Title

High-throughput purification of recombinant proteins using self-cleaving intein tags

Authors

Coolbaugh, M.J.¹, Shakalli Tang, M.J.², and Wood, D.W.^{2,*}

1. Process Development, Sanofi-Genzyme, 31 New York Ave, Framingham, MA 01701, USA
2. Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210

*Corresponding Author: Email: wood.750@osu.edu; Phone: +1-614-292-9636; Fax: +1-614-292-3769

Abstract

High throughput methods for recombinant protein production using *E. coli* typically involve the use of affinity tags for simple purification of the protein of interest. One drawback of these techniques is the occasional need for tag removal before study, which can be hard to predict. In this work, we demonstrate two high throughput purification methods for untagged protein targets based on simple and cost-effective self-cleaving intein tags. Two model proteins, *E. coli* beta-galactosidase (β Gal) and superfolder green fluorescent protein (sfGFP), were purified using self-cleaving versions of the conventional chitin-binding domain (CBD) affinity tag and the nonchromatographic elastin-like-polypeptide (ELP) precipitation tag in a 96-well filter plate format. Initial tests with shake flask cultures confirmed that the intein purification scheme could be scaled down, with >90% pure product generated in a single step using both methods. The scheme was then validated in a high throughput expression platform using 24-well plate cultures followed by purification in 96-well plates. For both tags and with both target proteins, the purified product was consistently obtained in a single-step, with low well-to-well and plate-to-plate variability. This simple method thus allows the reproducible production of highly pure untagged recombinant proteins in a convenient microtiter plate format.

Subject Category

Protein Purification Methods

Keywords

Intein Tag; High Throughput Protein Purification; Self-cleaving Tag; Elastin-Like Polypeptide; Affinity Methods; Recombinant Protein Purification

Abbreviations

CBD, chitin-binding domain; ELP, elastin-like-polypeptide; β Gal, beta-galactosidase; sfGFP, superfolder green fluorescent protein; IMAC, immobilized metal affinity chromatography; LIC, ligation-independent cloning

1 Introduction

One of the most commonly used methods for the purification of recombinant proteins produced in *E. coli* is the poly-histidine (His) tag combined with immobilized metal affinity chromatography (IMAC) [1]. Various His-tag configurations have been applied to high-throughput protein purification schemes, and this platform has become the primary method for protein purification for structural biology [2-7]. In particular, these approaches have several traits that make them especially appealing to structural biology labs: the method is general and can be applied to any protein of interest; the method is simple, requiring only a single column; the method is well-established and all required reagents are commercially available; and the His-tag is very small and often does not need to be removed in order to obtain an accurate structure [8]. However, in some cases the addition of a His-tag to a protein of interest has been shown to adversely affect the solubility of the protein, while in other cases the activity of the target protein can be affected [9, 10]. This is especially troublesome for proteins that express poorly in *E. coli*, including many human proteins. Although it is possible to recover insoluble proteins from inclusion bodies, refolding significantly increases the complexity of the process. One of the most commonly used ways to increase the production of soluble protein in *E. coli* is the use of large solubility tags, such as the maltose-binding protein (MBP) or N utilization substance A (NusA) [11]. The addition of these large, very soluble proteins to the N-terminus of a protein that is difficult to express often leads to increased soluble expression of the target. However, these solubility-enhancing tags introduce the requirement for an additional tag removal step before the native protein can be used for structure determination.

When using large, complex tags, the tag removal step becomes a critical component of downstream processing. The most commonly used method for tag removal involves the use of specific proteases, such as Tobacco Etch Virus (TEV) or Human Rhinovirus 3C (HRV3C) protease, among others, that cleave between the tag and the protein of interest [12, 13]. To simplify the construction of various expression vectors, a protease recognition site is typically included in most commercial expression vectors. Once the tagged target is purified, a protease is added to cleave the tag, and then removed in a subsequent affinity chromatography step. While this protease-based tag removal system is widely used, it is not without its drawbacks. Many highly specific proteases are very costly [14], and the requirement for protease removal adds to the complexity of the purification system. For these and other reasons, alternatives to protease-based tag removal have been developed.

One such protease-free method for tag removal involves the use of self-cleaving intein tags [15]. Inteins are naturally occurring self-splicing proteins that have been engineered to undergo isolated C or N-terminal cleavage [16]. In these systems, the protease recognition sequence is replaced by an intein, creating a self-cleaving tag module. After initial capture of the fusion protein, intein self cleavage is induced by the addition of a thiol group such as dithiothreitol (DTT) or by a change in pH or temperature [17]. The untagged, pure target protein can thus be eluted from the column in a single step. A number of different groups have reported successful use of intein-based systems for protein purification [18-21].

The use of the Δ I-CM intein, derived from the *Mycobacterium tuberculosis* (*Mtu*) RecA intein, has been well documented for simple purification and tag removal of proteins expressed in *E. coli* [22-27]. The

major advantage of the Δ I-CM intein is that the cleavage reaction is induced by a simple pH shift, rather than addition of thiol groups. This allows for the purification of proteins containing disulfide bonds, including antibody fragments [28]. The Δ I-CM intein has been used in conjunction with a number of different tags, including the affinity chitin-binding domain (CBD) tag and the resin-free elastin-like polypeptide (ELP) precipitation tag [29, 30]. Recent work has also incorporated the Δ I-CM intein into several ligation-independent cloning (LIC) vectors for simple and rapid cloning of target protein genes into the expression vector [31].

In this work, we demonstrate the efficacy of using the Δ I-CM self cleaving intein tag system to purify *E. coli* expressed proteins in a high-throughput, 96-well plate format. The Δ I-CM system is demonstrated using both a conventional affinity tag (CBD) as well as the resin-free ELP tag. The system was used to purify two model proteins, *E. coli* β -galactosidase (β Gal) and superfolder GFP (sfGFP). This work serves as the first demonstration of the use of self-cleaving inteins for high throughput protein purification as well as the first demonstration of the ELP precipitation tag in a 96-well plate format.

2 Materials and Methods

2.1 Vectors

For expression of β Gal fusion proteins, the vectors pET/CI- β Gal-LIC and pET/EI- β Gal-LIC, which contain the chitin binding domain (CBD) tag or elastin-like polypeptide (ELP) tag fused to the delta I-CM intein and *E. coli* β Gal, were constructed using ligation-independent cloning, as described previously [31]. For the expression of GFP fusion proteins, the gene encoding superfolder GFP (sfGFP) was amplified from the vector pET/sfGFP-LIC (Addgene plasmid 29772) using primers; F 5'-AT CGT GTA CAC AAC ATG CAG GGT ATG GTG AGC AAG GGC GAG GAG-3' and R 5'-GTA CAA GCT TGC CTG CAG TTA CTA CTT ATA CAG CTC GTC CAT G-3'. The forward primer contains the last four bases of the intein C-terminus (CAAC) as well as the first three amino acids of the *E. coli* RNA polymerase A. These additional amino acids increase the intein cleavage rate for the GFP target protein (data not shown), while the β Gal fusion protein contains no additional amino acids at the N-terminus of the β Gal protein. The resulting PCR product was cloned into previously constructed vectors (pET/ClaFGF and pET/EIMBP) using unique BsrGI and HindIII restriction sites. The resulting vectors were transformed into *E. coli* strain BLR (Novagen) for expression.

2.2 Protein Expression in Shake Flasks

Single isolated colonies were used to inoculate 3 ml inoculum cultures in LB medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, pH 7.0) supplemented with 100 μ g/ml ampicillin. The inoculum cultures were shaken at 250 RPM overnight (18-20 hours) at 37°C, and then diluted 1:200 into 200 ml of prewarmed (37°C) expression media supplemented with 100 μ g/ml ampicillin in a 1 liter baffled flask. Expression medium for the ELP fusion constructs was Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 2.31 g/l KH₂PO₄, 12.54 g/l K₂HPO₄), while the expression medium for CBD constructs was 2xLB (20 g/l tryptone, 10 g/l yeast extract, 10 g/l NaCl). The expression culture was shaken at 140 RPM for

approximately 2.5 hours, until the optical density of the cultures at 600 nm (OD_{600}) was between 0.5 and 0.8. The culture was then incubated in a 20°C shaking water bath at 140 RPM for 20 minutes. Target protein expression was then induced by addition of IPTG to a final concentration of 0.5 mM. Expression was carried out at 16°C, 140 RPM, for 20-24 hours post induction. After expression, cultures were chilled at 4°C for 30 minutes before harvest. Cultures were harvested by centrifugation at 5000xg for 10 minutes. Cell pellets were resuspended in 10% of the original culture volume of appropriate lysis buffer. For ELP fusions, this was 20 mM Tris, 2 mM EDTA, pH 8.5. For CBD fusions, this was 20 mM AMPD, 20 mM Bis-Tris, 500 mM NaCl, 2 mM EDTA, pH 8.5. After resuspension, lysozyme was added to a final concentration of 0.2 mg/ml and cells were frozen at -20°C for at least 24 hours.

2.3 Protein Expression in 24-Well Microtiter Plates

Overnight cultures were prepared in LB medium as described above. The overnight cultures were then diluted 1:100 into prewarmed expression media as described above, supplemented with 100 µg/ml ampicillin. The inoculated expression medium was then distributed evenly between two or four sterile 24-well plates at a volume of 2 ml per well. The 24-well plates were sealed with a breathable plate seal (AeraSeal, Excel Scientific) and agitated in a conventional shaking incubator (25.4 mm orbital diameter) at 250 RPM at 37°C for 3-4 hours, until visibly turbid (OD_{600} of approximately 1). The temperature was then reduced to 16°C for approximately 30 minutes, followed by addition of IPTG to a 1 mM final concentration. Expression was carried out at 16°C and 250 RPM agitation for 20-24 hours post induction. After expression, cultures were chilled at 4°C for 30 minutes before harvest. Cultures were harvested by centrifugation at 3000xg for 10 minutes. Cell pellets were resuspended in 10% of the original culture volume (200 µl) of appropriate lysis buffer. For ELP fusions, this was 20 mM Tris, 2 mM EDTA, pH 8.5. For CBD fusions, this was 20 mM AMPD, 20 mM Bis-Tris, pH 8.5. To each cell suspension, a 1:10 dilution (20 µl) of 10X BugBuster Reagent (Novagen) was added. Cell suspensions were transferred from the two 24-well plates to columns 1-6 of a 96-well plate. Lysozyme (Sigma-Aldrich) was then added to a final concentration of 0.2 mg/ml and cells were frozen at -20°C for at least 24 hours.

2.4 Sample Preparation

2.4.1 Shake Flask Cultures

Cell suspensions recovered from shake flasks were thawed at 20°C and then maintained at 4°C for the remainder of the purification procedure until intein cleaving. Thawed cell suspensions were lysed via sonication, using approximately 5 cycles of 30 seconds each at 4-5 Watts RMS. DNase 1 (Sigma Aldrich) and $MgCl_2$ were added final concentrations of 10 µg/ml and 10 mM, respectively, to degrade residual DNA and reduce sample viscosity. DNase treatment was carried out at 4°C for 15 minutes as per manufacturer's instructions. Samples were then clarified by centrifugation at 9,600xg for 15 minutes at 4°C. Clarified samples were finally filtered through 0.22 µm filters to remove any residual cell debris.

2.4.2 24-well Plate Cultures

Cell suspensions in 96-well microtiter plates were thawed at 20°C and gently agitated (~500 RPM on a titer plate shaker, orbital diameter 3 mm) for 10 minutes at 4°C to aid in chemical lysis. Chemical lysis

was visually confirmed by increased viscosity of the well samples. To remove residual DNA and reduce sample viscosity, DNase 1 and MgCl_2 were added to final concentrations of 10 $\mu\text{g}/\text{ml}$ and 10 mM, respectively. DNase treatment was carried out at 4°C for 15 minutes with gentle agitation at 4°C. Initial clarification was carried out by centrifugation at 3220xg (the maximum speed for the microtiter plate rotor) for 15 minutes at 4°C. The supernatant from this centrifugation was then applied to a lysate clarification filter plate (3 μm GF, 0.2 μm SUPOR; Pall Life Sciences) to remove residual cell debris.

2.5 CBD Affinity Tag Purification

Clarified samples were diluted 1:5 in CBD dilution buffer (25 mM AMPD, 25 mM Bis-Tris, 2.5 mM EDTA, 625 mM NaCl, pH 8.5), to produce a final buffer composition of 20 mM AMPD, 20 mM Bis-Tris, 2 mM EDTA, 500 mM NaCl, pH 8.5 (CBD Equilibration Buffer). 96-well filter plates (1.2 μm Supor membrane; Pall Life Sciences) were charged with 25 μl chitin beads (NEB). The chitin resin was washed two times with 200 μl of CBD Equilibration Buffer, where each wash was for 5 minutes, followed by centrifugation at 750xg for 2 minutes to recover the washed resin. To each washed resin pellet, 200 μl of diluted clarified lysate was added and incubated at 4°C and 750 RPM agitation for 30 minutes on a microtiter plate shaker to allow binding of the CBD tag to the chitin resin beads. After incubation, the plates were centrifuged for 2 minutes at 750xg and 4°C. The supernatant was collected for subsequent SDS-PAGE analysis. An additional 200 μl of CBD equilibration buffer was then added to the CBD resin pellet and incubated for 5 minutes at 4°C and 750 RPM to wash out any unbound proteins. Plates were then centrifuged for 2 minutes at 750xg and 4°C and resin samples were collected for SDS-PAGE analysis. The resin was then resuspended in 200 μl of CBD cleavage buffer (20 mM AMPD, 20 mM Bis-Tris, 2 mM EDTA, 500 mM NaCl, pH 6.2) and agitated for 5 minutes at 4°C and 750 RPM. The resin was then centrifuged at 750xg for 2 minutes and a resin sample was taken for subsequent SDS-PAGE analysis. The resin pellets were then again resuspended in an additional 200 μl of CBD cleaving buffer, and the suspensions were incubated at 750 RPM agitation overnight at 25°C (room temperature) to allow for intein self-cleavage. After the overnight incubation, eluted target protein was collected in the supernatant after centrifugation of the resin at 750xg and 4°C for 2 minutes. The chitin resin was regenerated by washing with 200 μl of 1% SDS for 20 minutes at 20°C, and a regeneration sample was collected via centrifugation for analysis of cleaving by SDS-PAGE analysis. Finally, the resin was regenerated by washing three times with 200 μl water, followed by two washes with 200 μl CBD Equilibration Buffer.

2.6 ELP Precipitation Tag Purification

Clarified samples were diluted 1:2 for shake flask cultures, or 1:2.5 for 24-well cultures, in ELP lysis buffer (20 mM Tris, 2 mM EDTA, pH 8.5), to ensure sufficient sample volume in each well. To each well of a 96-well filter plate (1.2 μm Supor Membrane), 180 μl of diluted lysate sample was added, followed by addition of 120 μl of ELP lysis buffer supplemented with 2 M ammonium sulfate, to give a final concentration of 0.8 M ammonium sulfate in each well. The plates were agitated at 20°C for 20 minutes and 750 RPM on a microtiter plate shaker to induce ELP tag precipitation. The plates were then centrifuged at 1500xg and 20°C for 2 minutes to pellet the ELP-tagged target protein, and the supernatant containing the soluble contaminant proteins was collected for subsequent SDS-PAGE

analysis. Each pellet on the microtiter plate was then washed by addition of 150 μ l of ELP lysis buffer supplemented with 0.8 M ammonium sulfate, and agitation at 750 RPM and 20°C for 5 minutes. The plates were then centrifuged at 1500xg for 2 minutes and supernatant (wash) was collected for subsequent SDS-PAGE analysis. The washed ELP pellets were then dissolved in 90 μ l of ELP Cleavage Buffer (1xPBS supplemented with 40 mM Bis-Tris, pH 6.2) with shaking at 4°C for 15 minutes. After dissolution, the plate was again centrifuged at 1500xg and 4°C for 2 minutes to remove any irreversibly aggregated contaminant proteins. This additional centrifugation has been shown to significantly increase the purity of the target with minimal impact on yield. The supernatant containing the soluble ELP precursor was recovered to a new microtiter plate and incubated overnight (approximately 20 hours) in ELP Cleavage Buffer to induce intein cleavage. After overnight incubation, 60 μ l of ELP Cleavage Buffer supplemented with 2 M ammonium sulfate was added to the samples and transferred to a 96-well filter plate. The plates were agitated at 20°C for 20 minutes at 750 RPM on a microtiter plate shaker to precipitate the cleaved ELP tag. The purified target protein was collected in the supernatant after centrifugation at 1500xg and 20°C for 2 minutes. Finally, 150 μ l of ELP Lysis Buffer was added to each well and incubated at 4°C for 5 minutes with agitation to redissolve the cleaved ELP tag for subsequent analysis via SDS-PAGE.

2.7 SDS-PAGE

Protein samples were diluted 1:1 in 2X Laemmli sample buffer with β -mercaptoethanol and denatured by heating at 95°C for 5 minutes. Samples were then fractionated in 12% Tris-Glycine gels or in 8-16% Precise Tris-Glycine gels (Pierce Biotechnology). After electrophoresis, samples were stained using Coomassie Brilliant Blue R-250.

2.8 Protein Quantification

Protein quantification was carried out using the Coomassie Plus Protein Assay kit (Pierce Biotechnology), per manufacturer instructions. Briefly, 140 μ l of Coomassie Plus reagent was added to 10 μ l of sample in a titer plate well then mixed at 600 RPM on a titer plate shaker for 30 seconds. The reaction was incubated at 20°C for 10 minutes to allow for color development, followed by measurement of the absorbance at 595 nm using a microtiter plate reader. The protein concentration was interpolated from a standard curve comprised of known amounts of bovine serum albumin (BSA). All samples were measured in triplicate.

2.9 Activity Assays

2.9.1 Beta-Galactosidase (β Gal)

β Gal activity was measured using a beta-galactosidase assay kit (Stratagene). The micro-assay format was used where the change in absorbance at 420nm is tracked over time. One unit of activity was defined as 1 nmol of *ortho*-nitrophenyl- β -galactoside (ONPG) hydrolyzed per minute at 37°C. Measurements of OD₄₂₀ were divided by 0.0045 to convert to nmol of hydrolyzed ONPG per ml reaction volume. The resulting concentrations were then plotted versus time, and the slope determined by curve fit using Microsoft Excel.

2.9.2 Superfolder Green Fluorescent Protein (sfGFP)

GFP activity was measured by measuring the fluorescence at 520 nm when excited at 490 nm. All samples were diluted 1:40 in 100 mM Phosphate Buffer, pH 8.0, prior to fluorescence measurement.

3 Results

3.1 Scale-Down of Intein Purification

3.1.1 Affinity Purification with Chitin Binding Domain Tag

The use of affinity chromatography for high-throughput protein purification is a well-established and commonly used technique that can be adapted for use with self-cleaving tags. To demonstrate the compatibility of intein systems with 96-well plate purifications, a chitin binding domain (CBD) tag was used as a model affinity tag. The CBD tag was chosen because it is commonly used with intein purifications and because it binds strongly to chitin beads, reducing the chance of tag leaching during purification [30]. A flowchart description of the CBD-intein purification is shown in Figure 1A. The purification procedure uses 96-well filter plates that have been charged with chitin beads. The workflow is similar to other 96-well plate affinity purifications, with the major modification being the addition of the on-resin intein cleaving step, which requires a pH shift from pH 8.5 to pH 6.2 and overnight incubation at 25°C.

In order to demonstrate the efficacy of this system, a purification was performed with *E. coli* beta-galactosidase (β Gal) as the target protein. The expression vector pET/CBD-I- β Gal was constructed previously using ligation independent cloning and transformed into *E. coli* strain BLR (Novagen) for expression. The main goals of this experiment were to determine if the purity of the final product was consistent with previous results (~90%) [30], to determine the yield of purified protein, and to determine the well-to-well and plate-to-plate variability of the purified product. To isolate the variability due to the purification process alone, a 200-ml shake flask culture was used to express sufficient protein for all wells of two plates, thus ensuring that each well would receive identical starting material. One half (100 ml) of the culture was harvested, resuspended in 5% original culture volume and diluted into 50 ml of start buffer, which was then distributed evenly to each well of two 96-well plates. The purification was carried out as described in the methods. The SDS-PAGE analysis depicted in Figure 1B indicates that the efficacy of the CBD-intein purification does not appear to be affected by the scale-down to a 96-well plate. In particular, the purity of β Gal in the elution lane (Lane 5) is >90%. Interestingly, there appears to be some residual β Gal in the regeneration lanes (Lane 6).

The well-to-well and plate-to-plate variability is summarized in Figure 2. The yield data in Figure 2A and 2B show that the yield from each individual well is tightly distributed for both plates, and that the average yield over the entire plate is very similar for Plate 1 and Plate 2. These yields are $20 \pm 2 \mu\text{g}$ and $21 \pm 1 \mu\text{g}$, respectively, with the uncertainty taken as one standard deviation. In addition, the distribution of individual well yields is well described by a normal distribution around the average plate yield, as

evidenced by a Shapiro-Wilk test of normality. This suggests a random distribution of individual well yields. Finally, the heat maps in Figure 2C and 2D suggest that there are no obvious edge effects or other patterns to the distribution of well yields.

3.1.2 Regeneration of Chitin Beads

The manufacturer states that it is possible to regenerate the chitin beads 3-4 times by washing with 1% SDS to remove the bound chitin tag. To confirm the reusability of the resin, the other half of the starting culture (corresponding to 100 ml of expression culture) was purified on the same plates after a regeneration cycle with 1% SDS. The average protein yield after regeneration was not significantly different than the original yield (21 ± 1 μ g vs. 20 ± 1 μ g, respectively). Purity analysis using SDS-PAGE was also essentially indistinguishable from the original purification (data not shown). In addition, this system has been used routinely for other applications in our lab with up to 5 regeneration cycles with no noticeable drop in performance.

3.1.3 Non-chromatographic Purification with ELP Tag

To our knowledge, the use of the elastin-like-polypeptide (ELP) tag for resin-free purification has never been demonstrated in a 96-well plate. This purification method is based on the highly selective and reversible precipitation of the ELP tag in response to changes temperature and buffer salt composition [29]. A flowchart representation of the ELP purification scheme is shown in Figure 3A. The major difference between the ELP and CBD schemes is the lack of a resin-charging step for the ELP method, since no affinity resin is necessary for purification. The addition of 0.8 M ammonium sulfate induces the selective precipitation of the ELP tag, which provides a basis for the purification of the tagged target protein. In the 96-well format, the precipitated target is effectively captured and washed using a simple 96-well filter plate. Since the ELP precipitate is unable to pass through the pores of the support membrane, the purification is similar to conventional affinity purifications. Initial data suggests that it is possible to capture the ELP precipitate with a 1.2 μ m Supor membrane filter plate (Pall), indicating that it is feasible to scale down the ELP intein purification (Data not shown).

Thus, to demonstrate this self-cleaving precipitation tag method, an ELP-based purification was performed in a manner similar to that of the CBD purification. The expression vector pET/ELP-I- β Gal was previously constructed using LIC and transformed into *E. coli* strain BLR (Novagen) for expression. Similarly to the CBD purification, protein expression was carried out in a 200-ml shake flask culture to remove the variability of differential expression from the analysis. After some minor optimization, the ELP system performed in agreement with previously published results, as evidenced by SDS-PAGE analysis in Figure 3B. The major modification to the ELP purification scheme was the increased concentration of ammonium sulfate to drive ELP precipitation, from a previously published value of 0.4 M ammonium sulfate to 0.8 M ammonium sulfate in this work. The reason for the increased ammonium sulfate requirement is likely due to the decreased mass of protein available for precipitation, requiring a larger driving force to cause ELP association. The results in Figure 3B indicate that the purification was successful, with a purity in the product lane (Lane 4) of ~90%.

The well-to-well and plate-to-plate variability for the ELP purification was comparable to that with the CBD purification, as summarized in Figure 4. The one major difference with the ELP purification was that the yield of purified β Gal was significantly lower than with the CBD purification ($6.8 \pm 0.6 \mu\text{g}$ vs. $20 \pm 1 \mu\text{g}$). This is primarily due to decreased expression of the ELP-I- β Gal precursor, which may be due in part to its large size ($\sim 180 \text{ kDa}$). As was the case for the CBD purification, the individual well yields were randomly distributed and there were no apparent patterns across the plate (Figure 4C and 4D).

3.2 Increasing Throughput of Expression and Purification

3.2.1 Scale-Down of Expression Culture

After determining that scaling down the intein purification process did not adversely affect the performance of the system, it was necessary to determine if the intein process could be incorporated into a high-throughput workflow using expression in multiwell plates. For this experiment, 24-well deep well plates were chosen over 96-well deep well plates because of their compatibility with standard incubator/shakers [7]. In particular, 96-deep well plates require much higher agitation rates to achieve adequate oxygen levels in the culture, which require specialized titer plate shakers. The use of four 24-well plates allows for the same throughput with larger culture volumes and conventional equipment. For this set of experiments, two different target proteins, β Gal and superfolder GFP (sfGFP), were combined with the two different tags, CBD and ELP, to yield a total of four constructs for study.

For the CBD constructs, expression was carried out in two 24-well plates, 2 ml per well, for a total of 48 expressions. These 48 expression wells were distributed across two 96-well purification plates, as shown in Figure 5A. This pattern was chosen to allow for determination of well-to-well and plate-to-plate variability. In addition, splitting the 2-ml cultures in this manner mimics what the purification would be if 0.5 ml culture volumes in 96-well deep well plates were used. Because expression of the ELP constructs was observed to be lower than the CBD constructs, the amount of starting material in each purification well was doubled by increasing the number of expression wells to 96, in four 24-well plates. The 96-expression wells were distributed across two 96-well purification plates, as shown in Figure 5B. This corresponded to 1 ml culture volume per purification well. This approach was used to increase the resolution of well-to-well variability and thus minimize confounding effects of expression differences between the chitin and ELP tagged proteins.

The primary change in scaling down the expression culture was in the cell lysis step during sample preparation. For the large-scale cultures, sonication was used to aid in final cell lysis. However, sonication does not scale well to 96-well plates. Therefore, 10x BugBuster reagent (Novagen) was added to each well to induce chemical cell lysis. Addition of DNase 1 (Sigma) was also used to degrade residual DNA and reduce sample viscosity. Following sample prep, the purification process was identical to that described in the previous section.

3.2.2 High-Throughput CBD Purification

The main purpose of these experiments was to confirm that the elution profile for the different target proteins was consistent across the purification plates, regardless of expression well, and to quantify the

well-to-well and plate-to-plate variability. Additionally, activity assays were performed to confirm that the purified products still retained activity and to quantify fractional recovery. Figure 6A and 6B show the purified product from 12 randomly selected wells for the CI-βGal and CI-sfGFP purifications, respectively. The purity profile is remarkably consistent and the purity is greater than 90% in all samples.

The quantitative data for these purifications are summarized in Table 1. The well-to-well variability, as described by the percent relative standard deviation (% RSD), is within acceptable ranges for both purifications. This is corroborated by the yield data shown in Figure 6C and 6D, which shows a tight distribution around the plate mean. Statistical analysis using a matched-pairs design (matching purification wells from the same expression well) indicated that there was no significant difference (significance level $\alpha=0.05$) between wells. This suggests that there is no significant effect of position in the purification plate on product yield. The heat maps in Figure 6E and 6F are also in line with this assertion, since there is no apparent pattern to the individual well yields. Additionally, the plate-to-plate variability, defined as the difference between the two plate mean yields divided by the overall average yield, was relatively low, ~5%.

Finally, both target proteins were obtained in an active form and the specific activity of each was in line with previous work from our lab [29]. The percent recovery for each was >60%, which was deemed to be acceptable based on the high level of purity for each protein, along with the single purification step process. Additional analysis indicated that the main source of product loss was premature intein cleavage during expression (Data not shown), which has been observed in a number of intein-based systems.

3.2.3 High-Throughput ELP Purification

The results for the high throughput ELP purification are very similar to the CBD purification and are summarized in Figure 7 and Table 1. In particular, it is interesting to note that purity profiles of the purified target proteins are very similar to the ones for the CBD purification, suggesting that the resin-free ELP purification is just as effective as the affinity CBD purification. The specific activity of the purified target, another measure of purity, is also comparable between the two purification methods. The biggest difference between the two purifications is the overall product yield (per expression volume), which is much lower for the ELP process. This is due to decreased expression of the ELP fusion protein in both cases.

The well-to-well variability for the ELP purifications was ~10%, which was very similar to the CBD purification. Similarly to the CBD purification, there was no significant effect of well position on yield, as evidenced by a matched-pairs analysis. There was also no clear pattern in well yields, as shown in the heat maps in Figure 7E and 7F. Additionally, the plate-to-plate variability for both ELP purifications is remarkably low (<1%), which is less than that observed for the CBD purifications.

4 Discussion

This work demonstrates the efficacy of scaling down self-cleaving intein tag purifications to a 96-well plate format as an alternative to His-tag/protease systems. The self-cleaving intein tag system was used to purify two different target proteins (β -galactosidase and superfolder green fluorescent protein) expressed in *E. coli* with two different purification tags; the chitin-binding domain affinity tag and the elastin-like polypeptide precipitation tag. This work further demonstrates the first use of the ELP tag for resin-free purification in a 96-well plate format.

Scaling down the previously described intein purification method did not have an adverse effect on the quality of the purified product proteins. For each combination of target and tag, the target protein was obtained in a highly pure (>90%) and active form, with a yield of >10 μ g per well. The only modification of the purification protocol from previously published results was in the concentration of ammonium sulfate that was required to induce sufficient precipitation of the ELP tag, from 0.4 M to 0.8M. This was most likely due to the decreased amount of ELP present, which then required a stronger driving force to initiate precipitation.

One of the limitations of the current work is the need for 24-well deep well plates for expression, rather than 96-well plates, limiting potential throughput. However, this limitation was due primarily to availability of appropriate equipment for execution of the study. There should be no reason why this method could not be adapted to 96-well plate expressions.

Interestingly, for both of the β Gal purifications (CBD and ELP), there was cleaved β Gal protein present in the regeneration steps. Ideally, all of the β Gal would be recovered after immediately after intein cleavage. The β Gal may be associating non-specifically with the membrane support of the filter plate, since this is the only factor in common between the two purifications, or it may arise due to tetramer formation between cleaved and uncleaved β Gal monomers. In this case, association of cleaved β Gal monomers with residual uncleaved monomers leads to retention of cleaved monomers on the purification medium, which only happens during purification of multimeric proteins. It may be possible to decrease this association by adding glycerol, salt or another excipient to the buffer, but this was not deemed necessary for this work. This was not observed for the sfGFP target, nor any other monomeric target we have used in our lab. This residual β Gal in the regeneration will adversely affect the fractional recoveries, which could explain why they are lower than the sfGFP recoveries. In addition, the yield of the β Gal derived from the ELP expressions was less than that from the CBD expressions. This may be due to the size of the overall construct (>150 kDa) which could put a metabolic burden on the cells. This effect may also be present for other large target proteins, and we are currently working to quantify these effects (manuscript in preparation).

Additionally, the well-to-well and plate-to-plate variability of product yield was relatively low for each of the purifications (~10% and ~5%, respectively). This level of precision was better than expected, considering that all of the liquid handling steps were performed manually using multi-channel pipettes by experimenters with no previous experience in high-throughput purification. This bodes well on two fronts. First, the barrier for incorporation of the technology described here into labs that do not have much prior expertise in high-throughput protein purification is low. Second, incorporation of this system into labs with well-established high-throughput protocols and state-of-the-art liquid handling systems

would almost certainly decrease this variability even further. Interestingly, the plate-to-plate variability of the ELP purification was extremely low, <1% for both target proteins. This could be due to the fact that the ELP purification is a resin-free purification, so there is no additional variability added by the charging of the filter plates with affinity resin.

5 Conclusions

Self-cleaving intein tags have the potential to be a powerful tool for high-throughput protein production. This simple, one-step system has been previously demonstrated to produce high levels of pure, untagged, native protein for a number of interesting targets. The current work extends our previous work by showing that incorporation of this purification scheme into a high-throughput format allows for high well-to-well reproducibility without sacrificing purity or robustness. This, this self-cleaving intein system could be used at every step of protein production for structural biology; from initial expression screening using LIC intein vectors with modular tag moieties, all the way to high-density fermentation cultures. Finally, use of the ELP tag could allow for significant savings by completely eliminating the need for affinity resins.

Funding

This work was supported by the National Science Foundation Grant #1264322 and the US Army Research Office Contract W911NF-11-1-0118.

Figure Legends:

Figure 1: 96-well chitin binding domain (CBD) purification. (A) Flowchart representation of the CBD purification scheme. Similar to traditional affinity chromatography, except with the addition of the intein-self cleaving step. (B) SDS-PAGE analysis of the scaled-down CBD purification. Numbers above each lane correspond to the numbers in the flowchart. The purity of the eluted product in Lane 5 is >90%. Interestingly, there is still some β Gal in the regeneration lane, Lane 6.

Figure 2: Yield and variability data for scaled down CBD purification. (A) Individual well yields for Plate 1. The middle (blue) line is the plate mean, while the upper (green) and lower (red) lines are the 95% confidence intervals. The well yields range from ~16-24 μ g. (B) Individual well yields for Plate 2. The well yields range from ~16-24 μ g. The distributions and averages are very similar for the two plates (C) Heat map of the well yields for Plate 1. There are no apparent edge effects or patterns in well yields. (D) Heat map of the well yields for Plate 2. Wells B5 and B6 were outliers ($>3\sigma$ away from mean) and were excluded from the analysis.

Figure 3: Demonstration of an elastin-like polypeptide (ELP) purification in a 96-well format. (A) Flowchart representation of the ELP purification scheme. Addition of ammonium sulfate induces reversible precipitation of the ELP tag. The tag is resolubilized in a low salt buffer. (B) SDS-PAGE analysis of the scaled-down ELP purification. Numbers above each lane correspond to the numbers in the flowchart. The purity of the eluted product in Lane 5 is >90% and is comparable to the CBD purification. The overall expression of the ELP-I- β Gal fusion is much lower than the CBD fusion.

Figure 4: Yield and variability data for scaled down ELP purification. (A) Individual well yields for Plate 1. The well yields range from ~5.5-8.5 μ g. The yields are much lower than for the CBD constructs, due to lower expression. (B) Individual well yields for Plate 2. Well yields range from ~5.5-8.5 μ g. Note that the average yield for the two plates is remarkably similar. (C) Heat map of the well yields for Plate 1. There are no apparent edge effects or patterns in well yields. Well E11 was an outlier ($>3\sigma$) and excluded from the analysis (D) Heat map of the well yields for Plate 2.

Figure 5: Schematic showing the distribution of expression wells to purification wells. (A) CBD purification schematic. Two 24-well plates are combined into a single 96-well plate (48 samples) for harvest. Each of the harvested expressions is then distributed to two 96-well filter plates for purification. Each expression well is distributed to 4 purification wells, 2 per plate. (B) ELP purification schematic. Four 24-well plates are combined into a single 96-well plate (96 samples) for harvest. Each of the harvested expressions is then distributed to two 96-well filter plates for purification. Each expression well is distributed to 2 purification wells, 1 per plate.

Figure 6: High-throughput CBD purification results. (A) SDS-PAGE analysis of the CI- β Gal purification. The purified product from 12 wells was randomly selected for analysis. The elution profile of the 116 kDa β Gal protein was very similar for all 12 samples, and the purity was >90% for each. (B) SDS-PAGE analysis of the CI-sfGFP purification. The purified product from 12 wells was randomly selected for analysis. The elution profile of the 27 kDa sfGFP protein was very similar for all 12 samples, and the purity was >90% for each. (C) Individual well yields for one of the CI- β Gal purification plates. Well yields range from ~13-

19 µg. (D) Individual well yields for one of the CI-sfGFP purification plates. Well yields range from ~15-25 µg. (E) Heat map for one of the CI-βGal purification plates. There are no apparent patterns in the yields. Wells H1 and H7 (from the same expression well) were taken as outliers ($>3\sigma$) due to no induction. (F) Heat map for one of the CI-sfGFP purification plates.

Figure 7: Figure 6: High-throughput ELP purification results. (A) SDS-PAGE analysis of the EI-βGal purification. The purified product from 12 wells was randomly selected for analysis. The elution profile of the 116 kDa βGal protein was very similar for all 12 samples, and the purity was $>90\%$ for each. (B) SDS-PAGE analysis of the EI-sfGFP purification. The purified product from 12 wells was randomly selected for analysis. The elution profile of the 27 kDa sfGFP protein was very similar for all 12 samples, and the purity was $>90\%$ for each. The sfGFP samples run as doublets on SDS-PAGE when constituted in the ELP elution buffer, for unknown reasons. (C) Individual well yields for one of the EI-βGal purification plates. Well yields range from ~10-14.5 µg. (D) Individual well yields for one of the EI-sfGFP purification plates. Well yields range from ~7-14 µg. (E) Heat map for one of the EI-βGal purification plates. There are no apparent patterns in the yields. (F) Heat map for one of the EI-sfGFP purification plates. Well D12 was taken as an outlier ($>3\sigma$).

References

- [1] M.R. Bell, M.J. Engleka, A. Malik, J.E. Strickler, To fuse or not to fuse: what is your purpose?, *Protein Sci*, 22 (2013) 1466-1477.
- [2] S.C. Almo, S.J. Garforth, B.S. Hillerich, J.D. Love, R.D. Seidel, S.K. Burley, Protein production from the structural genomics perspective: achievements and future needs, *Curr Opin Struct Biol*, 23 (2013) 335-344.
- [3] W.H. Eschenfeldt, N. Maltseva, L. Stols, M.I. Donnelly, M. Gu, B. Nocek, K. Tan, Y. Kim, A. Joachimiak, Cleavable C-terminal His-tag vectors for structure determination, *J Struct Funct Genomics*, 11 (2010) 31-39.
- [4] R.O. Frederick, L. Bergeman, P.G. Blommel, L.J. Bailey, J.G. McCoy, J. Song, L. Meske, C.A. Bingman, M. Ritters, N.A. Dillon, J. Kunert, J.W. Yoon, A. Lim, M. Cassidy, J. Bunge, D.J. Aceti, J.G. Primm, J.L. Markley, G.N. Phillips, Jr., B.G. Fox, Small-scale, semi-automated purification of eukaryotic proteins for structure determination, *J Struct Funct Genomics*, 8 (2007) 153-166.
- [5] Y. Kim, G. Babnigg, R. Jedrzejczak, W.H. Eschenfeldt, H. Li, N. Maltseva, C. Hatzos-Skintges, M. Gu, M. Makowska-Grzyska, R. Wu, H. An, G. Chhor, A. Joachimiak, High-throughput protein purification and quality assessment for crystallization, *Methods*, 55 (2011) 12-28.
- [6] J. Koehn, I. Hunt, High-Throughput Protein Production (HTPP): a review of enabling technologies to expedite protein production, *Methods Mol Biol*, 498 (2009) 1-18.
- [7] N.J. Saez, R. Vincentelli, High-throughput expression screening and purification of recombinant proteins in *E. coli*, *Methods Mol Biol*, 1091 (2014) 33-53.
- [8] R.C. Cheung, J.H. Wong, T.B. Ng, Immobilized metal ion affinity chromatography: a review on its applications, *Appl Microbiol Biotechnol*, 96 (2012) 1411-1420.
- [9] E.A. Woestenenk, M. Hammarstrom, S. van den Berg, T. Hard, H. Berglund, His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors, *J Struct Funct Genomics*, 5 (2004) 217-229.
- [10] S. Zhu, C. Gong, L. Ren, X. Li, D. Song, G. Zheng, A simple and effective strategy for solving the problem of inclusion bodies in recombinant protein technology: His-tag deletions enhance soluble expression, *Appl Microbiol Biotechnol*, 97 (2013) 837-845.
- [11] M. Hammarstrom, E.A. Woestenenk, N. Hellgren, T. Hard, H. Berglund, Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein, *J Struct Funct Genomics*, 7 (2006) 1-14.
- [12] D.S. Waugh, An overview of enzymatic reagents for the removal of affinity tags, *Protein Expr Purif*, 80 (2011) 283-293.
- [13] X. Zhao, G. Li, S. Liang, Several affinity tags commonly used in chromatographic purification, *J Anal Methods Chem*, 2013 (2013) 581093.
- [14] B.A. Fong, W.Y. Wu, D.W. Wood, The potential role of self-cleaving purification tags in commercial-scale processes, *Trends Biotechnol*, 28 (2010) 272-279.
- [15] J. Arnau, C. Lauritzen, G.E. Petersen, J. Pedersen, Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins, *Protein Expr Purif*, 48 (2006) 1-13.
- [16] D.W. Wood, W. Wu, G. Belfort, V. Derbyshire, M. Belfort, A genetic system yields self-cleaving inteins for bioseparations, *Nat Biotechnol*, 17 (1999) 889-892.
- [17] Y. Li, Self-cleaving fusion tags for recombinant protein production, *Biotechnol Lett*, 33 (2011) 869-881.
- [18] D. Guan, M. Ramirez, Z. Chen, Split intein mediated ultra-rapid purification of tagless protein (SIRP), *Biotechnol Bioeng*, 110 (2013) 2471-2481.
- [19] C. Shi, Q. Meng, D.W. Wood, A dual ELP-tagged split intein system for non-chromatographic recombinant protein purification, *Appl Microbiol Biotechnol*, 97 (2013) 829-835.

- [20] L.E. Smith, J. Yang, L. Goodman, X. Huang, R. Huang, J. Dressman, J. Morris, R.A. Silva, W.S. Davidson, G. Cavigliolo, High yield expression and purification of recombinant human apolipoprotein A-II in *Escherichia coli*, *J Lipid Res*, 53 (2012) 1708-1715.
- [21] L. Tian, S.S. Sun, A cost-effective ELP-intein coupling system for recombinant protein purification from plant production platform, *PLoS One*, 6 (2011) e24183.
- [22] B.A. Fong, A.R. Gillies, I. Ghazi, G. LeRoy, K.C. Lee, L.F. Westblade, D.W. Wood, Purification of *Escherichia coli* RNA polymerase using a self-cleaving elastin-like polypeptide tag, *Protein Sci*, 19 (2010) 1243-1252.
- [23] B.A. Fong, D.W. Wood, Expression and purification of ELP-intein-tagged target proteins in high cell density *E. coli* fermentation, *Microb Cell Fact*, 9 (2010) 77.
- [24] B.A. Fong, W.Y. Wu, D.W. Wood, Optimization of ELP-intein mediated protein purification by salt substitution, *Protein Expr Purif*, 66 (2009) 198-202.
- [25] A.R. Gillies, R.B. Mahmoud, D.W. Wood, PHB-intein-mediated protein purification strategy, *Methods Mol Biol*, 498 (2009) 173-183.
- [26] D.W. Wood, V. Derbyshire, W. Wu, M. Chartrain, M. Belfort, G. Belfort, Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor, *Biotechnol Prog*, 16 (2000) 1055-1063.
- [27] W. Wu, D.W. Wood, G. Belfort, V. Derbyshire, M. Belfort, Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing, *Nucleic Acids Res*, 30 (2002) 4864-4871.
- [28] W.Y. Wu, K.D. Miller, M. Coolbaugh, D.W. Wood, Intein-mediated one-step purification of *Escherichia coli* secreted human antibody fragments, *Protein Expr Purif*, 76 (2011) 221-228.
- [29] M.R. Banki, T.U. Gerngross, D.W. Wood, Novel and economical purification of recombinant proteins: intein-mediated protein purification using in vivo polyhydroxybutyrate (PHB) matrix association, *Protein Sci*, 14 (2005) 1387-1395.
- [30] M.J. Coolbaugh, D.W. Wood, Purification of *E. coli* Proteins Using a Self-Cleaving Chitin-Binding Affinity Tag, *Methods Mol Biol*, 1177 (2014) 47-58.
- [31] T.D. Warren, M.J. Coolbaugh, D.W. Wood, Ligation-independent cloning and self-cleaving intein as a tool for high-throughput protein purification, *Protein Expr Purif*, 91 (2013) 169-174.

Figure 1

A

B

Figure 2

Figure 3

A

B

Figure 4

Figure 5

Figure 6

Figure 7