Journal of Structural and Functional Genomics **6**: 13–20, 2005.
DOI 10.1007/s10969-005-0462-7
© Springer 2005

Pilot studies on the parallel production of soluble mouse proteins in a bacterial expression system

Nathan P. Cowieson[1], Pawel Listwan[2,3,5], Mareike Kurz[1], Anna Aagaard[1], Timothy Ravasi[1,3,5], Christine Wells[1,3,5], Thomas Huber[4], David A. Hume[1,3,5], Bostjan Kobe[1,2,5] & Jennifer L. Martin[1,2,5,*]

[1]*Institute for Molecular Bioscience;* [2]*School of Molecular and Microbial Sciences;* [3]*Cooperative Research Centre for Chronic Inflammatory Diseases;* [4]*Department of Mathematics;* [5]*ARC Special Research Centre for Functional and Applied Genomics, Queensland Bioscience Precinct, University of Queensland, 306 Carmody Road, St Lucia, Brisbane, Queensland 4072, Australia; *Author for correspondence (Fax: +61-7-3346-2101; e-mail: j.martin@imb.uq.edu.au)*

Received 12 October 2004; accepted in revised form 5 January 2005

Key words: bacterial expression, mammalian proteins, medium-throughput, protein expression screening, recombinatorial cloning

Abstract

We investigated the parallel production in medium throughput of mouse proteins, using protocols that involved recombinatorial cloning, protein expression screening and batch purification. The methods were scaled up to allow the simultaneous processing of tens or hundreds of protein samples. Scale-up was achieved in two stages. In an initial study, 30 targets were processed manually but with common protocols for all targets. In the second study, these protocols were applied to 96 target proteins that were processed in an automated manner. The success rates at each stage of the study were similar for both the manual and automated approaches. Overall, 15 of the selected 126 target mouse genes (12%) yielded soluble protein products in a bacterial expression system. This success rate compares favourably with other protein screening projects, particularly for eukaryotic proteins, and could be further improved by modifications at the cloning step.

Introduction

The use of high-throughput methods is becoming increasingly common for producing protein samples for structural biology and functional studies. The methods used vary depending on the scale of the project, with high-throughput programs (such as those at the Joint Centre for Structural Genomics (JCSG) [1–3] or RIKEN [4]), processing thousands of samples in a highly automated manner, using robotics and custom-built apparatus. With these approaches, it becomes inefficient to optimise experimental variables on a case-by-case basis. By contrast, medium-throughput approaches [5] process a smaller number of targets – tens to hundreds of samples – using a combination of robotics and manual methods. This allows for more flexibility and case-by-case optimisation, which can increase the overall percentage of starting clones that yield soluble protein. However, the yield improvement occurs at the expense of throughput.

The percentage of targets in a protein production pipeline that yield soluble protein in bacterial expression systems is dependent on target selection criteria, experimental methods, the amount of protein that is required [6] and on the host organism from which the targets were selected. This is best illustrated by comparing across several structural genomic target organisms the number of cloned targets from which soluble protein is expressed (TargetDB http://target-db.pdb.org/ accession date 9 September 2004) [7]. The highest success rate is achieved when

expressing *Escherichia coli* (*E. coli*) proteins in *E. coli*; 37% of 731 clones resulted in soluble protein. Other bacteria are slightly less successful, with 28% of 1733 *Mycobacterium tuberculosis* (*M. tuberculosis*) clones yielding soluble protein. Yeast is less successful again with 18% of 1221 *Saccharomyces cerevisiae* (*S. cerevisiae*) clones expressing soluble protein. Finally, proteins from mammalian hosts are the least amenable to bacterial over-expression, with only 14% of 1665 human clones and 17% of 717 mouse proteins expressing soluble protein. However, there are inconsistencies or ambiguities in the figures quoted in target databases. For example, there are more *S. cerevisiae* clones than initial targets. This could represent incomplete or erroneous data entry or that multiple clones were made from one target i.e. several different domains or different fusion tags. Similarly, more proteins were purified from human, *E. coli* and *Thermotoga maritima* (*T. maritima*) than were expressed in soluble form. This could indicate that some proteins were refolded from inclusion bodies or may suggest data entry inconsistencies.

Several properties of eukaryotic organisms contribute to the higher attrition rates in going from cloned targets to soluble protein. First, eukaryotic genes are relatively complex when compared to prokaryotic genes. This makes prediction of open reading frames more difficult and can result in spurious targets. In addition, the presence of introns makes it more difficult to obtain a template for amplifying eukaryotic genes. While prokaryotic structural genomics projects often use genomic DNA – which is relatively easy to prepare and which contains every gene – as a template for PCR [2, 5, 8, 9], eukaryotic projects most often rely on cDNA clone libraries as a template source [4, 10, 11]. For this reason, eukaryotic structural genomics projects have tended to work in close collaboration with curators of cDNA library resources. Second, eukaryotic proteins express more poorly than prokaryotic proteins in a prokaryotic host organism such as *E. coli*. This can be due to differences in the way nascent proteins are processed in the two organisms. Eukaryotic proteins fold more slowly than bacterial proteins [12], often require specific chaperones [13], post-translational modifications and processing in the Golgi or endoplasmic reticulum compartments, and frequently have different codon usage from a bacterial host [14]. For these reasons bacterial [4] and wheatgerm [15] cell free expression systems, yeast [16] and insect cells [17] are often used as alternatives to *E. coli* for expression of eukaryotic proteins.

One way to counteract the high attrition rates when expressing mammalian proteins in *E. coli* is to start with larger numbers of clones, as is done in the high-throughput approaches. In this way, the total number of soluble proteins produced is increased, although the percentage success remains the same. What is not yet clear is whether medium-throughput approaches – which are known to be useful for expressing bacterial proteins [5] – are suitable for the production of protein samples from mammalian hosts.

We developed a medium-throughput approach for the expression of mammalian proteins in *E. coli*. With this approach, we targeted proteins that are likely to play a role in the immune response and inflammation; the genes corresponding to the selected proteins had been shown, using microarray experiments, to be transcriptionally regulated following exposure of mouse macrophages to the bacterial cell wall component lipopolysccharide (LPS) [18, 19]. This population of proteins includes many of unknown structure and function. Producing these proteins and determining their structures will thus provide insight into their mechanisms of action and will also be an important starting point for future drug design research.

We selected the target proteins using a hierarchy of criteria. Firstly, targets were chosen based on biological criteria, so that only those genes exhibiting greater than twofold regulation in response to LPS treatment were considered. This is sufficiently above the noise level of microarray experiments to be statistically significant [18]. In addition, only genes with human orthologues (with a minimum of 70% sequence identity) were considered. This criterion filters out wrongly predicted open reading frames and simultaneously ensures that the selected proteins will be relevant to humans. Secondly, targets were filtered based on their predicted likelihood to express well in bacteria. To achieve this, protein sequence analysis algorithms (Table 1) were used to reject very large or very small proteins, and to exclude sorting signals and transmembrane regions from the expression constructs. Proteins annotated to be integral components of larger complexes were similarly

Table 1. Sequence-related criteria used for target selection.

Criteria	Action	Reason
Larger than 600 amino acids	Discard	Lower success rate of expression in bacteria
Less than 100 amino acids	Discard	Likely to be wrongly predicted open reading frames, or not contain defined structure
Presence of transmembrane helices predicted by TMHMM [19]	Avoid this region when cloning	Expected to be insoluble in aqueous solutions and technically difficult to work with
Presence of a signal sequence predicted by PSORT [20]	Avoid this region when cloning	Eukaryotic signal sequences are usually not cleaved by *E. coli* (leading to insolubility due to their hydrophobic character)
Part or all of the protein having greater than 30% sequence identity with a known protein structure	Discard or avoid homologous region when cloning	The structures may be predicted using homology modelling
Extracellular and cysteine-rich	Discard	Likely to contain structurally important disulphide bonds unlikely to form in *E. coli* expression systems

filtered out of the pipeline. Finally, to maximise the value of the determined structures, proteins for which structures could be modelled based on homology (>30% sequence identity to known structures) were removed. A summary of the selection criteria is shown in Table 1.

We present here the results of two rounds of parallel protein production. In the first round, 30 targets were chosen for the implementation and optimisation of techniques. These methods were then applied to a further 96 targets using a 96-well plate format. We show that the success rates of our approach compare favourably with other protein expression screening projects, particularly for eukaryotic proteins, and could be further improved by modifications at the cloning step.

Materials and methods

Target selection

Protein sequences corresponding to the longest open reading frame of RIKEN mouse cDNA [18] representative clones were selected. The sequences of these predicted proteins were analysed to assess their suitability for expression in *E. coli*. Transmembrane domains were predicted using the TMHMM [19] algorithm. Signal and sorting sequences were predicted using the PSORT [20] algorithm. Sequences with known structures were identified by sequence comparison with the sequences from the Protein Data Bank (BLAST [21]) as well as by threading and secondary structure prediction using 3D-PSSM [22]. Known domains within proteins were identified using PFAM [23].

Gene amplification

Target genes were amplified by PCR using cDNA prepared from LPS-stimulated mouse macrophage cells as a template. The reactions were catalysed by Triplemaster proofreading, blunt-ended polymerase mix (Eppendorf) using primers at a concentration of 2 ng/μl. The PCR conditions were: 95 °C for 2 min, 30 cycles of 95 °C for 30 s, 55 °C for 30 s, 68 °C for 1 min, and finally one cycle of 68 °C for 7 min. PCR products were purified from primers and buffer components using Montage 96-well PCR purification kits (Millipore).

Cloning

Purified PCR products were cloned into the Gateway 'entry vector' pENTR-D-TOPO (Invitrogen) following the manufacturer's instructions. Solutions containing the construct were then

transformed into chemically competent DH5α cells by heat shock. Colonies were picked by hand either into 3 ml cultures in 15 ml conical bottomed tubes for the 30 target trial or into 1 ml cultures in a 96 well 'deep-well' plate (Greiner) for the 96 target trial and grown overnight with shaking at 37 °C. Plasmids were purified from these cultures using either single tube Wizard plasmid purification kits (Promega) for the 30 target trial or Aurum 96-well plasmid purification kits (BioRad) for the 96 target trial. Genes cloned into pENTR-D-TOPO were recombined into the expression vector pDEST-17 (Invitrogen) by the Gateway LR reaction following manufacturer's instructions. Entry and expression vectors were assayed for correct insertion of the gene by digestion with the restriction enzyme BsrG1, followed by electrophoresis.

Expression and solubility testing

Expression vectors were transformed into chemically competent *E. coli* strain BL21(DE3)pLysS by heat shock. Proteins were expressed for 24 h at 25 °C in either 5 ml culture volumes in 50 ml conical tubes or in 0.5 ml culture volumes in a 96-well 2 ml deep-well plate (Greiner) using autoinduction media (W. Studier, personal communication). After pelleting and aspiration of media, cells were lysed by incubating for 10 min in B-PER cell lysis reagent (Pierce) containing 50 units/ml benzonase (Sigma). A sample of this whole cell extract was taken for SDS-PAGE analysis. Insoluble material was pelleted from the cell lysate by centrifugation at 4000 g, and a sample of the soluble extract taken for SDS-PAGE analysis. The soluble fraction was made up to a final concentration of 10 mM imidazole and incubated for 30 min with 50 µl TALON metal affinity resin (BD Bioscience) at 4 °C with agitation. The resin was pelleted and washed 3 times by resuspending in wash buffer (50 mM HEPES (pH 7.4), 300 mM NaCl, 10 mM imidazole), pelleting the resin and removing the buffer. Finally, protein bound to the resin was eluted by adding 50 µl of elution buffer (50 mM HEPES (pH 7.4), 300 mM NaCl, 250 mM imidazole). A sample of the eluted protein was taken for analysis by SDS-PAGE. The samples of whole cell extract, soluble protein and protein eluted from TALON resin were run side by side on 14% SDS-PAGE gels.

Results

A pilot study focussing on 30 mouse proteins was undertaken to investigate the feasibility of using medium-throughput techniques for the cloning and test expression of mouse genes in *E. coli*. Care was taken to use liquid handling steps that were readily applicable to processing larger numbers of targets. PCR primers were designed with calculated annealing temperatures above 55 °C. A single PCR programme was used for all genes without case-by-case optimisation, involving a 55 °C annealing step. Using this approach, 20 PCR products were obtained from 30 reactions (67% success rate). Lowering the annealing temperature or using a hot start PCR protocol did not increase the overall success rate.

PCR products were cloned into the vector pENTR-D-TOPO (Invitrogen) by a topoisomerase-mediated cloning strategy. The major advantage of this cloning method is that there is a minimal requirement for extra nucleotides in the PCR primers (only CACC at the 5′ end of the sense primer). Using this method, 75% of the PCR products were successfully cloned into the vector. However, a major limitation of this cloning strategy is a high rate of false positive clones, resulting in colonies that carry empty vector with no cloned gene insert. Indeed, approximately four colonies out of five contained empty vector. To achieve a 75% success rate for this step required plasmid purification and restriction digests for many independent colonies. Sequencing a sample of the empty vectors revealed that a deletion had occurred between the topoisomerase binding site and the middle of the toxic *ccdb* gene that prevents false positive transformation in these vectors. The mechanism by which this deletion occurs is unclear, though when a PCR product purification step was included, the percentage of empty vector clones fell from 90% to 75%, suggesting that the presence of primers may contribute to the problem.

The recombination based Gateway system (Invitrogen) was chosen for cloning the mouse targets. This system does not rely on restriction digestion and ligation and is therefore largely sequence-independent. In addition, cloned targets can be efficiently moved between expression vectors, allowing the incorporation of various fusion tags and the use of different expression hosts at a later time.

The Gateway LR reaction was used to recombine genes from pENTR-D-TOPO into the pET based expression vector pDEST-17 (Invitrogen). This vector allows expression of fusion proteins with amino terminal hexahistidine tags. This reaction was found to be extremely efficient. More than 98% of transformed colonies screened by restriction digestion contained vector with an insert of the expected size. In addition, all of the targets in entry vectors gave large numbers of colonies following the LR reaction and transformation into *E. coli*.

After transformation of the expression vectors into BL21(DE3)pLysS cells, colonies were picked into 5 ml autoinduction medium in 50 ml conical bottomed tubes. This medium, containing a balance of glucose and lactose, allows the cells to grow to high optical densities before inducing expression of the recombinant protein (Studier, personal communication). Cells were grown to saturation for 24 h at 25 °C. The bacteria were lysed using B-PER reagent (Pierce) that contains a non-ionic detergent. The detergent lysis step does not appear to alter the solubility profile of expressed proteins, since all of the proteins that were classified as soluble in the small scale experiments (for this and the 96-target trial) were also soluble in subsequent large scale expression and purification using detergent free buffers. Whole cell extracts, soluble fractions and protein eluted from metal affinity resin after small scale batch purification were analysed by SDS-PAGE for each target protein (Figure 1).

Fifteen target genes were analysed in this way and 11 (73%) of these expressed well in *E. coli*. Of the 11 proteins that could be over-expressed in *E. coli*, three (27%) were soluble. Protein test expressions at 37, 15 °C and in the codon-supplemented Rosetta(DE3)pLysS strain (Novagen) did not improve the success rates of protein expression or solubility.

The results from this small scale trial suggested that while the topoisomerase mediated cloning is convenient, the high rate of false positive clones and the need for optimised vector to insert ratios may be a problem for cloning on a larger scale. The study demonstrated that with common protocols, 10% of an initial starting set of mouse genes yielded soluble protein from a bacterial expression system. By extrapolation, a 96-well plate would be expected to give 9–10 soluble mouse proteins using the same protocols.

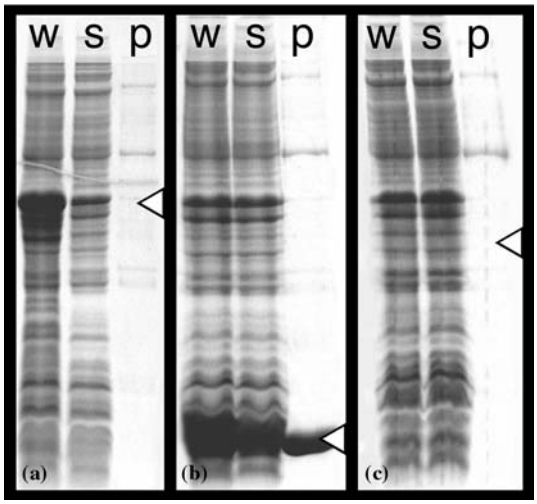

Figure 1. SDS-PAGE analysis of the expression results for selected proteins. W, whole cell extract; S, soluble fraction; P, protein eluted from metal affinity resin. Arrows indicate where the protein is expected to run according to the calculated molecular weight. Track (a) shows an example of protein that expressed but is insoluble, track (b) shows a protein that is expressed and soluble, and track (c) shows a protein that did not express.

To further test the methods and to evaluate the predicted success rate, a larger scale project was undertaken with 96 new targets. These targets were selected from the same population of LPS-induced genes as the initial 30 targets. PCR reactions were carried out in a 96-well, thin wall PCR plate. Reaction conditions were identical to those used for the 30-target trial. Under these conditions, 73% (70 of 96 targets) produced PCR products – a similar success rate to that found for the 30-target trial (68%).

PCR products were purified in 96-well format and cloned into the pENTR-D-TOPO vector by the same topoisomerase-mediated reaction. Clones were purified from transformed *E. coli* using a 96-well plasmid miniprep kit. After digestion analysis, one miniprep in five was found to contain a cloned gene. After many minipreps, 55 of the 70 PCR products (79%) were successfully cloned into plasmid vectors. All 55 were transferred to pDEST17 expression vectors by the LR reaction. Once again the success rate at this step was similar to that for the 30-target trial (75%).

Carrying out expression trials in 50 ml conical tubes becomes difficult with so many targets to be expressed. For this reason the trials were scaled

down to 500 µl autoinduction culture in 2 ml 'deep-well' 96-well plates. The cultures were grown for 24 h at 25 °C with very rapid shaking to ensure good aeration. Whole cell extracts, soluble fractions and metal affinity purified protein were analysed by SDS-PAGE as described above. Protein was expressed from 43 of the 55 constructs (78%) and 12 of these 43 (27%) were found to be soluble (73% and 27% for the 30-target trial). Overall, the success rate for soluble protein production was 12.5% for the 96-target trial and 10% for the 30-target trial.

Discussion

We evaluated a medium-throughput approach for the parallel production of mouse proteins expressed in *E. coli*. The protocol was scaled up in two phases. In the first phase, manual methods were applied to 30 protein targets to establish the validity of using common protocols (for PCR, cloning, expression and purification) for the parallel production of the mouse proteins in a bacterial expression system. In the second phase, the same protocols were used but reagent volumes were scaled down and the throughput was scaled up to allow 96 targets to be processed using microtitre 96-well plates. Figure 2 compares the results for the two trials. It can be seen that the efficiencies at each step were similar in both trials. This result indicates that there was no reduction in efficiency when scaling from manual to medium-throughput methods, indicating that the more automated approach for mammalian protein expression screening is a valid means to increase throughput.

The overall success of parallel throughput protein expression protocols – the total percentage of soluble proteins produced – is dependent on the efficiency or success rate at each step in the process. The number of soluble proteins can therefore be maximised by maximising the success rate at each step in the process. However, increasing the efficiency of any one step may come at the expense of throughput. The results of other structural genomics projects therefore provide a useful benchmark for determining which steps in the procedure would most benefit from further optimisation.

Table 2 shows the published number and percentage of targets that pass the cloning and solubility steps in the high-throughput protein production pipeline for several different organisms. For the first step evaluated, cloning success, there is a wide range of efficiencies ranging from a low of 56% for this project to a high of 89% for *E. coli*. This variable success rate is most likely to reflect the chosen cloning methods rather than any property inherent to the organism. Therefore, a success rate that is significantly lower than the average – such as for the mouse protein project described here (56%) – indicates that additional optimisation of this step is worthwhile. There are

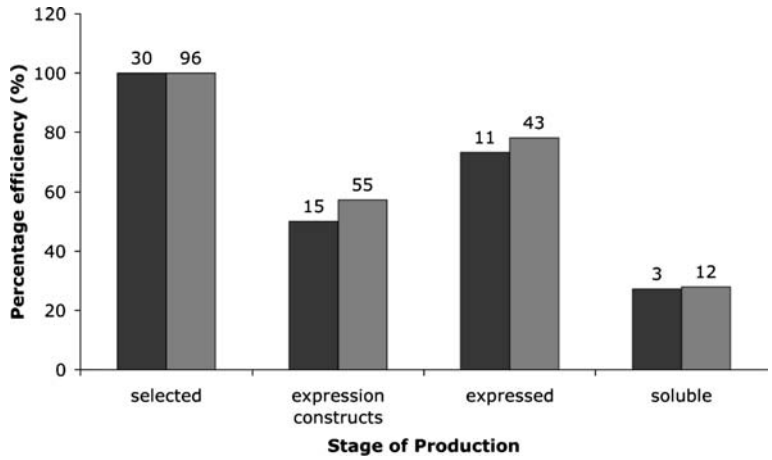

Figure 2. Comparison of protein expression screening success – manual versus 96-well plate. The percentage success for each step in the protein production pipeline is shown for both the initial trial of 30 targets and the subsequent trial with 96 targets. The number of targets represented at each stage is written above the bars. Results from the 30-target trial are represented by the black bars and results from the 96-target trial are in dark grey.

Table 2. Comparison of protein expression screening statistics for several organisms[a].

Target organism	Initiative	Targets	Expression constructs	Total expressed protein	High yield or purified protein	Reference
Homo sapiens	Berlin Structure Factory	599(100%)	359(60%) 60%	ND	113(19%) 32%	[11]
Caenorhabditis elegans	New York Structural Genomics Consortium	86(100%)	86(100%)*100%	53(62%) 62%	3(3.5%) 6%	[10]
Mus musculus	University of Queensland	126(100%)	70(56%)56%	54(43%)77%	15(12%)28%	This study
Thermotoga maritima	Joint Center for Structural Genomics	1877(100%)	1376(73%)73%	ND	542(29%)40%	[2]
Methanobacterium thermoautotrophicum	North Eastern Structural Genomics	393(100%)	393(100%)*100%	327(83%)83%	148(38%)45%	[9]
Escherichia coli	CNRS/ University of Marseille	108(100%)	96(89%)89%	70(64%)73%	32(30%)46%	[5]
Bacillus subtilis	Midwest Center for Structural Genomics	862(100%)	750(87%)87%	488(57%)65%	343(40%)70%	[22]

[a]The first number in each cell represents the total number of targets at each stage. The number in brackets is the cumulative success rate, that is the percentage success as a function of the total number of targets. The number on the right represents the success rate at each step, that is the percentage success as a function of the number of targets brought through from the previous step (for which there is data available). The cloning results marked with an asterisk are from projects that used target sets of precloned genes and these are 100% by definition. ND indicates no data available.

two obvious possibilities for increasing cloning efficiency in our system. First, the PCR amplification step could be improved by using cDNA clones from a mouse library, such as the FANTOM2 collection [20], rather than crude cDNA preparations. Second, the topoisomerase mediated cloning step, which proved to be problematic due to the high incidence of false positive clones, could be replaced with an alternative cloning strategy.

The next stage, expression of proteins from the expression constructs, had an average success rate of 72%, with our approach doing better than average with a 77% success rate. As might be expected, the highest success rates were found for the prokaryotic organisms, *E. coli*, *T. maritima* and *M. thermoautotrophicum* for which 65–83% of cloned genes gave rise to a protein product. By contrast, the higher organism *C. elegans* had only a 62% success rate. Given the higher than average success rate for our work at this step, further optimisation in the protocol is not likely to improve bacterial expression, although for individual proteins it might be worthwhile investigating alternative expression systems.

The step with by far the highest attrition rate for all the protein production projects – whether medium or high throughput, prokaryotic or eukaryotic – is the percentage of soluble proteins produced relative to the total number of proteins expressed (Table 2). This appears to be particularly problematic for the expression of proteins from multicellular eukaryotic organisms (human and *C. elegans*) in *E. coli* where only 6–32% of proteins are soluble (Table 2). By comparison, our success rate of 28% for soluble mouse proteins expressed in bacteria is excellent and may reflect the target selection criteria. Furthermore, given that no other eukaryotic project has solubility results significantly higher than ours suggests that it may be difficult to improve this step by modifying expression protocols. However, it may be possible to improve the percentage further, either by choosing alternate expression systems (cell-free, insect cell or mammalian expression) or by investigating the refolding of those proteins that are expressed at high levels but are insoluble. Both approaches are expensive in terms of time and money. However, a refolding approach could be amenable if it were automated to enable the rapid identification of proteins that can be refolded easily from inclusion bodies. Factorial refolding screens have been shown to be useful for refolding proteins [21], but it remains difficult to assay for folding in a high-throughput format.

In summary, we used a parallel production approach for bacterial expression in medium throughput to yield 15 soluble mouse proteins from a total of 126 starting targets (12%), using both manual and automated approaches. The success rate was similar to that observed in much higher throughput parallel production pipelines for eukaryotic proteins, but further improvement to the overall success rate for our approach may be obtained by improving the efficiency of the initial cloning step. In addition, parallel production projects may benefit from a method that will rapidly, and in an automated manner, identify insoluble proteins that can be refolded easily.

Acknowledgements

This work was supported by a University of Queensland Postdoctoral Fellowship to N.C.P. and an Australian Research Council grant to J.L.M. and B.K.

References

1. Stevens, R.C. and Wilson, I.A. (2001) *Science* **293**(5529), 519–520.
2. Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A. et al. (2002) *Proc. Natl. Acad. Sci. USA* **99**(18), 11664–11669.
3. Canaves, J.M., Page, R., Wilson, I.A. and Stevens, R.C. (2004) *J. Mol. Biol.* **344**(4), 977–991.
4. Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T. et al. (2000) *Nat. Struct. Biol.* **7 Suppl**., 943–945.
5. Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V. and Cambillau, C. (2003) *Acc. Chem. Res.* **36**(3), 165–172.
6. Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. (2002) *Proc. Natl. Acad. Sci. USA* **99**(5), 2654–2659.
7. Chen, L., Oughtred, R., Berman, H.M. and Westbrook, J. (2004) *Bioinformatics* **20**(16), 2860–2862.
8. Yokoyama, S. (2003) *Curr. Opin. Chem. Biol.* **7**(1), 39–43.
9. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H. and Edwards, A.M. (2000) *Prog. Biophys. Mol. Biol.* **73**(5), 339–345.
10. Huang, R.Y., Boulton, S.J., Vidal, M., Almo, S.C., Bresnick, A.R. and Chance, M.R. (2003) *Biochem. Biophys. Res. Commun.* **307**(4), 928–934.
11. Heinemann, U., Bussow, K., Mueller, U. and Umbach, P. (2003) *Acc. Chem. Res.* **36**(3), 157–163.
12. Widmann, M. and Christen P. (2000) *J. Biol. Chem.* **275**(25), 18619–18622.
13. Tresaugues, L., Collinet, B., Minard, P., Henckes, G., Aufrere, R., Blondeau, K. et al. (2004) *J. Struct. Funct. Genomics* **5**(3), 195–204.
14. You, J., Cohen, R.E. and Pickart, C.M. (1999) *Biotechniques* **27**(5), 950–954.
15. Endo, Y. and Sawasaki, T. (2004) *J. Struct. Funct. Genomics* **5**(1–2), 45–57.
16. Holz, C., Hesse, O., Bolotina, N., Stahl, U. and Lang, C. (2002) *Protein Expr. Purif.* **25**(3), 372–378.
17. Albala, J.S., Franke, K., McConnell, I.R., Pak, K.L., Folta, P.A., Rubinfeld, B., Davies, A.H., Lennon, G.G. and Clark R. (2000) *J. Cell. Biochem.* **80**(2), 187–191.
18. Wells, C.A., Ravasi, T., Sultana, R., Yagi, K., Carninci, P., Bono, H., Faulkner, G., Okazaki, Y., Quackenbush, J., Hume, D.A. and Lyons, P.A. (2003) *Genome Res.* **13**(6B), 1360–1365.
19. Wells, C.A., Ravasi, T., Faulkner, G.J., Carninci, P., Okazaki, Y., Hayashizaki, Y., Sweet, M., Wainwright, B.J. and Hume, D.A. (2003) *BMC Immunol.* **4**(1), 5.
20. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S. et al. (2002) *Nature* **420**(6915), 563–573.
21. Chen, G.Q. and Gouaux, E. (1997) *Proc. Natl. Acad. Sci. USA* **94**(25), 13431–13436.
22. Moy, S., Dieckman, L., Schiffer, M., Maltsev, N., Yu, G.X. and Collart, F.R. (2004) *J. Struct. Funct. Genomics* **5**(1–2), 103–109.