Journal of Structural Biology 172 (2010) 14–20

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Widening the bottleneck: Increasing success in protein expression and purification ☆

Ralph Hopkins, Dominic Esposito, William Gillette *

Protein Expression Laboratory, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA

ARTICLE INFO

Article history:
Available online 29 July 2010

Keywords:
Protein purification
Baculovirus
High throughput

ABSTRACT

The number of variables at play in the expression and purification of a single protein dwarf those involved in sequencing a genome. Although certain trends are apparent, there is no one-size-fits-all approach to the process of purifying proteins. Thus, whereas numerous genome sequencing projects are providing an overwhelming number of interesting open reading frames for structural biologists to study, fully realizing the potential of this resource is still only a distant hope. We will discuss several current approaches to high throughput expression and purification as well as strategies that have served us well to quickly identify lead protein expression constructs in the context of a core service protein expression and purification laboratory. The use of the baculovirus expression vector system and implementation of a purification screening method will be emphasized.

1. Introduction

Expressing and purifying human proteins, especially in *Escherichia coli*, the traditional host organism for high throughput (HTP) protein expression and purification, continues to be problematic for researchers. The low success rate (2–20%) when expressing eukaryotic proteins in *E. coli* (Service, 2002) stems from several well known problems including low yields due to toxicity, recombinant protein insolubility, and protein aggregation. It is this low success rate that drives the use of alternative expression systems, a wide variety of expression constructs, and numerous HTP approaches.

The baculovirus expression vector system (BEVS) has become an indispensable expression system for the production of proteins (Hunt, 2005, Aricescu et al., 2006). There are several advantages of insect cells over *E. coli* including improved solubility, incorporation of some post-translational modification and higher yields for secreted proteins (Jarvis, 2009). The use of viral promoters leads to soluble expression levels often equal to or greater than those reached in *E. coli*.

For the past 7 years we in the Protein Expression Laboratory (PEL) have striven to deliver a wide variety of eukaryotic proteins to researchers at the NIH. While there are no "magic bullets", we have evolved a pipeline using baculovirus expression in insect cells and micro scale protein purification testing that allows us to accurately predict success or failure quickly and cheaply. Combined with multiple expression formats that can be compared in parallel, our recent success at scale up has risen to 90%. This review will focus on our experiences and is by no means a comprehensive review of all available techniques.

2. The PEL approach

It should be noted at the outset that this review is intended to convey useful information to laboratories that might not have the resources for 'high-end' automation. Our lab falls into this category as do most of the NIH scientists to whom we provide services. Thus, many of the results we discuss were developed with minimal or no automation. The important exception to this will be discussed in detail, however, even in this case the cost was a fraction (~10%) of many fully automated platforms.

In part due to this financial limitation, but more importantly because we observed very quickly after our inception that success in protein production can be improved by embracing diversity in terms of constructs, expression systems, expression conditions, and protein homologs, the PEL has approached the protein production problem in an incremental fashion on many fronts. Accordingly, our processes have changed over time and it is the goal of the review to discuss what has, and equally important, what has not worked in our hands.

Abbreviations: HTP, high throughput; BEVS, baculovirus expression vector system; IMAC, immobilized metal ion affinity chromatography; LSD, large-scale direct; ED, early detection; PEL, Protein Expression Laboratory; GFP, green fluorescent protein; IEX, ion-exchange chromatography; HIC, hydrophobic interaction chromatography.

☆ This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

* Corresponding author. Address: SAIC-Frederick, Inc., Building 327, Room 6, Frederick, MD 21702, USA.
E-mail address: gillettew@mail.nih.gov (W. Gillette).

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.07.005

The most broad reaching and influential of these improvements has been the shift away from screening by expression profiling to screening by purification, a platform we refer to as purify first, or PF. It is this partially automated approach that has had a dramatic impact on our success by reducing the effort we spend on 'dead-end' constructs and in a synergistic fashion, has allowed us to 'widen the bottle neck' of our process. Specifically, we learn very quickly and for little cost what constructs/expression systems/expression conditions lead to purified protein. The higher throughput and efficiency of PF allows us to test a wider diversity of constructs/expression systems/expression conditions which translates into higher chance of success for any given protein.

3. BEVS cloning

Cloning genes for expression in the BEVS is a two step process: (1) creating a construct with the proper context (e.g. promoters, affinity and/or solubility tags) using standard cloning techniques for propagation in *E. coli* and (2) transferring the expression construct to insect cells for expression. Recombinational cloning has largely replaced restriction enzyme/ligase based cloning for the first step of this process (Walhout et al., 2000). There are several strategies available for transfer of the expression construct to the insect cells either via a bacmid DNA (e.g. Bac-to-Bac, Invitrogen) or directly into the viral genome (e.g. BacMagic, EMD Biosciences) (Hartley, 2003; Koehn and Hunt, 2009). Although the PEL creates many constructs for BEVS expression (~150 DNAs/year), the numbers are too low to justify the expense of automation. Rather, we have modified several aspects of the BEVS system to speed the process and increase the likelihood of finding optimal expression conditions.

4. BEVS expression

We use the Bac-to-Bac system to transfer expression constructs from *E. coli* to insect cells via an intermediate bacmid DNA. Once transfected, the bacmid DNA leads to the production of active virus in a lytic cycle. Using a modification to the transfection protocol, we bypass the traditional amplification step that may require 1–2 virus passages to achieve a high titer stock. In what we term 'large-scale direct' (LSD) transformation (for an outline of the protocol, see Fig. 1), a 100 ml high titer BV stock is created in 4 days (manuscript in preparation). Determining the titer of the viral stock allows for consistency in scale up expression experiments and we have greatly simplified this step by using a modified cell line which contains a fluorescent protein under the control of the viral polyhedrin promoter (Hopkins and Esposito, 2009). Virus titers can be determined in 3 days compared with the 6–8 days required for the cytopathic effect method. Also the detection of GFP-positive wells in an end-point dilution assay is much easier than detection of cell death, especially for the novice (Fig. 2).

Several years ago we observed distinct qualitative and quantitative differences between insect cell lines in terms of protein expression levels and the extent of proteolysis. Differences were also observed under certain culture conditions. Our lab routinely uses Sf9 cells and High Five™ (Invitrogen) cells for expression testing. In general, we find proteins expressed in High Five™ cells suffer more proteolysis, but this can sometimes be mitigated by lowering the incubation temperature from 27 to 21 °C. Protein expression levels are also frequently 2- to 5-fold higher in High Five™ cells than in Sf9 cells. However, this observation is not universal among researchers and several labs report good results with Sf9 and Sf21 cell lines. From these observations we developed a three culture (Sf9 cells incubated at 27 °C, High Five™ cells incubated at 21 °C, and High Five™ cells incubated at 27 °C), six sample time course (two harvests/per culture, 48 and 72 h post-infection) optimization method to determine the best conditions for protein expression (see Fig. 3 for a typical example). After using this platform to analyze the expression pattern of hundreds of proteins it became apparent that at least 90% of the time a single set of conditions was optimal: High Five™ cells, incubated at 21 °C and harvested 72 h post-infection. It should be noted that although incubation of High Five™ cells at the standard temperature of 27 °C does frequently give high expression levels, the window of time before the target protein becomes degraded is shorter than at 27 °C and this poses problems especially in scale up. This point is easy to observe in Fig. 3 as the target protein is essentially absent at 27 °C at 72 h. Given that the titers of the high titer stock were commonly within a fairly limited range ($2–4 \times 10^8$ PFU/ml), we reasoned that we could have high confidence in a single condition expression test using non-titered virus: a process we refer to as 'early detection' (ED). For an outline of the protocol, see Fig. 1. Thus, combining LSD and ED methods we reduce the time it takes to go from bacmid to samples for analysis from 3 weeks to 1 week.

Insect cells can also be used to express a secreted protein although the yields are not as high as obtained from *Pichia pastoris* or *Kluyveromyces lactis*. Nevertheless, we have observed yields up to 30 mg/l from insect cells and routinely use this as an option when working with secreted proteins. The major drawback to the approach is the incompatibility of insect cell culture media with immobilized metal ion affinity chromatography (IMAC). The use of a multi-histidine tag is extremely common as it affords a simple and relatively cheap affinity chromatography step. Purification of his-tagged secreted proteins from insect cell culture requires a buffer exchange step prior to chromatography (Hunt, 2005). Since signal peptides and cell lines have been reported to affect expression levels (Futatsumori-Sugai and Tsumoto, 2010), we routinely evaluate GP67, honey bee melittin, and native signal peptides for secretion constructs. As of yet, we have detected no predictable pattern and thus we screen these signal peptides as expression levels can be enhanced up to 3- to 4-fold.

5. Mammalian expression

Transient transfection of DNA constructs for cytoplasmic expression of proteins in mammalian cells has long been overlooked as a source of protein production. In a recent study (manuscript in preparation) we examined the effects of promoters, enhancers, UTR elements, other vector backbone components and fusion tags on the expression of different genes in mammalian hosts. The results validate our approach as certain combinations of elements led to a dramatic improvement in protein expression levels. Thus screening these conditions may lead to significant increases in protein production. Also, we observed that fusion to the green fluorescent protein (eGFP) can serve as a readout for protein expression which is highly scaleable for many proteins to large-scale production. Finally, we hope that given enough data, we may start to see trends in the "best" promoter or tag, which might limit the cloning options to a smaller initial set of test clones for a given protein.

6. Widening the bottleneck

6.1. Background

It is well established that protein production is the limiting factor in many biological experiments and especially so for large-scale structure based initiatives. A common strategy has been to choose a limited number of expression clones to simplify screening and purification. The success rate of this approach is typically less than 10% when applied to eukaryotic proteins expressed in *E. coli*