Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 36 (2004) 40–47
Protein Expression & Purification
www.elsevier.com/locate/yprep

# High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells

Stephen P. Chambers,* Douglas A. Austen, John R. Fulghum, and Walter M. Kim

*Vertex Pharmaceuticals Inc., 130 Waverly Street, Cambridge, MA 02139-4242, USA*

Received 25 November 2003, and in revised form 3 March 2004
Available online 10 May 2004

**Abstract**

We have constructed a dual expression vector for the production of recombinant proteins in both *Escherichia coli* and insect cells. In this vector, the baculoviral polyhedrin promoter was positioned upstream of the bacteriophage T7 promoter and the *lac* operator. This vector, designated pBEV, was specifically designed to exploit the advantages that both hosts would provide. This vector also facilitates one-stop cloning, thereby simplifying the expression process for automation, and the development of a high-throughput method for protein expression. Utilizing the multi-system vector pBEV, a high-throughput process was developed with expression in deep-well blocks and purification in micro-titer plates enabling the identification of expression and solubility in both *E. coli* and insect cells. In this study, using pBEV, we have successfully expressed and purified multiple human kinases produced in *E. coli* and insect cells. Our results validate expression screening as a strategy to rapidly triage proteins identifying the optimum expression system and conditions for production.


*Keywords:* Recombinant protein; Heterologous expression; High-throughput; Automated purification; Screening; Solubility

The technology responsible for producing recombinant protein is now highly developed with a plethora of vectors available for gene expression in a wide variety of prokaryotic and eukaryotic hosts. Since the choice of host for expression is critical in determining the quality and quantity of protein produced, multi-system vectors [1,2], including the recombination cloning vectors [3,4], have been specifically developed to allow access to multiple host and expression systems. In our construction of a multi-system vector, we have sought to exploit *Escherichia coli* and insect cells, two of the most common hosts used for the production of recombinant proteins. While successful *E. coli* expression results in overexpression, lacking the post-translational processing machinery of eukaryotic cells, *E. coli* expressed proteins are often insoluble and inactive [5]. Insect cell expression complements many of the deficiencies produced in *E. coli* expression, providing an effective alternative eukaryotic expression system with a formidable track record of producing authentic protein in a biologically active form [6]. Since both systems have inherent limitations, choosing a suitable host for successful expression, which is often protein specific, can be an empirical process. Therefore, the ability to rapidly express and analyze recombinant production from *E. coli* and insect cells in parallel is the key to fully exploiting multi-system vectors, allowing the screening and identification of the optimal expression system and conditions to maximize soluble protein production. While there are reports of screening soluble proteins expressed in *E. coli* [7,8] and high-throughput expression using *E. coli* [9], yeast [10], mammalian [11], and insect cells [12], our high-throughput process is distinguished by its combined exploitation of both *E. coli* and insect cell expression systems in parallel.

In this study, we describe pBEV, a vector combining the elements of the *E. coli* bacteriophage T7 [13] and the baculoviral polyhedrin (*polh*) [14] promoters in tandem. pBEV functions as both an *E. coli* expression vector and baculovirus transfer vector capable of generating

* Corresponding author. Fax: 1-617-444-6437.
*E-mail addresses:* chambers@vrtx.com, stephen_chambers@vrtx.com (S.P. Chambers).

1046-5928/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2004.03.003

recombinant baculovirus (vBEV) for the expression of foreign genes in insect cells. This vector allows one-stop cloning, greatly facilitating high-throughput cloning of genes and allowing a direct comparison of protein expression in *E. coli* and insect cells. We have demonstrated the utility of pBEV in a high-throughput expression process allowing the screening of 62 full-length kinases in both *E. coli* and insect cells for expression and solubility. This high-throughput expression and automated purification process has allowed the triage of proteins separating those that are difficult to express and purify, either expressed in low amounts or insoluble, from those that are well expressed and soluble.

## Materials and methods

### *Construction of expression vector*

The expression vector pBEV was constructed from pBacPAK8 (BD Biosciences-Clontech, Palo Alto, CA, USA) by inserting a 155 bp *Bgl*II–*Bam*HI fragment, isolated from pET15b (Novagen, Madison, WI, USA), into the *Bam*HI site of pBacPAK8. This fragment contains the T7*lac* promoter (the T7 promoter and *lac* operator sequences), the thrombin-cleavable His-tag sequence, and a portion of the polylinker. The *polh* promoter from pBacPAK8, now in series with the inserted T7*lac* promoter, was optimized for expression by removal of a portion of the sequence upstream of the T7*lac* promoter, reducing the distance between the *polh* start codon at position +1 and the ATG start codon to 86 bases. The resulting sequence of tandem promoters is a common feature present in all the pBEV constructs used in this study (Fig. 1).

In pBEV, substituting the low copy number origin of replication of pET15b, derived from pBR322, with the high copy number origin of pBacPAK8, derived from pUC18, increases the gene dosage and resultant recombinant expression in *E. coli* [15]. pBEV, unlike pET15b, does not contain the *lac* repressor gene in its backbone. This feature coupled with increased copy number, results in less stringent regulation of its T7*lac* promoter. Basal expression, directed by the T7*lac* promoter, was minimized using *E. coli* containing pLysS expressing T7 lysozyme, a natural inhibitor of T7 RNA polymerase [16]. In this study, all *E. coli* expression was performed in BL21 [$F^-$, *omp*T, *hsd*$S_B$ ($r_B^-$, $m_B^-$) *gal*, and *dcm*] (DE3) pLysS, which provides a protease deficient background for the expression of proteolytic sensitive proteins [17].

### *Cell growth in deep-well blocks*

Recombinant *E. coli* BL21(DE3) pLysS, in a 24-well block, aseptically sealed with AirPore tape sheets (Qiagen, Valencia, CA, USA) was grown using a HiGro incubator–shaker (Genomic Solutions, Ann Arbor, MI, USA) at 37 °C for 4.5 h in brain heart infusion (BHI) media (Becton–Dickinson, Sparks, MD, USA) supplemented with 100 μg/ml carbenicillin and 35 μg/ml chloramphenicol. The 5 ml cultures growing in a 24-well block were sampled every hour and absorbance ($A_{600\,nm}$) was recorded. Over the time period sampled, cell densities of the 24-well block were found to be comparable to an 800 ml culture grown in a 2 L fernbach flask (Fig. 2A).

*Trichoplusia ni* (High-5) cells infected at a multiplicity of infection (MOI) of 5 pfu/cell with recombinant baculovirus generated from pBEV were grown for 72 h at 28 °C in serum-free EX-CELL 405 media with L-glutamine (JRH Biosciences, Lenxa, KA, USA). The 3 ml culture grown in a 24-well block was sampled every 12 h and the number of viable cells was determined (Fig. 2B) using a Cedex analysis system (Innovatis GmbH, Bielefeld, Germany). Viabilities obtained from cells grown in the HiGro incubator–shaker appeared comparable to those obtained from cultures grown in a fernbach flask. It is noteworthy that the rate of infection also appears better in the 24-well block than in the flask, possibly due to the higher vortex action in the 24-well block produced by the HiGro incubator–shaker.

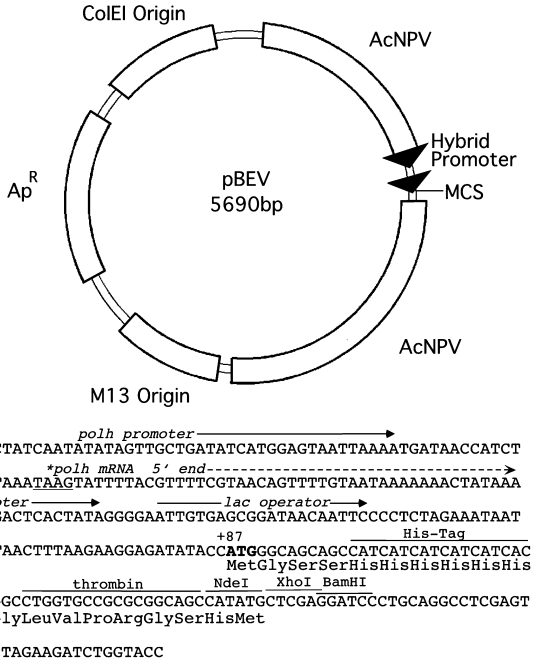


Fig. 1. Schematic map of the pBEV expression vector. The vector contains *polh*:T7*lac* promoter regions, multiple cloning site (MCS), flanking *Autographa californica* nuclear polyhedrosis viral (AcNPV) region for recombination, the ColE1 origin of replication derived from the high copy number cloning vector pUC, the M13 origin for preparation of single strand DNA for mutagenesis, and the β-lactamase gene for selection. Accompanying sequence below describes the hybrid *polh*:T7*lac* promoter, with the continuous lines identifying the T7*lac* promoter operator regions and the *polh* promoter, with the tetranucleotide TAAG underlined. Dotted line indicates the 5′-mRNA untranslated *polh* transcript. The distance from the site of the original *polh* ATG and the new ATG start codon (in bold) is indicated.

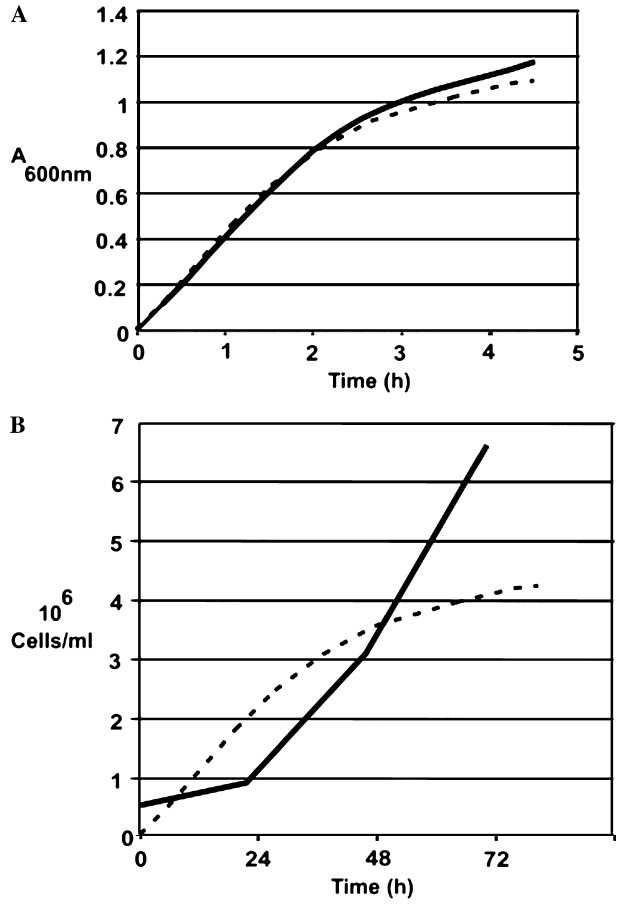


Fig. 2. Growth-curves of (A) *E. coli* and (B) insect cells in shake-flasks and deep-well blocks. A dotted line and a continuous line represent growth in shake-flasks and deep-well blocks, respectively.

*Expression in deep-well blocks*

Vectors transformed into *E. coli* BL21 (DE3) pLysS were grown overnight at 37 °C in 5 ml BHI medium in a 24-well block. Overnight cultures were pelleted at 2000*g* for 5 min using a micro-titer plate centrifuge and re-suspended in 1 ml BHI media. Five microliters of fresh BHI media was inoculated with 20 μl of re-suspended overnight culture and grown at 37 °C for 3–4 h in the 24-well block. Expression was induced at mid-log phase ($A_{600\,nm} \approx 1$) on the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were harvested 6–8 h after induction by centrifugation at 2000*g* for 5 min.

High-5 insect cells used for expression were grown to a density of $2.0 \times 10^6$ cells/ml in a fernbach flask. Insect cell expression in a 24-well block was then initiated by infecting 2.5 ml of the aforementioned cells with a high-titer baculovirus at an MOI of 5 pfu/cell. Cells were then grown in serum-free EX-CELL 405 with L-glutamine at 27 °C for 48–60 h following infection and harvested at 70–80% viability by centrifugation at 2000*g* for 5 min.

*Automated purification*

Purification was carried out with nickel–nitrilotriacetic acid (Ni–NTA) magnetic agarose beads (5% suspension) using the BioRobot 3000 automated liquid handling system (Qiagen, Valencia, CA, USA). A protocol adapted from the manufacturer's manual was used for the purification of cultures grown in the 24-well blocks. Following expression, the cell pellets were re-suspended in 400 μl lysis buffer; 10 mM Tris–HCl (pH 8.0), 50 mM $NaH_2PO_4$, 100 mM NaCl, 20% glycerol, 0.25% Tween 20, and 10 mM imidazole. Lysis in the presence of 0.1% benzonase solution (Novagen, Madison, WI, USA) was performed using a deep-well cup horn sonicator ($4 \times 1$ min bursts) (Misonix, Farmingdale, NY, USA). Cells were separated into soluble and insoluble fractions by centrifugation at 6000*g* for 5 min; the insoluble (pellet) fraction was then solubilized in 400 μl lysis buffer containing 8 M urea. The 400 μl fractions to be purified were transferred 200 μl at a time to a 96-well micro-titer plate containing 20-μl Ni–NTA magnetic–agarose beads, mixed for 1 min, and placed on a 96-well magnet for 1 min, and the supernatant was discarded before the remaining 200 μl was added. The beads were washed with 200 μl lysis buffer and the His-tagged proteins were eluted with 35 μl lysis buffer containing 1 M imidazole after placing the micro-titer plate on the magnet for 1 min. Purified fractions were analyzed and quantified by SDS–PAGE or Western blot.

*Determination of protein yield and analysis*

The levels of protein expression in the soluble and insoluble fractions following centrifugation and purification were estimated by comparison to a range of known protein concentration standards run in parallel on an SDS–PAGE and visualized following staining with Coomassie blue. The molecular weight and pI of full-length kinases (including His-tag) were calculated from their DNA sequence. Codon adaptive index (CAI) for *E. coli* and insect cell expression was calculated using EMBOSS [18]. The Wilkinson–Harrison solubility model was used to predict the solubility of proteins expressed in *E. coli* [19].

## Results and discussion

*Design of pBEV expression vector*

The T7*lac* and *polh* promoters in pBEV direct recombinant expression in *E. coli* and insect cells, respectively. The T7*lac* promoter in pBEV is identical to that present in pET15b, containing a 25 bp *lac* operator sequence immediately downstream from the 17 bp T7 promoter region. The T7*lac* promoter has been constructed with

its ribosome-binding site (RBS) positioned at the optimal distance from the initiation codon of the target gene to ensure efficient translation in *E. coli*. Unmodified in construction, the T7*lac* promoter in pBEV was therefore unlikely to present problems for *E. coli* expression.

Of greater concern was the *polh* promoter's effectiveness in vBEV to direct insect cell expression given the DNA manipulations that had taken place downstream of the promoter. The *polh* promoter had been extensively mapped identifying the absolute requirement of the tetra-nucleotide TAAG sequence, the site of the RNA transcription initiation [20]. Less well understood is the role of the 50 bp region between the RNA starting point and the ATG initiation codon. Possible roles for the untranslated leader region of the *polh* promoter included a putative RBS [21] and additional promoter elements influencing steady-state RNA levels as opposed to initiation of translation [22]. The uncertainty regarding the role of the un-translated RNA leader region has resulted in the whole region being regarded as constituting the *polh* promoter.

Baculoviral vectors employing the *polh* promoter have therefore tended to extend the un-translated leader region to avoid any disruption of the proposed *polh* promoter region. In pVL1392, the *polh* ATG was mutated to ATT and a poly-linker inserted, relocating the ATG 35 bp downstream of the original start codon, producing an un-translated leader region of 85+ bp without adversely interfering with expression levels [23]. Also, there is anecdotal evidence that genes with un-translated leader sequences up to 150 bp have been well expressed using an insect cell expression system [24]. We have explored these observations surrounding the *polh* promoter by extending the (50 bp) un-translated region of *polh* promoter with the (86 bp) encoding the T7*lac* promoter, thereby producing a tandem promoter, designated *polh*:T7*lac* (Fig. 1), capable of functioning in *E. coli* and insect cells.

### *polh promoter vs polh:T7lac promoter*

A comparison was undertaken to contrast the performance of the *polh*:T7*lac* promoter in pBEV with, its progenitor, the *polh* promoter in pBacPAK8. Full-length cyclin activating kinase (CAK1) isolated from *Candida albicans* [25] was cloned into pBEV and pBacPAK8. pBEV-CAK1 and pBacPAK8-CAK1 were co-transfected with linear *Autographa californica* nuclear polyhedrosis viral (AcNPV) DNA into *Spodoptera frugiperda* (Sf9) insect cells to generate the baculovirus recombinants vBEV-CAK1 and vBacPAK8-CAK1.

The results of side-by-side expression and purification of CAK1 produced using vBEV and vBacPAK8 are shown in Fig. 3. Under identical expression and purification conditions, similar levels of CAK1 were produced in insect cells using vBEV-CAK1 and vBacPAK8-CAK1.

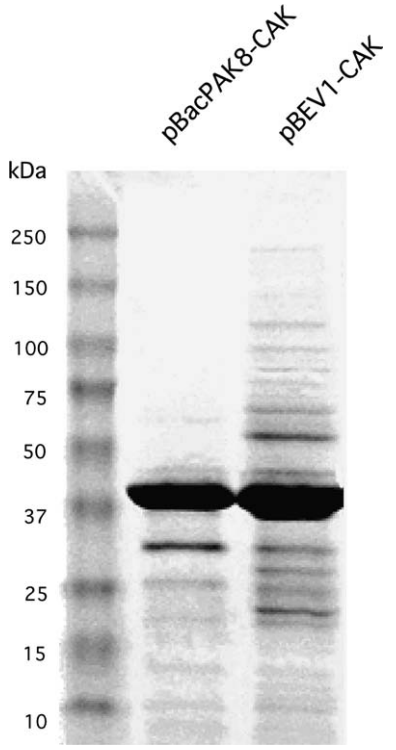


Fig. 3. Comparative expression and purification of CAK1 expressed using vBEV and vBacPAk8. SDS–PAGE analysis of CAK1 following its expression and purification indicated equivalent production between the two systems.

Demonstrating that the addition of the T7 promoter region in the *polh*:T7*lac* promoter of pBEV is not deleterious to *polh* promoter directed expression in insect cells.

### *Expression in deep-well plates vs shake flasks*

A key component in the successful utilization of the pBEV expression vector in achieving high-throughput expression was miniaturization. Consequently, an important part of our validation process was establishing that expression with pBEV could be reduced in volume while still providing results that replicated those seen at larger volumes and was not simply an artifact of scale. This was achieved by analyzing recombinant proteins produced in *E. coli*, purified from 5 ml volume of cultures grown in a 24-well block and a fernbach flask (5 ml from a 1 L culture). Recombinant baculovirus-mediated proteins, produced in insect cells, were purified from 2 ml volume of cultures infected in a 24-well block and a fernbach flask (2 ml from a 800 ml culture). Following lysis, cells were separated into insoluble and soluble fractions and the recombinant protein was purified and analyzed. Purification from equal volumes of cells allowed a direct comparison between the soluble expression in both 24-well blocks and flasks.

The results show that the expression levels and solubility of recombinant proteins expressed in *E. coli* and

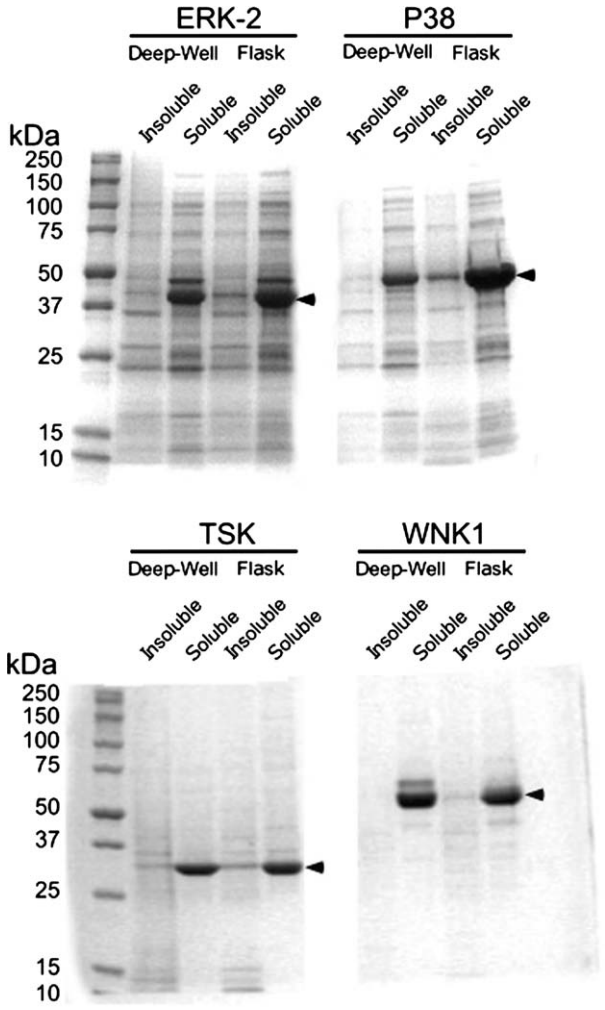


Fig. 4. Expression using pBEV expression vector in *E. coli* and insect cells. Full-length extracellular signal-regulated kinase 2 (ERK-2) and mitogen-activated protein kinase p38α (P38) were expressed and purified from 5 ml *E. coli* cultures grown in 24-well blocks and 2 L fernbach flasks (5 ml extracted from a 1 L culture). Truncated T-cell specific kinase (TSK:G354-L620) and with no K (lysine)-1 (WNK1:P180-G602) were expressed and purified from 2 ml High-5 insect cell cultures infected in 24-well blocks and 2.8 L fernbach flasks (2 ml extracted from 700 ml cultures). Insoluble and insoluble fractions were analyzed for expression and recombinant protein identified.

insect cells are comparable whether produced in a shake flask or a 24-well block (Fig. 4). Although we have employed a crude arbiter of solubility, in examining hundreds of proteins we have consistently found that proteins expressed and soluble in the 24-well blocks were subsequently re-confirmed as soluble when production was scaled-up.

The levels of protein expressed and purified from the soluble or insoluble fractions of *E. coli* and insect cells grown in a deep-well block ranged from 0.1 μg/ml, detected using antibodies recognizing the His-tag epitope, to 20–80 μg/ml readily identified on a Coomassie-stained gel.

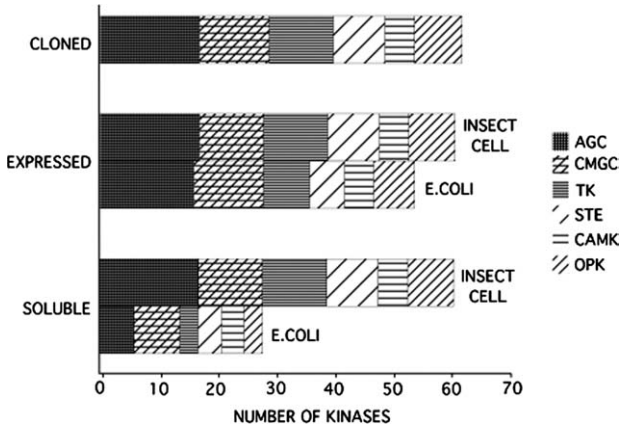


Fig. 5. Histogram of kinases cloned, expressed, and soluble in *E. coli* and insect cells. Kinases examined in this study fall into six major groups, based on sequence and structural similarities, within the kinase superfamily. They are (1) AGC containing PKA, PKG, PKC families; (2) CAMK, calcium/calmodiulin-dependent protein kinase; (3) CK1, casein kinase-1; (4) CMGC containing CDK, MAPK, GSK, CLK families; (5) STE homologs; (6) TK, tyrosine kinase (including tyrosine-like kinase). The remainder OPK—other protein kinases comprise those not falling into previous major groups [31].

### *Screening for expression and solubility*

Once validated and assembled, the high-throughput expression platform was used for the parallel production of 62 full-length human kinases (non-receptor type), ranging in size from 35 to 163 kDa, cloned into pBEV, and expressed in *E. coli* and insect cells. Screening for expression and solubility in either expression system identified those proteins that were successfully expressed and soluble, readily distinguishable from those that either failed to express, were insoluble, or exhibited partial solubility (Fig. 5). The definitions employed for expression and solubility allowed simple classification of the screening results. Successful expression was defined as protein production at or greater than 0.1 μg/ml: yields below this level were considered to be below the limit for practical purification in either native or denaturing conditions.

The solubility of protein expressed was classified into three categories. The first, termed soluble, resulted in the majority of expressed protein being found in the soluble fraction following fractionation. The second termed partially soluble, with protein distributed equally between the soluble and insoluble fractions. The last category termed insoluble had the bulk of the protein expressed found in the insoluble fraction following expression and purification.

In screening 62 kinases while the majority of which were successfully expressed in *E. coli*, many were not expressed in a soluble form and incapable of being purified in their native state. Of the 54 proteins (87%) expressed in *E. coli*, only 29 proteins (54%) were soluble,

with the remaining 25 (46%), either insoluble or exhibiting only partial solubility. Within the larger kinase superfamilies examined, there does appear to be a trend towards greater soluble expression in *E. coli* in the following order:

CAMK > CMGC ≫ STE > AGC > TK.

In developing the technology for high-throughput protein production, many previous studies have used bacterial proteins <23 kDa. Most of these proteins are cytoplasmic and, not surprisingly, soluble when expressed in *E. coli* [26] and in the case of thermophilic bacterial proteins robustly expressed [27]. This bias has resulted in a 46–93% success rate obtaining soluble expression of prokaryotic proteins in *E. coli* [26] compared with 13% soluble expression of eukaryotic proteins in *E. coli* [28]. The percentages reflect the close evolutionary relationship between the genes expressed and the natural host. The complexity of eukaryotic proteins would appear best served when produced in eukaryotic hosts, with successful reports of high-throughput expression of eukaryotic proteins using yeast [10] and insect cells [12]. In our expression screen of 62 human kinases in insect cells, all but one of the kinases screened were expressed and soluble. The 99% success rate expressing human kinases in insect cells is significantly higher than the 54% success rate achieved when expressing the same proteins in *E. coli* and demonstrates the benefit of insect cells in the production of eukaryotic proteins.

*Analysis of expression screening*

Unlike insect cell expression, the variability in expression and protein solubility exhibited in *E. coli* provided data from which to identify biophysical characteristics potentially responsible for any observed differences in protein expression and solubility (Fig. 6). Analysis of the data generated by the *E. coli* expression screen revealed a correlation between successful expression in *E. coli* and decreasing protein size (Fig. 7A); this had been previously observed in the expression of human proteins in *E. coli* [29]. Protein solubility in *E. coli* also appeared directly related to the size of the protein expressed (Fig. 7B), with a preference for proteins <50 kDa being soluble. Reduction in expression and solubility with protein size has been observed in the expression of the thermophilic bacterium *Thermotoga maritima* genome in *E. coli* [30]. This limitation of *E. coli* expression will have significant consequences in the expression of the human genome. With an average molecular weight of 52 kDa [31], the majority of the human proteome produced in *E. coli* is likely to be insoluble, if successfully expressed.

Of the eight proteins that failed to express in *E. coli*, four were large proteins (>100 kDa) containing a high proportion of rare *E. coli* codons. The accompanying

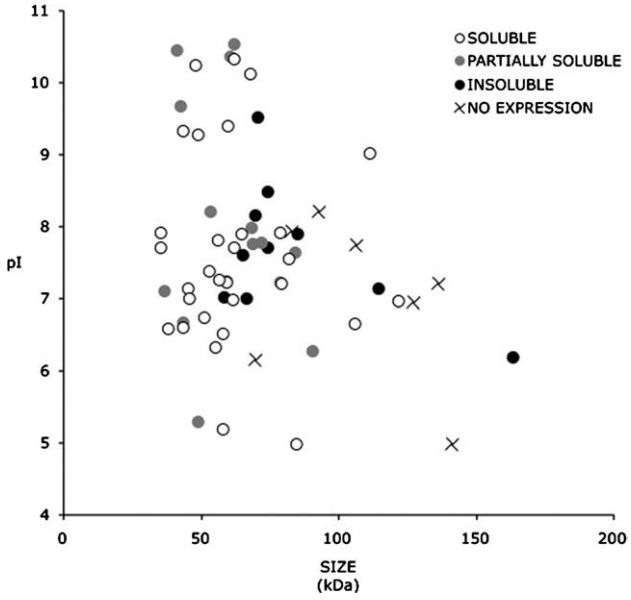


Fig. 6. Predicted molecular weight and isoelectric point for kinases expressed in *E. coli*. Scatter plot identifying kinase expression resulting in soluble (○), partially soluble (●), insoluble (●), and no expression (×).

low CAI is often cited for failed or low expression of mammalian genes in *E. coli* [32]. The remaining four proteins that failed to express in *E. coli* were moderately sized proteins (<100 kDa) with higher CAI values, suggesting other factor(s) also impact expression efficiency. The only kinase that failed to express in insect cells was DYRK3, a moderately sized protein (67.9 kDa) with a high p*I* (10.12), which was successfully expressed and soluble in *E. coli*.

The experimental results also failed to conform to the Wilkinson–Harrison model, a predictive algorithm for soluble *E. coli* expression that is based on protein parameters [19]. Equally confounding was the lack of accuracy of the CAI in determining successful expression in either *E. coli* or insect cells. The unpredictable outcome of protein expression and the failure of the available models to forecast protein expression behavior necessitate experimentation. The strategy of parallel processing of *E. coli* and insect cell expression, rapidly generating empirical data, allows the identification of the most tractable protein and expression system. Subsequent, comparative analysis of the expression data can be used for both target prioritization in production and downstream insurance of the maximum efficiency of resources. The overall poor performance of the *E. coli* expression demonstrated in our study is negated through our ability to identify soluble well-expressed protein by screening. This tactic enables *E. coli*, producing soluble well-expressed protein, to make a crucial contribution to our overall strategy.

To obtain high-throughput, we eschewed the customized approach where protein expression is optimized for

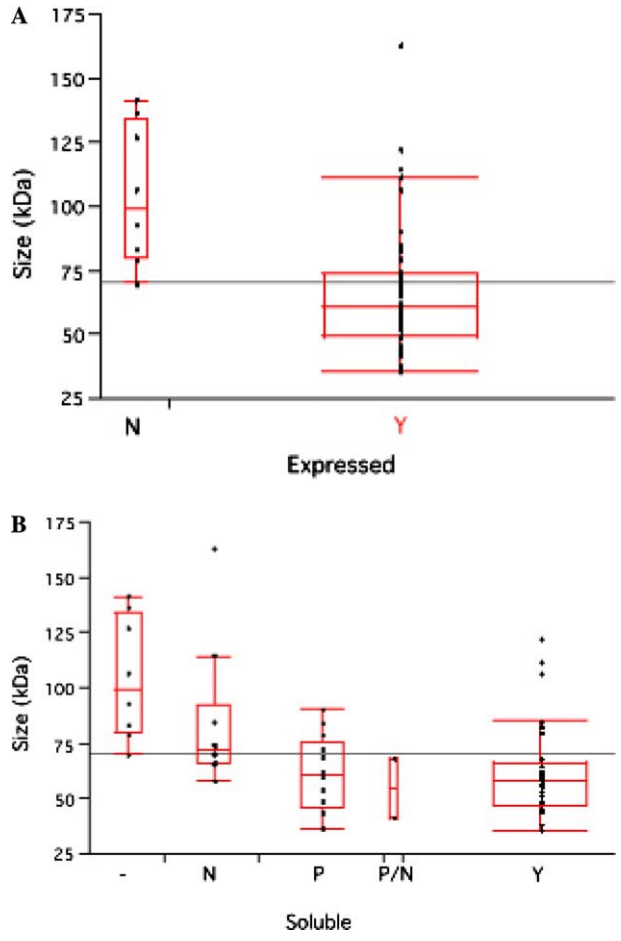


Fig. 7. Data analysis of expression and solubility in *E. coli*. Analysis revealed correlation between protein size, its ability to be expressed (A), and its solubility (B) in *E. coli*. Protein solubility identified as soluble (Y), partially soluble (P), and insoluble (N). Statistical analysis performed using JMP-4 software (SAS Institute, Cary, NC, USA).

each individual protein. We instead chose a limited set of conditions designed to produce optimum levels of expression that were amenable to scale-up for the majority of proteins, aware that a minority of proteins would be poorly expressed. The results obtained reflect the conditions of the experiment and while the unsuccessful outcome of expression could potentially be reversed through further optimization, our overall goal precludes time-consuming customization of production and places a premium on the predictable delivery of biological reagents.

pBEV, and its accompanying expression platform, has been successfully deployed to express thousands of cDNAs in *E. coli* and insect cells, successfully generating hundreds of proteins for both enzyme characterization [33] and structure determination [34–36].

## Acknowledgments

We acknowledge members of the Molecular Cloning group at Vertex Pharmaceuticals for cloning the cDNA used in this study and thank Jocelyn Corradi and Jim Hogan for their technical assistance. We also wish to thank Ted Fox and Scott Raybuck for valuable discussions and critical reading of the manuscript.

## References

[1] A. Lueking, C. Holz, C. Gotthold, H. Lehrach, D. Cahill, A system for dual protein expression in *Pichia pastoris* and *Escherichia coli*, Protein Expr. Purif. 20 (2000) 372–378.
[2] R. Novy, K. Yaeger, S. Monsma, M. Scott, in: InNovations, 1999, pp. 1–5.
[3] A.J. Walhout, G.F. Temple, M.A. Brasch, J.L. Hartley, M.A. Lorson, S. van den Heuvel, M. Vidal, GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes, Methods Enzymol. 328 (2000) 575–592.
[4] Q. Liu, M.Z. Li, D. Leibham, D. Cortez, S.J. Elledge, The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes, Curr. Biol. 8 (1998) 1300–1309.
[5] F.A. Marston, The purification of eukaryotic polypeptides synthesized in *Escherichia coli*, Biochem. J. 240 (1986) 1–12.
[6] J.M. Vlak, R.J. Keus, Baculovirus expression vector system for production of viral vaccines, Adv. Biotechnol. Process. 14 (1990) 91–128.
[7] R. Wetzel, L.J. Perry, C. Veilleux, Mutations in human interferon gamma affecting inclusion body formation identified by a general immunochemical screen, Biotechnology (N. Y.) 9 (1991) 731–737.
[8] R.K. Knaust, P. Nordlund, Screening for soluble expression of recombinant proteins in a 96-well format, Anal. Biochem. 297 (2001) 79–85.
[9] K. Bussow, E. Nordhoff, C. Lubbert, H. Lehrach, G. Walter, A human cDNA library for high-throughput protein expression screening, Genomics 65 (2000) 1–8.
[10] C. Holz, O. Hesse, N. Bolotina, U. Stahl, C. Lang, A micro-scale process for high-throughput expression of cDNAs in the yeast *Saccharomyces cerevisiae*, Protein Expr. Purif. 25 (2002) 372–378.
[11] J.A. Heyman, J. Cornthwaite, L. Foncerrada, J.R. Gilmore, E. Gontang, K.J. Hartman, C.L. Hernandez, R. Hood, H.M. Hull, W.Y. Lee, R. Marcil, E.J. Marsh, K.M. Mudd, M.J. Patino, T.J. Purcell, J.J. Rowland, M.L. Sindici, J.P. Hoeffler, Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation, Genome Res. 9 (1999) 383–392.
[12] J.S. Albala, K. Franke, I.R. McConnell, K.L. Pak, P.A. Folta, B. Rubinfeld, A.H. Davies, G.G. Lennon, R. Clark, From genes to proteins: high-throughput expression and purification of the human proteome, J. Cell Biochem. 80 (2000) 187–191.
[13] F.W. Studier, B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, J. Mol. Biol. 189 (1986) 113–130.
[14] G.E. Smith, M.D. Summers, M.J. Fraser, Production of human beta interferon in insect cells infected with a baculovirus expression vector, Mol. Cell. Biol. 3 (1983) 2156–2165.
[15] S.P. Chambers, J.K. Brehm, N.P. Michael, T. Atkinson, N.P. Minton, Physical characterisation and over-expression of the *Bacillus caldotenax* superoxide dismutase gene, FEMS Microbiol. Lett. 70 (1992) 277–284.
[16] F.W. Studier, Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system, J. Mol. Biol. 219 (1991) 37–44.
[17] A.I. Derman, W.A. Prinz, D. Belin, J. Beckwith, Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*, Science 262 (1993) 1744–1747.
[18] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite, Trends Genet. 16 (2000) 276–277.

[19] D.L. Wilkinson, R.G. Harrison, Predicting the solubility of recombinant proteins in *Escherichia coli*, Biotechnology (N. Y.) 9 (1991) 443–448.
[20] S.M. Thiem, L.K. Miller, Identification, sequence, and transcriptional mapping of the major capsid protein gene of the baculovirus *Autographa californica* nuclear polyhedrosis virus, J. Virol. 63 (1989) 2008–2018.
[21] Y.W. Hu, C.Y. Kang, Enzyme activities in four different forms of human immunodeficiency virus 1 pol gene products, Proc. Natl. Acad. Sci. USA 88 (1991) 4596–4600.
[22] B.G. Ooi, C. Rankin, L.K. Miller, Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene, J. Mol. Biol. 210 (1989) 721–736.
[23] V.A. Luckow, M.D. Summers, High level expression of nonfused foreign genes with *Autographa californica* nuclear polyhedrosis virus expression vectors, Virology 170 (1989) 31–39.
[24] D.R. O'Reilly, L.K. Miller, V.A. Luckow, Baculovirus Expression Vectors: A Laboratory Manual, W.H. Freeman, New York, 1992.
[25] V. Tsakraklides, M.J. Solomon, Comparison of Cak1p-like cyclin-dependent kinase-activating kinases, J. Biol. Chem. 277 (2002) 33482–33489.
[26] A. Yee, X. Chang, A. Pineda-Lucena, B. Wu, A. Semesi, B. Le, T. Ramelot, G.M. Lee, S. Bhattacharyya, P. Gutierrez, A. Denisov, C.H. Lee, J.R. Cort, G. Kozlov, J. Liao, G. Finak, L. Chen, D. Wishart, W. Lee, L.P. McIntosh, K. Gehring, M.A. Kennedy, A.M. Edwards, C.H. Arrowsmith, An NMR approach to structural proteomics, Proc. Natl. Acad. Sci. USA 99 (2002) 1825–1830.
[27] D. Christendat, A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J.R. Cort, V. Booth, C.D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K.L. Maxwell, N. Wu, L.P. McIntosh, K. Gehring, M.A. Kennedy, A.R. Davidson, E.F. Pai, M. Gerstein, A.M. Edwards, C.H. Arrowsmith, Structural proteomics of an archaeon, Nat. Struct. Biol. 7 (2000) 903–909.
[28] 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.
[29] H.T. Ding, H. Ren, Q. Chen, G. Fang, L.F. Li, R. Li, Z. Wang, X.Y. Jia, Y.H. Liang, M.H. Hu, Y. Li, J.C. Luo, X.C. Gu, X.D. Su, M. Luo, S.Y. Lu, Parallel cloning, expression, purification and crystallization of human proteins for structural genomics, Acta Crystallogr. D. Biol. Crystallogr. 58 (2002) 2102–2108.
[30] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline, Proc. Natl. Acad. Sci. USA 99 (2002) 11664–11669.
[31] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[32] S.P. Zhang, G. Zubay, E. Goldman, Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates, Gene 105 (1991) 61–72.
[33] C.H. Gross, J.D. Parsons, T.H. Grossman, P.S. Charifson, S. Bellon, J. Jernee, M. Dwyer, S.P. Chambers, W. Markland, M. Botfield, S.A. Raybuck, Active-site residues of *Escherichia coli* DNA gyrase required in coupling ATP hydrolysis to DNA supercoiling and amino acid substitutions leading to novobiocin resistance, Antimicrob. Agents Chemother. 47 (2003) 1037–1046.
[34] E. ter Haar, J.T. Coll, D.A. Austen, H.M. Hsiao, L. Swenson, J. Jain, Structure of GSK3beta reveals a primed phosphorylation mechanism, Nat. Struct. Biol. 8 (2001) 593–596.
[35] G.M. Cheetham, R.M. Knegtel, J.T. Coll, S.B. Renwick, L. Swenson, P. Weber, J.A. Lippke, D.A. Austen, Crystal structure of aurora-2, an oncogenic serine/threonine kinase, J. Biol. Chem. 277 (2002) 42419–42422.
[36] W. Meng, L.L. Swenson, M.J. Fitzgibbon, K. Hayakawa, E. ter Haar, A.E. Behrens, J.R. Fulghum, J.A. Lippke, Structure of mitogen-activated protein kinase-activated protein (MAPKAP) kinase 2 suggests a bifunctional switch that couples kinase activation with nuclear export, J. Biol. Chem. 277 (2002) 37401–37405.