Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 47 (2006) 16–24
Protein Expression & Purification
www.elsevier.com/locate/yprep

Automated purification of recombinant proteins: Combining high-throughput with high yield

Chiann-Tso Lin [a], Priscilla A. Moore [a], Deanna L. Auberry [a], Elizabeth V. Landorf [b], Terese Peppler [b], Kristin D. Victry [a], Frank R. Collart [b], Vladimir Kery [a,*]

[a] *Pacific Northwest National Laboratory, Richland, WA 99352, USA*
[b] *Argonne National Laboratory, Argonne, IL 60439, USA*

Received 2 August 2005, and in revised form 28 October 2005
Available online 15 December 2005

Abstract

Protein crystallography, mapping protein interactions, and other functional genomic approaches require purifying many different proteins, each of sufficient yield and homogeneity, for subsequent high-throughput applications. To fill this requirement efficiently, there is a need to develop robust, automated, high-throughput protein expression, and purification processes. We developed and compared two alternative workflows for automated purification of recombinant proteins based on expression of bacterial genes in *Escherichia coli* (*E. coli*). The first is a filtration separation protocol in which proteins of interest are expressed in a large volume, 800 ml of *E. coli* cultures, then isolated by filtration purification using Ni–NTA–Agarose (Qiagen). The second is a smaller scale magnetic separation method in which proteins of interest are expressed in a small volume, 25 ml, of *E. coli* cultures then isolated using a 96-well purification system with MagneHis Ni^{2+} Agarose (Promega). Both workflows provided comparable average yields of proteins, about 8 μg of purified protein per optical density unit of bacterial culture measured at 600 nm. We discuss advantages and limitations of these automated workflows, which can provide proteins with more than 90% purity and yields in the range of 100 μg to 45 mg per purification run, as well as strategies for optimizing these protocols.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Automated; Filtration; High-throughput; Magnetic separation; Protein expression; Protein purification; Recombinant protein

Answering biological and medical questions about protein function in the new era of systems biology and functional genomics requires the application of high-throughput approaches for parallel production of many different purified proteins from the same organism. Protein crystallography, generating antibodies or other affinity reagents (aptamers) to proteins of interest, and in vitro mapping of protein–protein interactions are technologies that need not only large numbers of purified proteins but also sufficient quantities of them. Despite the continued miniaturization of these technologies, tens or hundreds of micrograms of purified proteins are needed to achieve meaningful outcomes. Sufficient quantities of desired homogeneous proteins can efficiently be generated by high-throughput processes performed in an automated fashion.

Genome-scale in vivo [1–3] and in vitro [4] expression approaches have been described that allow high-throughput automated cloning and expression of soluble proteins of any organism in a well defined expression system. Recombinant expression in *Escherichia coli* (*E. coli*) is the oldest system, and because of its simplicity, low cost, and robustness, still remains the most popular of the expression systems [5].

Purification of expressed proteins is most efficiently accomplished by using a variety of affinity tags that are engineered at the N- or C- terminus of expressed proteins [6]. Although automated, high-throughput, proteome-scale purification processes have already been described [7,8] for

* Corresponding author. Fax: +1 509 372 6455.
E-mail address: Vladimir.Kery@pnl.gov (V. Kery).

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.11.015

small microgram quantities of purified proteins, the parallel high-throughput purification of larger amounts of proteins are still a considerable technical and technological challenge. While the purification of a single protein is a relatively simple procedure, each labor intensive step in the parallel purification of many proteins can quickly become a bottleneck for the whole process.

In this work we compare a large-scale, lower-throughput, filtration-based purification process with a lower-scale but higher-throughput, magnetic bead-based purification process. We discuss the bottlenecks associated with automated high-throughput protein purification as well as approaches for optimizing workflow logistics for maximum protein yield.

Materials and methods

High-throughput cloning and protein expression profiling

From the annotation of *Shewanella oneidensis* (*S. oneidensis*), hypothetical proteins were chosen and sorted into cytoplasmic, helical membrane, or periplasmic proteins by sequence analysis. The signalP algorithm [9] was used for analysis of the initial 70 N-terminal amino acids of each protein. The program parameters were selected for a gram negative organism and any proteins identified as containing signal sequences by the Neural Net (NN)[1] or Hidden Markov Model modes (HMM) were removed from the heterologous expression targets set. Proteins containing transmembrane regions were identified by application of the transmembrane TMHMM program [10]. Proteins identified as containing transmembrane helical segments were removed from the expression target set. PCR primers were designed using a publicly available tool [11] to design primers (http://tools.bio.anl.gov.8080/publicExpressPrimer-Tool/index.html) for amplification of DNA coding regions.

The open reading frames (ORFs) of hypothetical proteins from *S. oneidensis* were amplified from genomic DNA with KOD DNA polymerase using conditions and reagents provided by the vendor (Novagen, Madison, WI). The amplified fragments were cloned into a pMCSG7 vector using a modified ligation independent cloning protocol [12]. All primers contained ligation independent cloning regions compatible with the pMCSG7 cloning vector [13]. This process generated expression clones producing a fusion protein with an N-terminal hexahistidine (His6) tag and a TEV protease recognition site (ENLYFQ↓S). The fusion proteins were over-produced in an *E. coli* BL21 strain harboring a plasmid encoding three rare *E. coli* tRNAs (Arg [AGG/AGA] and Ile [ATA]). After IPTG induction, targets were analyzed for expression and solubility via denaturing gel electrophoretic (SDS–PAGE) analysis. Targets were scored as positive for expression and solubility if a detectable fusion protein of the correct molecular weight was observed on Coomassie-stained gels. Proteins were then expressed at the production scale in *E. coli* BL21(DE3).

Protein expression in a Bactolift

A Bactolift is an air spurge fermentor (Lofstrand LTD, Gaithersburg, MD) in which agitation is achieved by bubbling air through the culture using a flow-controlled air pump and ceramic air disposers. For 800 ml cultures, 23 g of Q-Biogene Luria broth medium (LB) and 800 ml of deionized water was added to a 1 L polypropylene bottle. This medium was then inoculated with cell cultures in LB and 50 μg/ml carbenicillin to a starting density of 0.1 OD unit measured at 600 nm (0.1 OD_{600}). The culture was grown at 37 °C until the OD_{600} reached 0.6–0.9 (~2 h) at which point 800 μl of 1 M IPTG stock solution was injected to a final inducer concentration of 1 mM. The cultures were maintained under constant airflow (1 L/min) at 30 °C for another 3 h and then harvested. Cell pellets were resuspended in precooled lysis buffer (20 mM Hepes, 2 mM $MgCl_2$, 1 mg/ml lysozyme, 5 μg/ml RNase, and 10 μg/ml DNase, with protease inhibitors (Roche), pH 8.0) at 1:2, w/v.

Protein expression in Falcon tubes

Protein expression for magnetic bead based protein purification was performed in 50 ml Falcon tubes containing 25 ml LB medium and 50 μg/ml of carbenicillin. The tubes were incubated at 37 °C and 220 rpm to 0.6–0.9 OD_{600} then induced with 1 mM IPTG and incubated overnight at 30 °C. Aliquots of cell suspension corresponding to 30 OD_{600} were dispensed to 15 ml Falcon tubes and spun down at 2500 rpm for 15 min. The bacterial pellets were stored at −80 °C until lysis. Protein expression experiments were performed in batches of 24 or 48 tubes per batch.

Sonication mediated cell lysis and filtration protein purification process

Cell suspensions from a Bactolift were homogenized by sonication. Cell debris was removed using high-speed centrifugation (190,000*g* in a Beckman-Coulter L8-70 M centrifuge with 50.2 Ti rotor) at 4 °C for 1 h. The clear lysate was then adjusted with concentrated imidazole and glycerols to final concentrations of 10 mM and 5%, respectively. Cell lysate aliquots were then applied to mini columns prepacked with 1.5 ml nickel nitrilotriacetic acid gel (Ni–NTA Superflow Agarose, Qiagen, Valencia, CA) and equilibrated with binding buffer (20 mM Hepes, 500 mM NaCl, 10 mM imidazole, 2 mM $MgCl_2$, and 5% glycerol, pH 8.0). Liquid handling procedures were performed either using a Qiagen BioRobot 3000 according to the manufacturer's instructions, or the liquids were allowed to flow through the columns by gravity. After loading lysates, the columns were washed with a 10-fold volume of washing buffer (20 mM Hepes, 80 mM imidazole, and 500 mM NaCl, pH 7.0) and

[1] *Abbreviations used:* OD, optical density; ORFs, open reading frames; LB, Luria broth medium; LIMS, laboratory information management system; NN, Neural Net; HMM, Hidden MarkovModel modes.

eluted with 3 ml of elution buffer (20 mM Hepes, 150 mM NaCl, 500 mM imidazole, and 10% glycerol, pH 7.5). The eluted protein solutions were then dialyzed against 20 mM Hepes and 150 mM NaCl, pH 7.5, and the protein concentration determined with a Bradford Protein Assay (Bio-Rad, Hercules, CA). Protein homogeneity was examined by separating proteins on precast SDS 4–12% polyacrylamide gels (Invitrogen, Carlsbad, CA) then sequentially staining the gels with InVision His-tag (Invitrogen) and Coomassie Blue (Gel Code Blue, Pierce, Rockford, IL) stains. Routinely, parallel purification of multiple proteins on 12 or 24 columns was performed. Data about the protein purification and sample processing were maintained using spreadsheets and the in-house developed laboratory information management system (LIMS).

Reagent mediated lysis followed by protein purification using magnetic beads

Lysis with FastBreak cell lysis reagent

Cell suspensions were centrifuged at 2000*g* for 10 min. After removal of supernatant, the resulting pellets were resuspended in 1.1 ml binding buffer (see above) supplemented with protease inhibitors, DNase and RNase (5 μl each of 10 mg/ml stock solution), and lysozyme (0.1 mg/ml). One tenth of the total sample volume of the FastBreak reagent was added to the suspension, homogenized with a pipette, and then the tube was rotated for 10 min at room temperature. The mixture was subjected directly to the automated purification on MagneHis beads.

Lysis with MagneHis cell lysis reagent

Another set of cell pellets were prepared as described above except that 1.0 ml of the binding buffer previously supplemented with protease inhibitors, DNase, RNAse, and lysozyme with the MagneHis lysis reagent was added directly to the pellets in the centrifuge tube. Subsequent procedures were identical to those treated with FastBreak reagent.

Automated purification using MagneHis magnetic Ni^{2+}–agarose A

Automated purification of proteins was performed on the liquid handler, Biomek FX (Beckman-Coulter, Fullerton, CA), using magnetic Ni^{2+}–agarose A particles (Promega, Madison, WI) and a MagnaBot magnetic plate holder (Promega). Cell lysates were loaded at 1 ml per well of a 96-deep well plate to which 100 μl of MagneHis agarose beads in a 50% slurry were transferred. The plate was incubated for 10 min at room temperature using the Biomek FX's shake and pause mode (30 s shakes followed by 1 min pauses). To wash the beads, the plate was moved to a MagnaBot to capture the magnetic beads while aspirating the unbound lysate from the wells. The plate was then returned to the shaker and the beads in each well were mixed with 0.5 ml of washing buffer (20 mM Hepes, pH 7.0, 500 mM NaCl, 5% glycerol, 2 mM $MgCl_2$, and 80 mM imidazole). This washing process was repeated four times. After washing, 200 μl of elution buffer (20 mM Hepes, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 5% glycerol, and 500 mM imidazole) was added to each well and allowed to shake and incubate for 10 min. After capturing the beads using the MagnaBot, the clear eluates were transferred into a new, clean 96-well plate and analyzed as shown in the previous paragraph.

Results

Expression of soluble proteins

A subset of the 236 hypothetical *S. oneidensis* proteins predicted to be cytoplasmic proteins was selected to evaluate the potential for heterologous expression of the target set in *E. coli*. Expression was confirmed for 61% of the hypothetical target set by denaturing gel electrophoresis. Approximately 50% of the clones expressing fusion products generated a visible amount of soluble protein detected on the Coomassie blue stained gels. The soluble proteins were grouped into several relative categories based on qualitative assessment of band intensity, determined via the Coomassie blue staining after denaturing gel electrophoresis. The solubility assignments at this analytical scale show a general correlation to the yield of soluble protein obtained in preparative-scale cultures purified using standard column-based purification methods. This was in agreement with a previous large-scale study [1].

Steric factors affecting fusion protein binding to nickel beads

The optimal amount of beads needed for maximum yield of purified protein depends on the binding efficiency of the hexahistidine-tagged fusion proteins to the nickel beads. Protein-bead binding is affected by multiple factors such as steric availability of the affinity tag, the native size of the tagged protein, and the expression level of soluble tagged protein in the production organism. We evaluated the first two factors by measuring saturation curves of three purified *S. oneidensis* proteins for both Ni–NTA and MagneHis beads. Both beads appear to have similar capacity for the same proteins, each binding less than 50% of total protein of interest available in the media. Lower binding is partially caused by relatively high contribution of the nonspecific binding measured by binding of bovine serum albumin to the beads and subtracted from the binding of specific proteins.

Although the molecular sizes of the proteins were quite similar (SO4449, 20 kDa; SO0159, 16 kDa; and SO1611, 17 kDa) we observed profound differences in binding of the different native proteins to the same beads (Fig. 1). This phenomenon could most likely be explained by differences in the native sizes of the proteins. While the native size of the SO4449 protein is close to its predicted monomer size (20 kDa), the other two proteins, SO1611 and SO0159, form oligomers with native sizes above 100 kDa calculated from the native gel. Differences in binding of those two proteins to the same beads could likely be attributed to steric hindrances

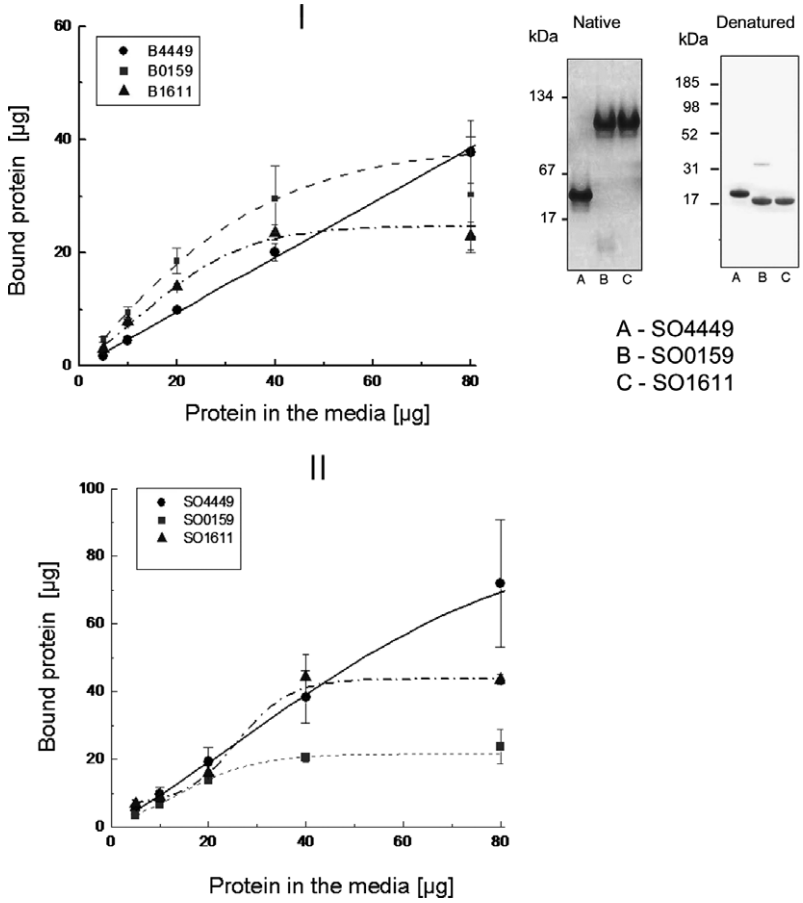

Fig. 1. Saturation curves for Ni–NTA (I) and MagneHis beads (II). Beads, 10 µl of 50% slurry pre-washed with binding buffer, were incubated with purified His-tag fused proteins in various volumes, ranging from 20 to 640 µl at a constant protein concentration of 0.25 µg/µl at room temperature for 30 min. The beads were removed and the amount of protein bound to the beads was determined from the residual protein concentration in the supernatants. Non-specific binding to the beads was evaluated by using bovine serum albumin and subtracted from the binding curves. Native and denatured Coomassie blue stained gels of the proteins are also shown.

around the hexahistidine affinity tag due to protein folding in spatially close areas. Interestingly, we observed reversed saturation curves for binding of hexahistidine-tagged baits SO0159 and SO1611 to MagneHis beads compared to the Ni–NTA beads. While SO0159 binds to Ni–NTA with higher efficiency than SO1611 the binding efficiency of the same proteins to the MagneHis beads is in the reversed order. This is likely due to an effect of the beads on the protein binding efficiency, which indeed translated to the yield of purified proteins. We purified SO1611 and SO0159 using the filtration process and Ni–NTA with yields of purified protein 7 µg/OD_{600} and 16 µg/OD_{600}, respectively, but when using MagneHis beads in the magnetic process we purified them with the yields of 10 µg/OD_{600} and 7 µg/OD_{600}, respectively. Of course protein binding to the beads needs to be completely reversible to correlate the yield of purified protein and its binding efficiency, which may not always be the case.

Effect of protein size on the yield of purified proteins

We observed some individual cases of the effect of native size on binding of purified proteins to the beads but is the dependence of the yield of protein purification on the size of expressed proteins statistically significant? To explore this issue we selected 14 proteins purified to homogeneity of >90% and determined their Stokes radii using gel filtration (Fig. 2). We first observed that the Stokes radii reflecting the native sizes of the proteins broadly varied for different proteins of the same molecular size. For example, the proteins with the small molecular sizes of 14 and 12 kDa were having the largest radii 52 and 54 Å, respectively. On the other hand multiple proteins with similar Stokes radius of ~40 Å were having broadly different molecular sizes (16, 37, and 47 kDa, respectively).

No correlation between the yield of purified proteins and their molecular size was observed. This was true even if larger set of proteins was taken into the considerations (results not shown). Similarly, no significant correlation was observed between the protein purification yield and the Stokes radii (native size of the proteins). Therefore, although we demonstrated some individual native size effects on protein binding to the beads and yield of purified proteins no generic conclusions could be made.

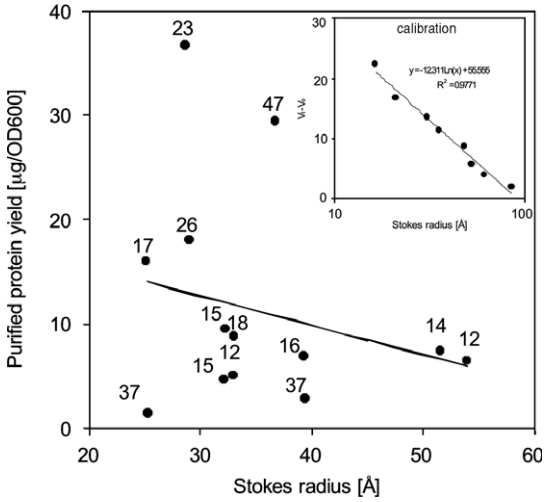

Fig. 2. Yield of purified hexahistidine fusion proteins as a function of a Stokes radius. Inset: calibration curve for the gel filtration column using protein standards ribonuclease A (16.4 Å) chymotrypsinogen A (20.9 Å), ovalbumin (30.5 Å), albumin (35.5 Å), aldolase (48.1 Å), catalase (52.2 Å), ferritin (61.0 Å), and thyroglobulin (85 Å). Blue dextran (2 MDa) was used for determination of the void volume of the column V_o. Proteins, 50 µg each that purified with >90% homogeneity were loaded on the column of Superdex 200 in 100 µl Tris buffered saline, pH 7.5, and were filtered through the column of using an ÄKTA high-performance chromatographic system (Amersham Biosciences) at the flow rate of 0.5 ml/min at 4 °C. Numbers at the points mean molecular sizes of the proteins calculated from their sequences.

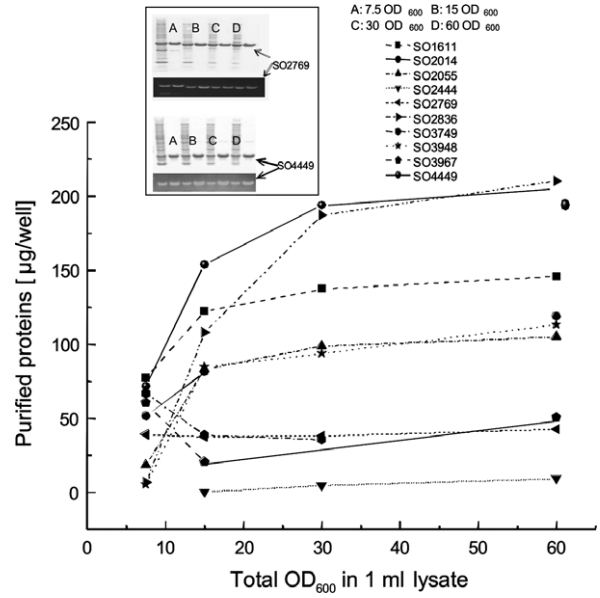

Fig. 3. Yield of purified hexahistidine fusion proteins as a function of cell density. Inset: homogeneity of SO2769 and SO4449 proteins purified from cell lysates of different densities. Cell pellets corresponding to various cell densities measured by OD_{600} were treated with 1 ml lysis reagent and forwarded for automated purification. Aliquots of 5 µg of purified proteins were applied to SDS-NuPAGE gels. The lanes on the left side of the purified lanes are total lysate of the respective ODs. The His-tag proteins were visualized with InVision His-tag (Invitrogen) (lower gel view) and Coomassie blue staining (upper gel view).

Optimizing the amount of cell lysate

The filtration process is not limited in the amount of cell lysate that can be loaded on the affinity columns. However, the plate wells in the magnetic high-throughput process are volume-limited, requiring optimization of the amount of the cell lysate per limited volume to obtain the maximum yield of purified protein.

As we showed in Fig. 1, each hexahistidine-tagged protein binds to the MagneHis beads differently depending on mutual steric relationships of the bead surface and the native protein size. Other factors affecting binding could be protein conformation around the tag and ionic and hydrophobic interactions. Therefore, we optimized the average amount of cell lysate per well with 50 µl of the MagneHis bead suspension using multiple clones of *S. oneidensis* genes in our production *E. coli* BL21(DE3) strain.

Fig. 3 shows that different clones indeed provided different yields of purified proteins reaching the maximum yield at around 200 µg of purified protein per well at a lysate load of 30 OD_{600}. Some of the proteins, SO2769 and SO3967, yielded less than 50 µg of purified product. Interestingly, the homogeneity of some of those low-yield purified proteins was still quite high (SO2769), while others were significantly contaminated by impurities (e.g., SO3967, results not shown). If the bead capacity for hexahistidine-tagged proteins was the only constraint influencing the yields of purified proteins, proteins with lower expression levels would eventually match the yields of the well-expressed proteins if beads were loaded with sufficient amounts of cell lysates.

The limited yield of purified proteins below the saturation limit of the beads must, therefore, be explained by factors other than the binding capacity of the beads for the fusion protein. For example, our observations that some of the expressed proteins are very difficult to elute under native or even under strongly denaturing conditions suggest that very strong and/or irreversible binding of some proteins may influence the final yield. Alternatively, competing hexahistidine-like proteins from the expression cell lysate could bind reversibly to the beads, thus limiting binding of proteins expressed at low levels in the lysate. In this case, increasing the amount of lysate would not lead to an increased amount of purified protein. Indeed, some of the purified protein yields in Fig. 3 were flat regardless the amount of lysate loaded (SO2444, SO2769, and SO3967).

Sonication mediated cell lysis and filtration protein purification process

Parallel expression of milligram amounts of proteins is typically performed by growing the production cells in 2 L Fernbach flasks on a shaker/incubator. We opted for a more space efficient and better cell growth condition-controlling device called a Bactolift in which up to twenty 800 ml cell cultures can be grown in parallel directly in 1 L centrifuge bottles on a small cart. Forced sterile air flow at

1 L/min is distributed into the bottles to sufficiently aerate the cells. Our production *E. coli* cells grew equally well in LB media in both the Fernbach flasks and on the Bactolift reaching the stationary phase at an average density of ~6 OD_{600} in overnight culture at 30 °C. We typically harvested the cells after 3 h of growth in the middle of the logarithmic phase at an average (average of 47 experiments) of 2.8 ± 0.8 OD_{600} yielding about 5.2 ± 0.8 g of wet cells per 800 ml of media. After sonication, an average of 105 ± 24 mg total protein was measured in the lysate. Total yield of purified protein was strongly dependent on the clone and the individual expressed protein level, ranging anywhere between 0 and 45 mg per culture with an average of 13 mg per culture. Total yield of purified proteins per culture could easily be doubled by growing the cells to the stationary phase but this may result in more risk of protein denaturation or cleavage in vivo. In our study set of 47 *S. oneidensis* proteins, 31 were purified to greater than 90% homogeneity.

Reagent mediated lysis followed by protein purification using magnetic beads

We aimed to increase the throughput of protein purification while maintaining the maximum yield of purified proteins per clone. As mechanical sonication is not easy to accomplish in parallel, we needed to replace it with a more high-throughput, yet efficient, lysis protocol. We tested two lysis protocols using Promega lysis reagents. While the concentrated FastBreak reagent was added directly to the cell culture suspension, the lysis reagent in the MagneHis lysis protocol was diluted and applied to the cell pellet obtained by centrifugation of the cell suspension. We observed that spinning cells to produce the cell pellets followed by re-suspension in the diluted MagneHis lysis reagent in the lysis buffer provided consistently larger yield of purified proteins than adding the concentrated FastBreak lysis reagent to the cell suspension (results not shown).

Homogeneity of the proteins purified by 96-well automated protein purification was evaluated by SDS–PAGE followed by consecutive staining using InVision His-tag stain then Coomassie blue stain (Fig. 4). This allowed simultaneous evaluation of the purity and the proteolytic cleavage of the tagged proteins on one gel. About 30% of proteins were purified to >90% homogeneity and about 40% of proteins were purified to >80% homogeneity (Fig. 4 and Table 1). We did not purify any protein from 14 clones. Some of the proteins did not have predicted size. For example, SO0342 (Fig. 4, #9) should be 39 kDa but two bands at

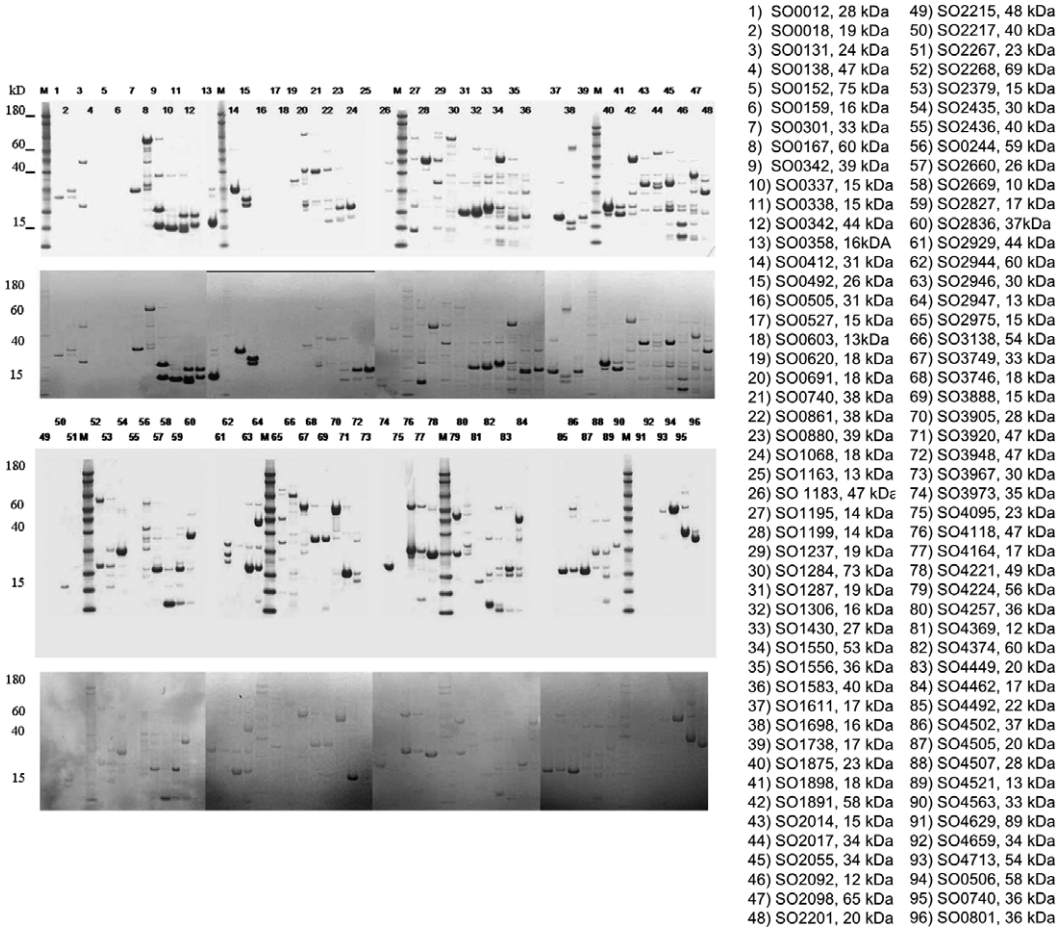

Fig. 4. Homogeneity of proteins purified on a Biomek FX. Purified proteins, 3 μg/lane, were separated using SDS denaturing electrophoresis on NuPAGE gels. The proteins were stained either with Coomassie brilliant blue (upper panel) or with the hexahistidine specific fluorescent stain, InVision His-tag (Invitrogen).

15 and 20 kDa were observed instead. Likewise, SO1853 (Fig. 4, #36) should be 40 kDa but two bands at 15 and 30 kDa were found instead. A likely explanation is that the change occurred due to a protease-mediated cleavage of the fusion proteins in vivo during protein expression or in vitro during the protein purification process. Other inconsistencies could be explained by cross-contamination of some samples with clone expressed/purified in a neighbor well (e.g., #76 and #77 for clones SO4118 and SO4164, respectively, in Fig. 4). Human error is much more likely to occur in the high-throughput processes than in "traditional" experiments of the few tubes at the time. Consequently, a rigorous quality control based on sequence determination (e.g., mass spectrometry) of the purified proteins needs to be introduced into the high-throughput protein purification process to confirm identity of purified proteins.

Statistical evaluation of the workflows

Statistical evaluation of our high-throughput protein purifications is summarized in Table 1. It appeared that growing the expression cells in a smaller, 25 ml, volume yielded more homogeneous cultures with a lower spread of cell densities between cultures than those grown in a larger, 800 ml volume on the Bactolift. The total yield of purified protein per clone for the larger-scale and the lower-scale methods was an average of 12.9 and 0.25 mg, respectively. Calculating the purified protein yields per unit of optical density shows that both types of purification were quite comparable, 8.7 and 8.8 μg/OD_{600} for the larger-scale and lower-scale methods, respectively, indicating that the yields of purified proteins per unit of cell density were not dependent on the scale of expression or the purification mode (filtration through Ni–NTA or using MagneHis).

We also performed a test of reproducibility for our purification workflows. A high expression clone (SO0138) and a low expression clone (SO2379) were each grown in ten culture tubes in a 25 ml volume and purified using our standard automated procedure. We obtained an average yield of 5.68 ± 0.22 and 0.39 ± 0.06 μg/OD_{600} for the SO0138 and SO2379, respectively. In the larger scale we tested reproducibility of protein purification on clones SO0256 and SO4449 grown in triplicate 800 ml cultures. We obtained average yields of 7.5 ± 0.6 μg/OD_{600} for each protein. The homogeneity of the purified proteins (>90%) was equivalent for all purified samples (results not shown). Our results indicate that our automated protein purification workflows produced purified proteins with high reproducibility in both the purified protein yield and the homogeneity.

Table 1
Comparison of the filtration-based protein expression and purification workflow with the workflow using magnetic separation

	Filtration	Magnetic separation
Culture volume (ml)	800	25
Lysate volume (ml)	10	1 ml for 30 OD_{600}
Purification method	1.5 ml Ni–NTA	50 μl MagneHis particles
Total number of samples	47	96
Total average OD_{600} at harvest	1608 ± 850 ($n = 39$)	96.4 ± 25.5 ($n = 96$)
Total average protein (mg)	12.9 ± 12.5 ($n = 46$)	0.25 ± 0.06 ($n = 96$)
Yield μg/OD_{600} at protein Purity >90% μg/OD_{600}	Average = 8.7 ± 7.8 ($n = 31$)	Average = 8.8 ± 1.9 ($n = 30$)
Yield μg/OD_{600} at protein Purity >80%	Average = 8.5 ± 7.7 μg/OD_{600} ($n = 34$)	Mean = 8.8 ± 1.8 μg/OD_{600} ($n = 39$)

Discussion

High-throughput methods that produce a number of different proteins in parallel, each with sufficient yield and homogeneity, are essential for functional proteomic studies. Several in vitro and in vivo protein expression systems have been developed [3], but the in vivo *E. coli* expression system is by far the most common due to its low cost, scalability, and speed. The main drawback of this expression system is often the low solubility of expressed proteins mainly from species phylogenetically distant from *E. coli*. Although the hexahistidine affinity tag can add to the lower solubility of in vivo expressed proteins in *E. coli* [3,14], the increased yield of purified proteins [8], simplicity of the purification procedure, and its versatility are just a few of the many features that make this tag so popular and useful in high-throughput purification methods.

Recently four different expression vectors with N- or C-terminal hexahistidine tags were tested for expression of soluble proteins in *E. coli* [14]. In a study set of 20 human proteins the authors found that the hexahistidine tags have noticeable effect on protein solubility regardless the tag placement. This negative solubility effect could be alleviated by introducing a second "solubilizing" tag. This strategy of testing multiple affinity tags on the same set of requested proteins was recommended to improve protein solubility and consequently to increase the yields of purified proteins [15]. Large and highly soluble tags (e.g., maltose binding protein and glutathione *S*-transferase) typically perform the best [3,15]. However, the size limitation for the expressed protein with a larger tag will have to be negotiated.

An automated, high-throughput process for the production of bacterial expression clones with a linear throughput of 400 targets per production run has recently been developed at the Argonne National Laboratory [12]. The clones are produced in a plate-based format using automated methods and a molecular biology robotic system. Our high-throughput strategy implements automated methods at the front end of the process that enable the generation of large numbers of expression clones in microwell plates. A solubility screen then provides a plate map that identifies the location of wells containing clones producing soluble proteins. This process provides an 80% success rate for the identification of clones producing soluble protein and results in a significant decrease in the level of effort required for the labor intensive components of validation and preparation of freezer stocks. Although many of these proteins can be

produced in a soluble form using novel expression strategies, these novel approaches and the requirements for production need to be validated and refined by experimentalists. The need for purified proteins can only be met by the development of high-throughput protein production and purification strategies that parallel current capabilities for high-throughput cloning and expression.

We used the pMCS7 vector and the N-terminal hexahistidine affinity tag to express and purify *S. oneidensis* proteins in *E. coli*. Although the clones were pre-selected for production cloning and expression by an estimate of solubility, only about 30% of proteins were purified to more than 90% homogeneity. Others were either not pure enough or were not expressed in sufficient quantity to get meaningful yields. This is not surprising and agrees with other studies using the hexahistidine tag in which only 29% of a set of 37 human proteins [16], and 13% of a set of 32 human proteins [7] were purified to homogeneity.

There are only few reports evaluating how the protein structure can affect the protein purification yield; for the most part, these reports address the purification of individual proteins. Our targeted experiments showed that the native size of the fusion proteins, their structure and folding around the hexahistidine affinity tag, and the nature of the beads can significantly affect protein-bead binding and the yield of purified proteins. However, no statistically significant dependence of the yield on both molecular and native protein sizes could be found in a larger set of proteins. This suggests that there are multiple steric and molecular effects reflected into the protein purification yield and more high-throughput studies are needed to understand how their combination influences protein purification.

Purifications of only a few micrograms per well using automated high-throughput protein purification processes were recently published [7,8]. We optimized the amount of the cell lysate per well in our automated microplate process to achieve maximal average yield up to hundreds of micrograms (250 μg) of purified protein per well. Both our workflows provided purified proteins with high reproducibility in both yield and purity.

The protein purification process itself significantly contributes to the yield and quality of purified proteins. In this work we described two parallel protein purification workflows, the larger-scale, lower-throughput filtration-based process and the lower-scale, higher-throughput magnetic separation process. These processes yielded from 0.1 to 45 mg of purified proteins per clone. The automated filtration process is not much time- and labor-saving compared to the same, manually performed process. On the other hand, the lower-scale 96-well automated purification process based on magnetic separation is quite efficient with minimal error rate. Labor-saving lysis reagents that avoid using mechanical lysis of the production cells and centrifugation of the cell debris greatly contributed to the efficiency of the magnetic purification workflow. The recent introduction of the 96-well pre-cast SDS–polyacrylamide gels and the simple fluorescent staining of hexahistidine-tagged proteins by Invitrogen further increased efficiency of the high-throughput protein purification process and offered higher degree of quality control.

Culturing and handling the production cells is the last bottleneck of the process. Minimizing the volume of the cell cultures yet preserving or increasing the total cell mass significantly increases the efficiency of protein expression. This can be accomplished by using rich media and by growing cells to the cell densities optimal for maximum yields of the soluble expressed proteins. Recently, we achieved cell densities up to 30 OD_{600}/ml using rich media (2× YT) and sufficiently aerating the cell cultures. This provided the basis for expressing enough cells in 1 ml of cell cultures for purification of up to an average of 0.25 mg purified protein per clone for 96 proteins. Both protein expression and purification could then be done in one microplate.

In this work, we described two different platforms for automated high-throughput protein purification based on two different heterogeneous phase separation principles, filtration, and magnetic separation. We discussed factors affecting the yield and homogeneity of purified proteins. Both platforms are easily adaptable to expression systems other than *E. coli* or in vitro expression systems, and they can be further scaled up to meet the protein yield demands of modern high-throughput functional genomic and proteomic applications.

Acknowledgments

This work was supported by US Department of Energy, Genomics: Genomes to Life project Grant #KP1102010. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under Contract DEAC05-76RLO 1830. We thank to Loel E. Kathmann for editing and correcting the manuscript.

References

[1] S. Moy, L. Dieckman, M. Schiffer, N. Maltsev, G.X. Yu, F.R. Collart, Genome-scale expression of proteins from *Bacillus subtilis*, J. Struct. Funct. Genomics 5 (2004) 103–109.

[2] C.H. Luan, S. Qiu, J.B. Finley, M. Carson, R.J. Gray, W. Huang, D. Johnson, J. Tsao, J. Reboul, P. Vaglio, D.E. Hill, M. Vidal, L.J. Delucas, M. Luo, High-throughput expression of *C. elegans* proteins, Genome Res. 14 (2004) 2102–2110.

[3] P. Braun, J. LaBaer, High throughput protein production for functional proteomics, Trends Biotechnol. 21 (2003) 383–388.

[4] T. Sawasaki, T. Ogasawara, R. Morishita, Y. Endo, A cell-free protein synthesis system for high-throughput proteomics, Proc. Natl. Acad. Sci. USA 99 (2002) 14652–14657.

[5] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411–421.

[6] K. Terpe, Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems, Appl. Microbiol. Biotechnol. 60 (2003) 523–533.

[7] P. Braun, Y. Hu, B. Shen, A. Halleck, M. Koundinya, E. Harlow, J. LaBaer, Proteome-scale purification of human proteins from bacteria, Proc. Natl. Acad. Sci. USA 99 (2002) 2654–2659.

[8] C. Scheich, V. Sievert, K. Bussow, An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography, BMC Biotechnol. 3 (2003) 12.
[9] H. Nielsen, S. Brunak, G. von Heijne, Machine learning approaches for the prediction of signal peptides and other protein sorting signals, Protein Eng. 12 (1999) 3–9.
[10] E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175–182.
[11] J.R. Yoon, P.D. Laible, M. Gu, H.N. Scott, F.R. Collart, Express primer tool for high-throughput gene cloning and expression, Biotechniques 33 (2002) 1328–1333.
[12] L. Dieckman, M. Gu, L. Stols, M.I. Donnelly, F.R. Collart, High throughput methods for gene cloning and expression, Protein Expr. Purif. 25 (2002) 1–7.
[13] L. Stols, M. Gu, L. Dieckman, R. Raffen, F.R. Collart, M.I. Donnelly, A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site, Protein Expr. Purif. 25 (2002) 8–15.
[14] 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.
[15] Y.P. Shih, W.M. Kung, J.C. Chen, C.H. Yeh, A.H. Wang, T.F. Wang, High-throughput screening of soluble recombinant proteins, Protein Sci. 11 (2002) 1714–1719.
[16] M. Hammarstrom, N. Hellgren, S. van Den Berg, H. Berglund, T. Hard, Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*, Protein Sci. 11 (2002) 313–321.