/Stul, Stul/Xhol and Xhol/PstI, respectively, were ligated into the NdeI/PstI digested pTWN1 in one step to produce the pCIG plasmid. The three other target genes included the C. thermocellum cellulobiose phosphorylase (CBP) [22,23], phosphoglucomutase (PGM, CT1265), and putative α-glucan phosphorylase (GNP, CT0932).

2.4. Recombinant protein expression

The protein expression plasmids were transformed into the strain E. coli BL21 (DE3). Two hundred millilitres of LB medium supplemented with 100 μg/mL ampicillin in 1-L Erlenmeyer flasks were incubated with a rotary shaking rate of 160 rpm at 37 °C until an A600 reached between ∼0.6 and 0.8. The recombinant protein expression was induced by adding IPTG (0.20 mM, final), and then the cultures were incubated at the decreased temperature of 18 °C for 9–12 h. The cells were harvested by centrifugation at 4 °C, washed once by 50 mM Tris–HCl buffer (pH 8.5), re-suspended by ∼30 mL of 50 mM Tris–HCl buffer (pH 8.5). The cell pellets were lysed in an ice bath by ultra-sanitation by Fisher Scientific Sonic Dismembrator Model 500 (3-s pulse, total 90 s, at maximum strength). After centrifugation, the supernatant of cell lysate was used for protein purification.

2.5. Self-cleavage efficiency of intein at different temperature and pH

1.2 mL of 10 mg of RAC/mL was mixed with 8 mL of cell lysate containing 0.246 mg CIG/mL at room temperature for 30 min. After centrifugation and washing once in 6 mL of 50 mM Tris–HCl, 0.5 M NaCl, and 1 mM EDTA buffer (pH 6.5 or 7.0) at 4 °C, the RAC pellet with adsorbed CIG was suspended in 8 mL of 50 mM Tris–HCl, 0.5 M NaCl, and 1 mM EDTA buffer (pH 6.5 or 7.0). The suspension solution was incubated at different temperatures (4 °C, 18 °C, 23 °C, 30 °C, 40 °C, 50 °C, and 60 °C). The fluorescence of cleaved GFP in the supernatant was measured and calculated for self-cleavage efficiency.

2.6. Protein purification

Thirty-five millilitres of cell lysate (1.65 mg of crude protein containing 0.33 mg of CIG per mL) was mixed with 5.25 mL of 10 mg RAC/mL at room temperature for 30 min. After centrifugation, the pellet was suspended in 27 mL of 50 mM Tris–HCl buffer (pH 6.5) containing 0.5 M NaCl and 1 mM EDTA. After centrifugation, a 35 mL of the cleavage buffer (50 mM Tris–HCl, 0.5 M NaCl, and 1 mM EDTA buffer, pH 6.5) was added to suspend the pellet. After incubation at 40 °C overnight or for 4–8 h followed by centrifugation, the cleaved GFP was obtained in the supernatant. For heat-labile proteins, lower temperatures (e.g., 4 °C or 18 °C or room temperature) could be used for intein self-cleavage. Similarly, the crude proteins from the cell
lysate containing CBP, PGM, and GNP were purified by RAC adsorption followed by intein cleavage. The amount of target protein in the lysate was roughly estimated by the enzyme activity or sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The amount of RAC addition can be estimated at a ratio of 200 mg of target protein to 1 g of RAC.

2.7. Enzyme assays and protein analysis

GFP fluorescence was detected (excitation at 485 nm, emission at 528 nm) by the Bio-Tek multi-detection microplate reader (Winooski, VT, USA). CBP activity was measured as described elsewhere [22–26]. Putative α-glucan phosphorylase activity was determined by measuring the increase in inorganic phosphate from 10 mM soluble starch to 10 mM glucose 1-phosphate in a 50-mM HEPES buffer (pH 7.0) at 37 °C. One unit of CBP or GNP activity was defined as the amount of enzyme that produced 1 μmol of phosphate per 15 min under the above conditions. PMG activity was measured in a 50 mM Hepes buffer (pH 7.5) with 5 mM glucose-1-phosphate, 0.5 mM Mn²⁺, 5 mM Mg²⁺, 1 mg/mL BSA, and 0.1% Triton X-100 at 60 °C. The protein mass concentration was determined by the Bradford method based on a standard protein of bovine serum albumin. 12% SDS-PAGE was performed in the Tris–glycine buffer as described elsewhere [27].

3. Results

A simple protein separation method was proposed based on high-affinity adsorption of the CBM-tag protein on the surface of RAC (Fig. 1). RAC has a much higher binding capacity (e.g., 365 mg of CBM-GFP per gram of RAC) than that of crystalline cellulose (14.8 mg of CG per gram of Avicel) and than those of any commercial protein purity resins (e.g., 10–40 mg of protein per gram of resin). The dissociation constant based on RAC is 168 L/g RAC, indicating high affinity adsorption of the family three CBM on RAC. The binding capacity based on the RAC volume is approximately 15 mg of protein per mL of bed volume. The crude protein solution containing the fusion proteins of CBM-intein-target protein was mixed well with the RAC absorbent. After centrifugation, the supernatant containing most of impure proteins was decanted. The impure proteins that stayed in the porous RAC were washed away in a buffer once or several times. The cleaved soluble target protein can be obtained through intein self-cleavage in the cleaving buffer (low pH and high salt concentration). After centrifugation, the purified cleaved target protein can be obtained in the supernatant. The above protein purification processes involving only solid/liquid separation (centrifugation or filtration) can be easily scaled up.

The protein purification process for the GFP-intein-CBM fusion protein was optimized. One gram of RAC has a maximum binding capacity of 365 ± 20 mg of CIG, determined by the Langmuir isotherm as described elsewhere [15,28]. In order to bind ~95% of the target protein, a 1.7-fold overload of RAC was needed (i.e., 215 mg of adsorbed protein/g of RAC). After centrifugation, the impure protein in the supernatant was decanted. The RAC pellet with the adsorbed CIG was suspended in the buffer once or twice to remove the impure protein remaining in the adsorbent matrix.

Self-cleavage efficiency on the surface of RAC was investigated at different pH and temperature. It was found that cleavage rates at pH 6.5 were much faster than those at pH 7.0 regardless of temperatures (data not shown). Fig. 2 shows self-cleavage efficiency at pH 6.5 and various temperatures. The highest cleavage rates were obtained at 40 °C. The efficiencies were 82% at hour 4 and 90% at hour 12. Higher temperatures (e.g., 50 °C and 60 °C) resulted in lower efficiency due to intein denaturation. Lower temperatures (e.g., 23 °C or 4 °C) need longer reaction time for cleavage and the efficiencies were relatively low. Sometimes low-temperature self-cleavage must be conducted for non-thermostable protein purification [20].
Fig. 3. Images of the cleaved GFP purification by RAC adsorption followed by intein self-cleavage. (1) The cell lysate of E. coli BL21 (pCIG); (2) cell lysate + RAC after centrifugation; (3) the pellet—adsorbed target protein on RAC; (4) the supernatant after centrifugation; (5) the pellet suspended and precipitated in ~0.9 mL of the washing buffer (pH 8.0); (6) the washed RAC pellet; (7) used washing buffer; (8) the pellet suspended in the cleavage buffer (pH 6.5); (9) the RAC pellet with a small amount of the non-cleavage CIG; (10) the cleaved GFP in the supernatant; (11) once-washed RAC pellet; (12) RAC in 50 mM Tris buffer (negative control-1); (13) RAC in water (negative control-2) and (14) 50 mM Tris buffer (negative control-3) (all images were taken where the pHs were or were adjusted to ~8.0).

Fig. 3 shows the GFP protein allocation during each step of the purification process. After sonication, the cell lysate containing CBM-intein-GFP showed a strong green fluorescence signal (tube 1). After RAC addition, CIG adsorption on RAC, and centrifugation, the CIG was bound in the RAC pellet (tube 2 and tube 3). The supernatant that did not contain CIG had no fluorescence (tube 4). The pellets were washed in 0.9 mL of the buffer (pH 8.0) (tubes 5, 6 and 7). The RAC pellets were suspended in the cleavage buffer (pH 6.5) (tube 8). After incubation at 40 °C for several hours, the cleaved GFP was released into the supernatant (tube 10), and the RAC still adsorbed some non-cleavage CIG (tube 11). The cleaved GFP yield was 69% and the purified fold was 6.15 (Table 2). Fig. 4 shows the SDS-PAGE analysis results at key steps of protein purification process. The purified GFP after self-cleavage and centrifuge in the supernatant was the sole band (lane 8). Fig. 5 presents the SDS-PAGE analysis for purifying the recombinant phosphoglucomutase from C. thermocellum by using the similar method.

In addition, three other enzymes CBP, GNP, and PGM were purified using the similar methods (Table 3). The relatively lower yields can be further enhanced by optimization of adsorbent addition amount, cleavage conditions, washing conditions, and so on.

4. Discussion

On the laboratory level, product purity, yield, and costs, as well as purification speed are several important considerations. Affinity chromatography is widely used in laboratories, but is prohibited for large-scale protein purity because of batch operation, complex scale-up, low throughput, low adsorption capacity of affinity resin, slow association/dissociation rates of the target protein on resins, potential column fouling, and flow-rate limitations. On the industrial scale, process scalability and waste treatment are two additional factors.

Utilization of the ultra-high-binding capacity adsorbent RAC for capturing CBM-tag proteins has several distinctive advantages:

(1) fairly high protein purity (Figs. 4 and 5);
(2) reasonably high protein yield using ultra-high binding capacity RAC (Table 3), where high protein yields were implemented to efficiently wash the target protein in the matrix by multi-step washing;

Table 2
The optimized CBM-intein-GFP protein purification process.

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Volume (mL)</th>
<th>Sp. fluorescence (mL−1)</th>
<th>Total fluorescence</th>
<th>Protein (mg/mL)</th>
<th>Total protein (mg)</th>
<th>Sp. fluorescence</th>
<th>Yield (%)</th>
<th>Purification fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soluble cell lysate</td>
<td>20</td>
<td>4.097</td>
<td>81.933</td>
<td>1.84</td>
<td>36.79</td>
<td>2.227</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Adsorbed protein</td>
<td>2.4</td>
<td>3.303</td>
<td>60.583</td>
<td>0.241</td>
<td>4.15</td>
<td>4.883</td>
<td>74</td>
<td>2.19</td>
</tr>
<tr>
<td>Truncated GFP</td>
<td>17.2</td>
<td>3.303</td>
<td>56.817</td>
<td>0.241</td>
<td>4.15</td>
<td>13.706</td>
<td>69</td>
<td>6.15</td>
</tr>
</tbody>
</table>

* The adsorbed protein is calculated based on the difference between the initial protein used the supernatant protein after RAC adsorption.
(3) shorter purification time due to faster protein adsorption and intein cleavage;
(4) ultra-low-cost absorbent RAC (e.g., $\sim$2 US cents per gram of RAC on the laboratory scale and as low as 0.02 US cents per gram of purified protein when large scale RAC production is achieved [29]). Given 200 mg of the bound target protein per gram of RAC, the fraction of protein purification costs based on RAC resin could be as low as 10 or 0.1 US cents per gram of purified protein now or in the future, respectively. For large scale recombinant protein production and separations, fermentation costs, target protein fermentation titer, and buffer cost will be responsible for the major fractions of production costs;
(5) simple solid/liquid separation (centrifugation or filtration), which can be scaled up easily and does not need costly instruments; and
(6) minimal waste treatment because of non-toxic chemicals (such as heavy metal ions) consumed or needed and use of biodegradable and abundant absorbent—RAC.

In a word, a simple protein purification method by using the low cost ultra-high capacity absorbent RAC has been developed to capture CBM-tag proteins. Four cleaved target proteins by intein self-cleavage were obtained by a simple solid/liquid unit operation, centrifugation. This new technology will play an important role in future biocommodity production catalyzed by low-cost enzymes.

Acknowledgments

This work was supported primarily by the United States Department of Agriculture and partially by the Institute for Critical Technology and Applied Science (ICTAS) and the Air Force Young Investigator Award.

References


Table 3
The purified protein quantification and activity assay

<table>
<thead>
<tr>
<th>Protein</th>
<th>MW (kDa)</th>
<th>Purified protein (mg/200 mL broth)</th>
<th>Specific activity</th>
<th>Yield (%)</th>
<th>Purification fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>27</td>
<td>2.1</td>
<td>13170/mg</td>
<td>69</td>
<td>6.15</td>
</tr>
<tr>
<td>CBP</td>
<td>93</td>
<td>0.1</td>
<td>458 U/mg</td>
<td>22</td>
<td>31.8</td>
</tr>
<tr>
<td>PGM</td>
<td>65</td>
<td>3.0</td>
<td>644 IU/mg</td>
<td>21</td>
<td>2.60</td>
</tr>
<tr>
<td>GNP</td>
<td>98</td>
<td>0.8</td>
<td>1732 U/mg</td>
<td>60</td>
<td>2.70</td>
</tr>
</tbody>
</table>