Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A flexible and economical medium-throughput strategy for protein production and crystallization

Nicole Moreland,[a]* Rachael Ashton,[a] Heather M. Baker,[a] Ivan Ivanovic,[a] Sean Patterson,[b] Vickery L. Arcus,[a] Edward N. Baker[a,c] and J. Shaun Lott[a,c]

[a]School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand, [b]Beckman Coulter NZ, PO Box 109518, Newmarket, Auckland, New Zealand, and [c]Centre for Molecular Biodiscovery, University of Auckland, Private Bag 92-019, Auckland, New Zealand

Correspondence e-mail: n.moreland@auckland.ac.nz

Received 3 June 2005
Accepted 25 July 2005

Large-scale structural genomics centres rely heavily on robotics to ensure that maximum throughput is achieved. However, the size and cost of these approaches is out of the reach of most academic structural biology efforts. A major challenge for such groups is to adapt current high-throughput schemes to a reasonable scale with the resources available. A flexible medium-throughput approach has been developed that is suitable for typical academic research groups. Following nested PCR, targets are routinely cloned into two Gateway expression vectors (pDEST15 for an N-terminal GST tag and pDEST17 for an N-terminal His tag). Expression of soluble recombinant protein in *Escherichia coli* is rapidly assessed in 96-well format. An eight-probe sonicator is utilized and a six-buffer lysis screen was incorporated to enhance solubility. Robotics is reserved for crystallization, since this is the key bottleneck for crystallography. Screening proteins with a 480-condition protocol using a Cartesian nanolitre-dispensing robot has increased crystallization success markedly, with an overall success rate (structures solved out of proteins screened) of 19%. The methods are robust and economical – with the exception of the crystallization robot, investment in additional equipment has been minimal at US$9000. All protocols are designed for individuals so that graduate students and postdoctoral fellows gain expertise in every aspect of the structural pipeline, from cloning to crystallization.

1. Introduction

The relatively young field of structural genomics has developed as a direct result of the dramatic advances made in genome sequencing at the end of the 20th century. A small prokaryotic genome can now be sequenced in days, and this explosion of genomic information has fuelled the establishment of structural genomics programmes worldwide (Thornton, 2001). Structural genomics may be defined as the large-scale, systematic, determination of three-dimensional structures for all gene products in an organism. In reality, this is often refined to a representative group of proteins from an organism, chosen to address narrower goals such as metabolic pathways, protein families or novel folds (Todd *et al.*, 2005). Structural genomic approaches are tempered further by current technology and practical constraints – only a proportion of proteins can be expressed in a soluble form and, in turn, only a fraction of these will crystallize or be amenable to NMR analysis. For these reasons, structural genomics initiatives are somewhat targeted in nature and are undertaken by large consortia and academic groups alike, the key difference being that of scale (Vincentelli *et al.*, 2003).

Large structural genomics centres rely heavily on robotics to ensure that maximum throughput is achieved. With sufficient resources, gene cloning, protein-solubility screening, purification and crystallization can all be largely automated. However, the size and cost of these approaches is out of the reach of most academic structural biology laboratories. A major challenge for academic laboratories is to adapt current high-throughput schemes to a manageable scale with the resources available (Sulzenbacher *et al.*, 2002). With such a large array of robotic options available, academic groups need to carefully consider which, if any, part of their system will be automated. Automation is not only expensive but requires expertise in engineering, data processing and tracking, tasks which are not readily handled in traditional academic departments (Rupp, 2003). The needs of graduate students must also be considered: any academic 'throughput' approach must still ensure that graduates gain an understanding of structural biology. An overarching consideration is that of throughput *versus* output. Large consortia are likely to accept a high level of attrition, whereas hypothesis-driven research groups in academia may be less tolerant and would benefit more from a targeted approach with higher output.

Here, we present approaches to protein production and crystallization developed for a moderate-sized academic structural biology group. We are currently involved in two medium-scale structural genomics projects, with focused target sets from *Mycobacterium tuberculosis* (see http://www.tbgenomics.org/) and the mammalian pathogenic orf virus. Further projects encompass target sets from *Staphylococcus aureus*, *Streptococcus pyogenes* and the model plant species *Arabidopsis thaliania*, together with traditional single-target structural biology projects. Such a diverse target pool requires adaptable systems which accommodate the broad needs of each set of proteins. We utilize nested PCR and Gateway cloning strategies and have developed robotic free protocols to screen for protein solubility in a 96-well format. Robotics is reserved for crystallization, where the greatest benefits for output accrue. Protocols are designed for individuals, so that graduate students and postdoctoral fellows gain expertise in all aspects of structural biology, from cloning to crystallization.

2. Target selection, cloning and expression

In contrast to large-scale structural genomics programmes, where vast numbers of targets are selected by software with strict bioinformatic filters (see http://www.nigms.nih.gov/psi/), we select and process limited target sets driven by focused goals. Examples amongst the ~20 target sets currently being processed are a set of DNA-repair proteins from *M. tuberculosis* and a set of putative virulence proteins from *S. pyogenes*. This structural genomics approach has different needs with respect to technology and experimental protocol. Large-scale programmes process several hundreds of genes simultaneously by performing PCR and cloning in 96-well plates using robotics. In contrast, our medium-throughput (MT) cloning strategy has been designed for an individual to readily process eight ORFs (or constructs) using eight-tube strips and 24-well plates. As the targets are of particular functional interest, considerable effort is made to ensure targets move through the pipeline with the maximum chance of success.

Our MT approach utilizes *Escherichia coli* as a protein-expression host – fast cultivation, easy handling, high protein yields and a large range of plasmid vectors are the major advantages of using the *E. coli* system (reviewed by Makrides, 1996). While the insect-cell/baculovirus expression system is available in-house, these methods have not been incorporated into our MT strategy to date. This system is instead reserved for more challenging targets where post-translation modification is required to obtain fully active recombinant proteins or as a salvage strategy for high-priority targets that have proved intractable in *E. coli*.

2.1. Gateway cloning

The limitations of restriction-enzyme-based cloning, in particular the restriction-site analysis required for primer design and the reliance on digestion of the PCR-amplified target gene before cloning, have resulted in our adoption of the Gateway (Invitrogen) *in vitro* recombination cloning method (Hartley *et al.*, 2000). The decision to utilize the Gateway system rather than another recombinational cloning method such as BD In-Fusion Cloning (Clontech) was based on the availability of clone libraries. Evaluations of the Gateway and In-Fusion systems have shown them to be largely equivalent in terms of efficiency (Marsischky & LaBaer, 2004), but with the majority of our collaborators using the Gateway system and the availability of Gateway custom clone sets from The Institute for Genomic Research, the ability to readily share and obtain Gateway clones was seen as a key benefit.

Gateway cloning relies on the presence of *att* recombination sequences on either side of the target gene DNA to transfer target sequences into expression vectors. A consequence of including the 25 bp *att*B recombination sequence is the addition of an eight-amino-acid linker between the target protein and the N-terminal tag. While some groups have found the additional *att* amino acids affect solubility (Luan *et al.*, 2004) and crystallization (Schormann *et al.*, 2004; Chance *et al.*, 2002), others have reported good success rates for solubility (Vincentelli *et al.*, 2003; Cowieson *et al.*, 2005) and crystallization (Sulzenbacher *et al.*, 2002) when the additional amino acids are present. We have chosen to include an rTEV protease-cleavage site allowing the expressed protein to be released from the affinity tag and *att* amino acids prior to crystallization if necessary. We routinely clone targets into two expression vectors encoding hexahistidine (His_6) and glutathione *S*-transferase (GST) purification tags to maximize the likelihood of solubility.

We use a two-stage nested PCR strategy to flank the target sequences with the *att*B recombination sequences and to encode an rTEV cleavage site in between the *att*B sequence and the original ORF translational start (Fig. 1). Gene-specific

primers used in the first round of PCR contain a 12 base-pair overlap with the generic primers used in the second round. PCR products are recombined (BP reaction) with pDONR221 (Invitrogen) to generate Entry clones. Subsequent LR reactions routinely utilize pDEST17 and pDEST15 N-terminal His_6-tag and GST-tag expression vectors, respectively (Invitrogen). The relative ease with which an Entry clone can be recombined with any number of destination vectors *via* the LR reaction adds flexibility in downstream functional analysis. We often utilize pDEST14 for native expression and the ProQuest yeast two-hybrid bait and prey vectors pDEST32 and pDEST22 (Invitrogen) to investigate protein–protein interactions.

2.2. Cloning and expression workflow

As shown in Fig. 2, all reactions are performed in eight-tube strips and multi-well plates. The vast majority of targets amplify on the first attempt using nested PCR across an annealing-temperature gradient (Eppendorf Mastercycler Gradient) in eight-tube strips. On the rare occasion that the target sequence is not correctly amplified, reaction conditions are customized in individual PCR tubes, so that >95% of targets are subject to Gateway cloning without a major investment of effort. The Gateway cloning method has proven to be robust and cost-effective in our laboratory. An aliquot of nested PCR product can be used directly in BP cloning reactions without quantification and the BP and LR reactions are very stable, remaining viable after >48 h on the bench. We propagate the cloning vectors ourselves using *E. coli* DB3.1 cells (Invitrogen) and routinely use one-quarter of the manufacturer's recommended reaction scale, thus reducing the cost of each cloning reaction to approximately US$4.80.

While large-scale programmes omit screening more than one colony to select for positive clones, we have found it necessary to retain this step in order to ensure a near to 100% cloning success rate. In our experience, the selection of three colonies from each BP or LR reaction plates is sufficient to identify a positive clone. DNA is extracted from colonies by Miniprep (QIAprep Spin Miniprep kit) and clones are screened for insert with *Bsr*GI digests (Novagen). We sequence all of our Entry clones to ensure that targets are cloned in-frame and without mutation.

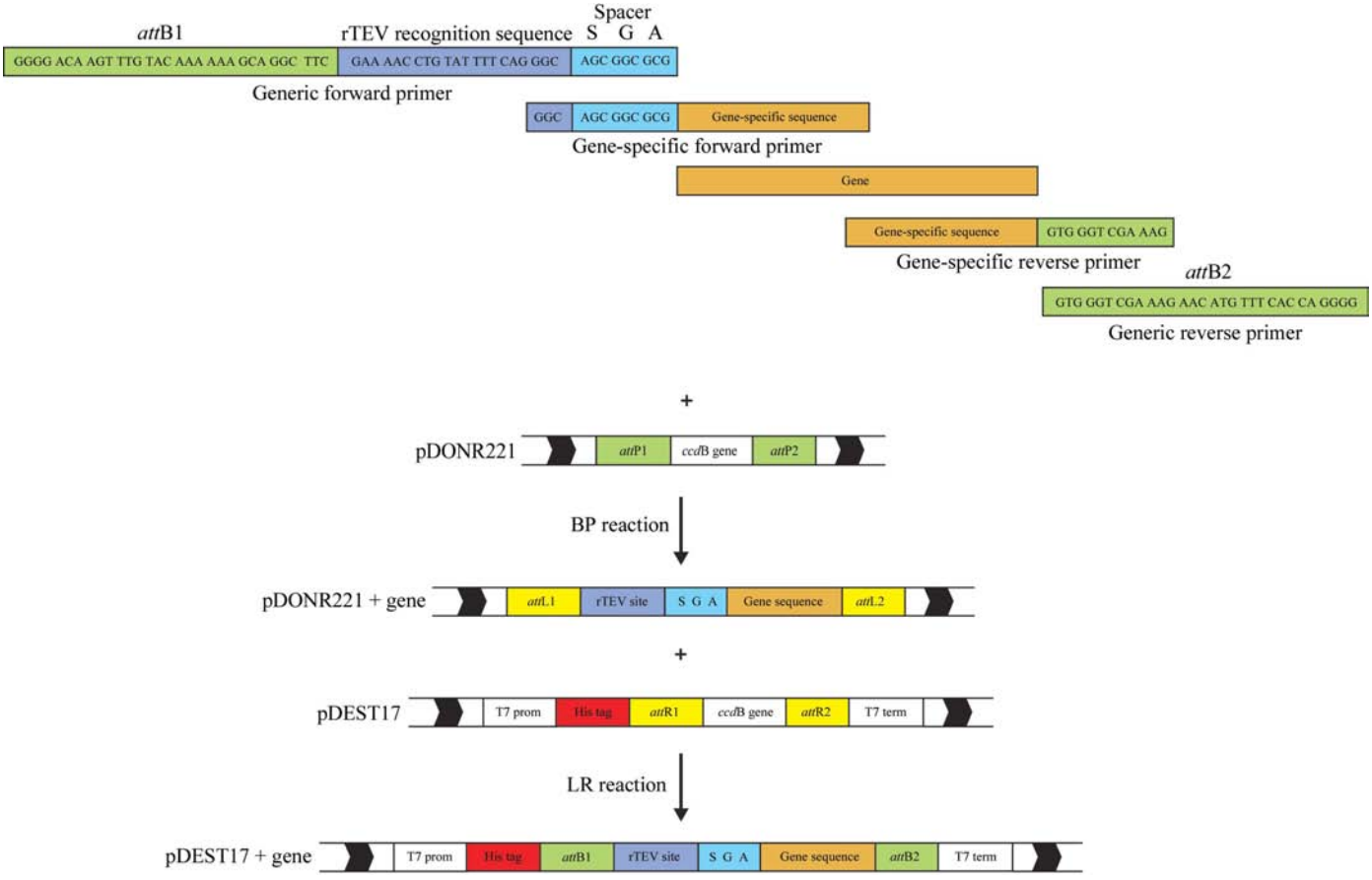


Figure 1
Nested PCR and recombinant cloning using the Gateway technology. Nested PCR is performed so that the resulting constructs comprise the target ORF flanked by the *att*B recombination sequences. Gene-specific primers used in the first round of PCR contain a 12-base-pair overlap with the generic primers used in the second round. The generic forward primers include 21 nucleotides that encode the rTEV peptide-recognition site followed by three spacer amino acids (SGA) for optimal rTEV cleavage of the recombinant protein. The PCR product is subcloned into pDONR221 (BP reaction) by incubating with BP Clonase for $\geq$1 h at 298 K. Following selection of a positive Entry clone (detailed in Fig. 2), the ORF is transferred into at least two destination vectors (LR reaction). Entry clones are incubated with pDEST17 (or pDEST15) and LR Clonase for $\geq$1 h at 298 K.

3. Protein expression and solubility screening

A significant hurdle in expressing proteins for structural biology is that of protein solubility in the heterologous host. A disadvantage of the *E. coli* expression system is that recombinant proteins are often found to be insoluble when the host cells are lysed. This observed insolubility can arise from a number of factors. In many cases, proteins which are unable to be folded in the cell accumulate as insoluble aggregates in inclusion bodies (Mukhopadhyay, 1997). However, as individual proteins each have a characteristic solubility and behave according to their own chemistry, altering the buffer used to lyse the cells can affect solubility. Screening varying concentrations of salt, pH ranges and a variety of additives can result in enhanced yields of active soluble protein (Lindwall *et al.*, 2000).

We have developed a quick, robust and robot-free means of screening for soluble recombinant protein after overexpression in a 96-well format. Rather than chemically lyse cells in one standard buffer, we utilize an eight-probe sonicator and have incorporated a lysis-buffer screen to maximize output. Investment in equipment has been minimal at US$9000 for multi-channel pipettes, bench-top Eppendorf Thermomixers and an eight-probe sonicator attachment. Our protocol enables an individual to rapidly assess 16 recombinant proteins (eight ORFs cloned into pDEST15 and pDEST17) for solubility in six different lysis buffers. The most commonly used *E. coli* strains are BL21(DE3) and the rare-codon variant Rosetta(DE3) (Novagen). In a target-dependent manner, a researcher may choose to clone into BL21(DE3)pLysS for tighter control of protein expression or into BL21(DE3)pGroELS for the coexpression of the *E. coli* chaperones GroEL and GroES. The 96-well format can be adapted to allow several *E. coli* host strains to be tested in parallel.

3.1. 96-Well expression and solubility

We utilize the auto-induction protocols developed by Studier (2005) in our solubility screen. Each of the 16 expression clones is grown overnight in PA-0.5G, a minimal medium that does not induce protein expression. These PA-0.5G cultures are used to seed a 96-well culture plate with six duplicate 1 ml cultures of each of the 16 clones as shown (Fig. 3). Expression and solubility in ZYP-5052 (an auto-induction medium) is assessed at three different incubation temperatures (310, 297 and 291 K), as decreasing the temperature of expression has been demonstrated to improve the solubility of many expressed proteins (reviewed by Sørensen & Mortensen, 2005). After incubation, cells are harvested by centrifugation and pellets are resuspended in 300 µl of the initial lysis buffer screen as detailed in Fig. 3. An eight-probe sonicator (Sonics Vibra Cell) is used for lysis and samples from soluble and pellet fractions are visualized by SDS–PAGE.

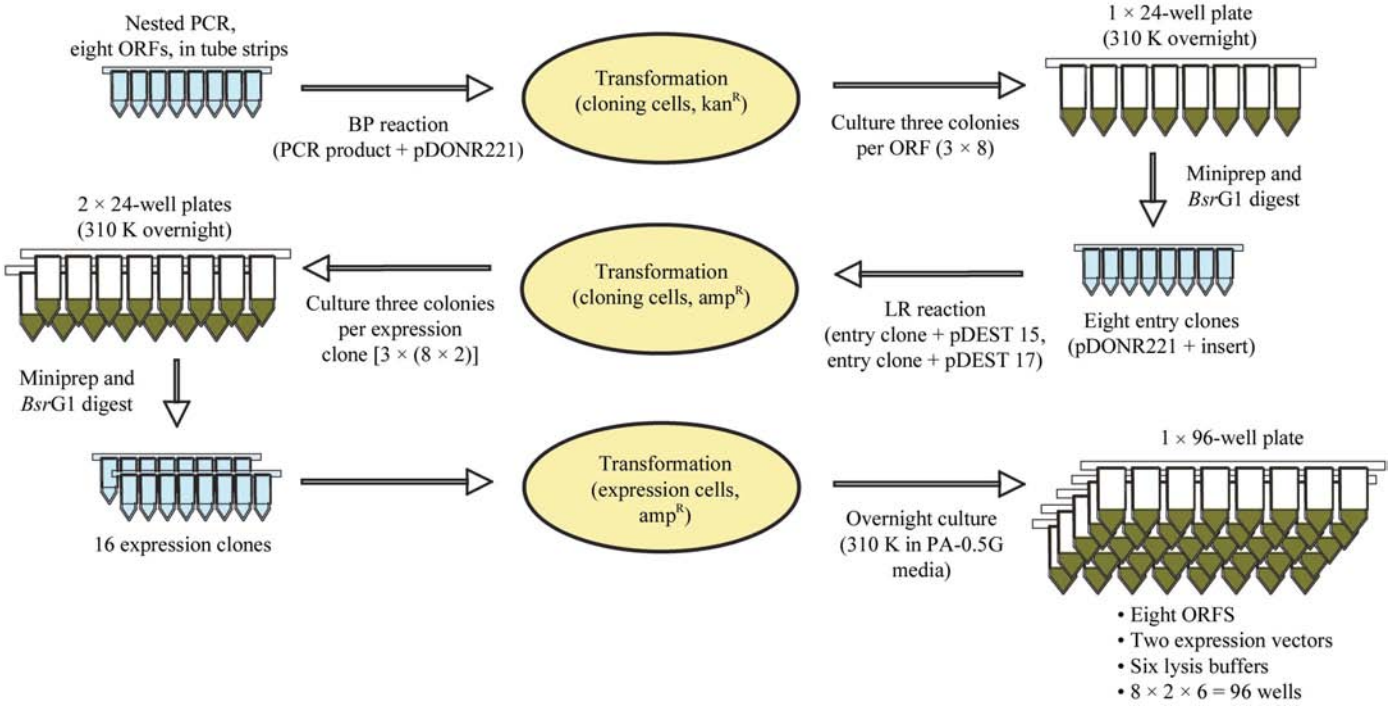


Figure 2
Medium-throughput workflow for cloning and protein expression. *E. coli* plasmid-maintenance strains (Top10 or DH5α) are used for transformations following BP and LR reactions. Positive recombinant clones are selected by plating on kanamycin (Entry clones) or ampicillin plates (Expression clones). To ensure a 100% cloning success rate, three colonies are screened from each plate. Colonies are grown overnight at 310 K in Luria–Bertani broth and plasmids are extracted using the QIAprep Spin Miniprep kit. Extracted DNA is digested with *Bsr*GI for 1 h at 310 K to identify clones containing the desired ORF. *E. coli* expression strains [BL21(DE3) or Rosetta(DE3)] are transformed with 16 expression vectors for solubility screening in 96-well format (detailed in Fig. 3). All reactions are performed in eight-tube strips and multi-well plates.

3.1.1. Validation of the 96-well solubility screen. In order to ensure that scaling down to 1 ml auto-induction cultures in 96-well culture plates did not alter solubility, the protocols were validated using 12 targets with known solubility characteristics. These targets had been cloned using either Gateway or traditional restriction-enzyme methods. All had been extensively screened for solubility following growth in individual 13 ml culture tubes with Luria–Bertani broth. The 96-well methods and initial lysis buffer screen accurately reproduced solubility for the *M. tuberculosis* positive controls Rv2874, Rv1636c, Rv0550c and Rv1680 and the *M. smegmatis* positive control PhnD, demonstrating the reliability of the protocols. Similarly, all of the targets previously identified as insoluble (Rv0407, Rv0549c, Rv0555, Rv2829c, Rv2830c, Rv2546c and Rv2545c) remained insoluble using the 96-well methods.

The cloning and expression methods described are now routinely used by researchers in our laboratory. The proportion of soluble proteins identified per 96-well plate screen is dependent on the target source. On average, researchers working on *M. tuberculosis* identify three soluble targets in each set of eight using the initial lysis buffer screen (37.5%). This is in line with solubility success rates observed by other member laboratories of the TB Structural Genomics Consortium (see http://www.tbgenomics.org/). Summary statistics for the first 52 targets processed using the 96-well protocols are shown in Table 1. These encompass target sets from three different species. 47 (90%) of these targets were successfully cloned into Gateway expression vectors following nested PCR and generation of a sequence-verified Entry clone. Of these, 41 (79%) expressed and 23 (44%) are soluble and suitable for large-scale purification.

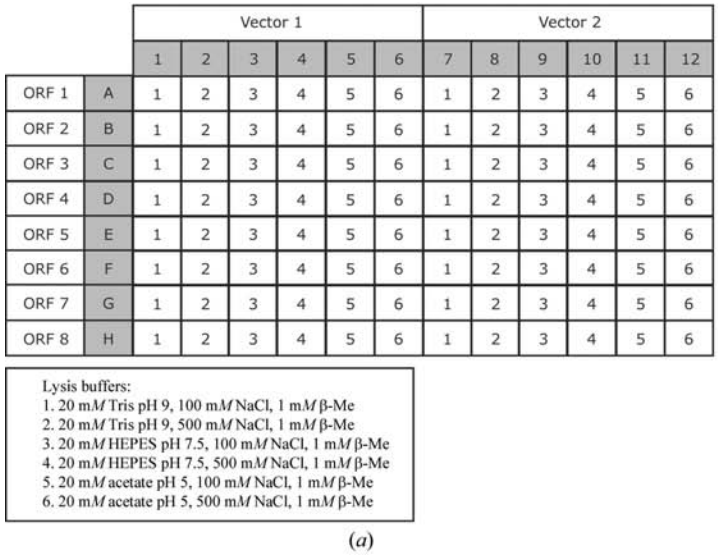
Table 1
Summary statistics for cloning, expression and solubility screening.

The targets encompass ORFs from three different species.

	Number	Targets selected (%)
Targets	52	100
Nested PCR	51	98
Gateway cloned	47	90
Expression	41	79
Soluble	23	44

3.2. Solubility salvage pathways

A number of strategies are used in our laboratory to salvage targets that are insoluble in the initial lysis-buffer screen. Lysis buffers used in subsequent 96-well solubility trials can be varied depending on prior knowledge or predicted physical characteristics of individual targets such as isoelectric point and cofactor requirements. Buffers may be adjusted around a particular pH or salt concentration and additives such as detergents or cosolvents may be included. We have had recent success using osmotic shock as a rescue strategy for several *M. tuberculosis* proteins. *E. coli* grown in the presence of high salt (0.5 *M* NaCl) and 1 m*M* betaine produces osmolytes which may activate chaperone-mediated refolding of the target protein, thus aiding solubility (Das *et al.*, 2005). If neither of these rapid approaches is successful, then a researcher may consider a protein-refolding screen or the expression of individual domains of a particular ORF. While refolding strategies are often automated and used heavily by large-scale structural genomics consortia (Vincentelli *et al.*, 2004; Tresaugues *et al.*, 2004), in our laboratory they are used on a case-by-case basis. Currently used refolding screens may be adapted to higher-throughput 96-well methods down-



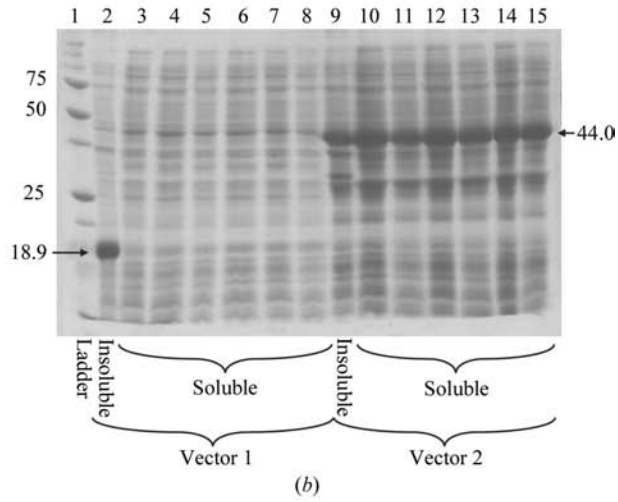


Figure 3
96-Well solubility screen. (*a*) Following centrifugation (3000 rev min^{-1}, 30 min, 277 K) cell pellets from 1 ml of ZYP-5052 culture are resuspended in 300 µl of initial lysis-buffer screen as shown in the grid. Samples are lysed using an eight-probe sonicator in four 15 s bursts. Soluble and insoluble fractions are separated by a final centrifugation step (4500 rev min^{-1}, 60 min, 277 K). (*b*) Samples from soluble and pellet fractions are visualized on 15-lane SDS–PAGE. Each row from the 96-well plate is loaded on one gel. In this example, the *M. tuberculosis* ORF Rv1160 was cloned into pDEST17 (Vector 1) and pDEST15 (Vector 2). Insoluble and soluble fractions (15 µl) were loaded onto a 15% SDS–PAGE gel. Molecular weights are labelled in kDa.

Table 2
Crystallization throughput statistics for the first 12 month period that robotics were fully functional in the laboratory.

For the purposes of the analysis a target is defined as a unique ORF or domain, so that different constructs of a particular ORF are separate targets. Target-derivative statistics are shown in parentheses. These include selenomethionine-labelled proteins and targets that were screened with different substrates.

	Number	Proteins screened (%)
Targets screened (including derivatives)	69 (79)	100
Targets crystallized (including derivatives)	44 (51)	64
Crystals with X-ray data (including derivatives)	20 (25)	29
Structures	13	19

stream, but the expansion in effort and resources required negates this at present.

4. Protein purification

Large-scale production of soluble proteins does not readily lend itself to parallel processing (Lesley, 2001). Centrifugation and lysis of large cultures is not easily integrated into high-throughput approaches and with limited resources automation is difficult to achieve. We scale up the production of soluble targets identified in the 96-well screen individually. Cultures (500 ml) are grown in 2 l baffled flasks in autoinduction media, lysed using a cell disrupter or sonicator and purified by affinity chromatography (using the His_6 or GST purification tags). If N-terminal tag removal is desired, samples are incubated with rTEV at 277 K overnight, followed by a further Ni–NTA affinity purification step to remove His-tagged rTEV. Gel filtration is routinely used as the final purification step and together with dynamic light scattering (DynaPro, Protein Solutions) serves as a characterization step prior to crystallization. Mass spectrometry (Voyager-DE PRO Biospectrometry), circular-dichroism spectrometry (π^* – 180, Applied Photophysics) and one-dimensional NMR spectrometry (Bruker 400 and 600 MHz systems) are used for further characterization of targets.

5. Crystallization

Success in protein crystallization has traditionally been regarded as the major bottleneck in structural analysis by X-ray crystallography. Although many attempts have been made to understand crystallization, little progress has been made in the rational prediction of the conditions under which a particular protein will crystallize (Mittl & Grütter, 2001). The labour-intensive and inherently empirical nature of crystal trials, the relative ease with which crystallization experiments can be automated and reports of increased crystallization success rates using nanolitre technology (Sulzenbacher *et al.*, 2002) spurred our investment in a Cartesian HoneyBee nanolitre-dispensing robot for crystallization (Genome Solutions). This is the only part of our MT system to be automated – the largest investment in equipment made to circumvent our major bottleneck.

5.1. Crystal screens and robotics

The opportunities for data mining from high-throughput structural genomics projects (Page *et al.*, 2003), coupled with the development of new crystallization screens, led us to design a 480-condition protocol (five 96-well plates) for initial screening of target proteins. Using 100 + 100 nl drops enables a comprehensive initial screen to be performed using less than 60 µl of protein solution or approximately 0.6 mg of protein at a concentration of 10 mg ml^{-1}. The protocol has been produced by combining our most successful in-house and commercially available screens, arranged in order so that the most successful screens are used first if there is insufficient protein for all five plates. The composite screen, after removal of duplicates, includes the Top67 screen of Page *et al.* (2003), PEG/Ion and sparse-matrix Hampton Crystal Screens I and II (Hampton Research), PEG/pH and ammonium sulfate screens based on orthogonal arrays (Kingston *et al.*, 1994), the Stura Footprint screen (Stura *et al.*, 1992), an MPD screen expanded from the Hampton MPD screen and the Precipitant Synergy (Majeed *et al.*, 2003) and Clear Strategy (Brzozowski & Walton, 2001) screens.

The Cartesian nanolitre-dispensing robot is complemented by a MultiPROBE II HT/EX liquid-handling robot (Perkin–Elmer) for preparing and dispensing screens. Custom-built reagent racks on the MultiPROBE hold 160 stock solutions which are used to automatically assemble the initial screens in 96-deep-well plates. Each 96-condition screen can be assembled from stock solutions in approximately 3 h without user intervention. All screen assembly and dispensing is carried out in a temperature-controlled environment (291 K) and volatile components are added last in order to minimize evaporation variability. Crystallization precipitant solutions (50 µl per well) are then transferred from the deep-well plates into 96-well Intelli-Plates (Hampton Research). Intelli-Plates were chosen for their superior optical properties and their ability to yield reproducible drop shape. The Intelli-Plate is manually transferred to the Cartesian robot, where 100 nl drops of protein and well solutions are dispensed for sitting-drop vapour-diffusion crystallization (Fig. 4). To minimize evaporation, crystallization experiments are set up under controlled temperature conditions (291 K) and relative humidity (85%). While some crystals from the initial 480-condition robot screen are of suitable size and quality for diffraction, in the majority of cases optimization is required. We use both robotics with 100 nl drops and larger volume drops set up manually in this optimization step. The same stock solutions that are used to construct the initial 480-condition robot screens are used to maximize reproducibility.

5.2. Crystallization throughput

The robotic crystallization facility has been fully operational within the laboratory for a year. During this period, 31 users have screened 69 proteins across our diverse target pool (Table 2). 44 (64%) of these proteins produced crystals using the 480-condition protocol, examples of which are shown in

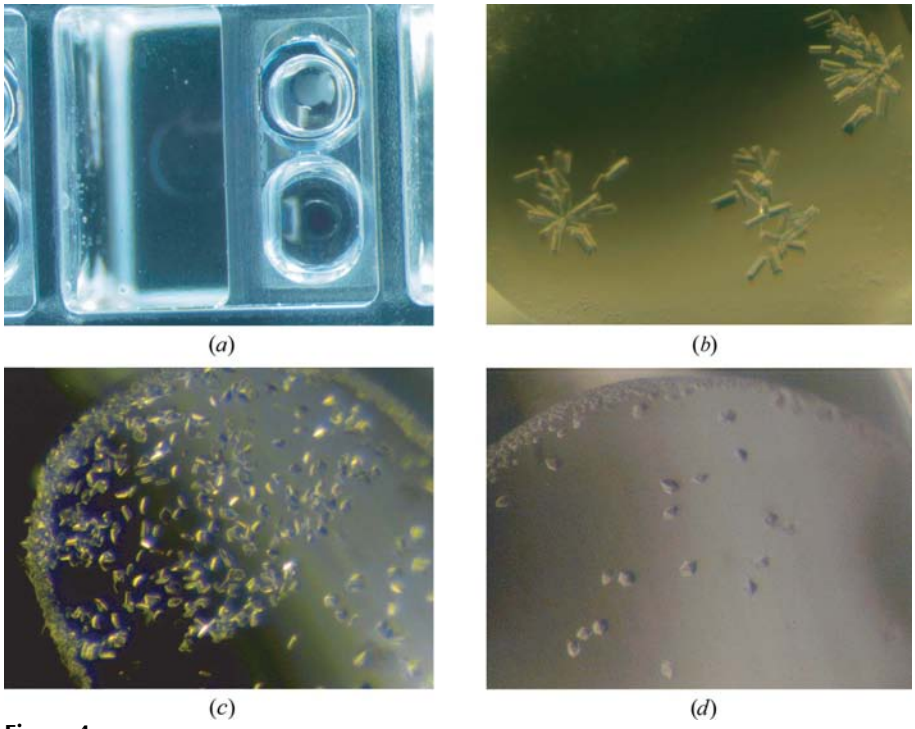


Figure 4
Crystallization with 100 nl drops. (*a*) One of the 96 wells in an Intelli-Plate. The main well on the left contains the well solution and the top right shelf contains the crystallization drop. (*b*) Rv3214 crystals in the Top67 screen. (*c*) PAE 2754 crystals in the Top67 screen. (*d*) Rv1284 in the PEG screen.

Fig. 4. Optimization of these initial conditions resulted in diffraction-quality crystals and X-ray data for 20 of the 69 (29%) using either the home source [MAR345 image-plate detector (MAR Research) with a Rigaku RU-300 X-ray generator] or synchrotron-radiation sources (the Advanced Light Source at Lawrence Berkeley National Laboratory and the Stanford Synchrotron Radiation Laboratory). Structures have been solved for 13 of the 69 proteins screened (19%) in the 12-month period and this overall success rate is in line with that observed in large-scale consortia (O'Toole *et al.*, 2004). The success rate for obtaining initial crystallization conditions has substantially improved with robotics and 100 nl drops (64% compared with ∼45% for previous manual methods) and the major emphasis is now on optimization to produce diffraction-quality crystals.

6. Summary and conclusions

As the power and speed of structure determination by X-ray crystallography increases, the bottlenecks in obtaining protein structures are increasingly found at the protein-production and crystallization stages. Efficient methods to identify soluble targets and move these targets through crystallization trials are essential tools for any structural biologist. The challenge for smaller groups, and especially for those operating in an academic setting, is to identify approaches and technologies that are cost-effective in meeting the twin goals of research output and graduate-student development.

We have developed a simple medium-throughput approach which increases success and is suitable for a typical academic research group. Following nested PCR, targets are routinely cloned into two Gateway expression vectors that enable affinity-tagged protein production (pDEST15 for an N-terminal GST tag and pDEST17 for an N-terminal His tag; Invitrogen). Expression of soluble recombinant protein in *E. coli* is rapidly assessed in 96-well format. We utilize an eight-probe sonicator and have incorporated a six-buffer lysis screen to enhance solubility. The only investment in robotics to date is for crystallization, since this is the key bottleneck for crystallography. Screening target proteins against a 480-condition protocol using a Cartesian nanolitre-dispensing robot has increased crystallization success markedly, as well as saving intensive manual effort.

The strategies presented here are robust and economical – with the exception of the crystallization robot, investment in additional equipment has been minimal at US$9000. The resulting system is flexible and could be adapted to suit most academic structural biology environments.

We thank Nathan Cowieson, Pawal Listwan and Jenny Martin (Institute for Molecular Bioscience, University of Queensland) for advice with Gateway cloning. We thank Jodie Johnston for providing controls for the solubility screen and Harriet Watkins, Richard Bunker and Jessica Chaston for providing crystal pictures. Brent Segelke and Tim Lekin (Lawrence Livermore National Laboratory, California) are acknowledged for providing invaluable advice in the establishment of the robotic crystallization facility. This work was supported by the New Economy Research Fund of New Zealand, the Health Research Council of New Zealand and Centres of Research Excellence funding to the Centre for Molecular Biodiscovery.

References

Brzozowski, A. M. & Walton, J. (2001). *J. Appl. Cryst.* **34**, 97–101.
Chance, M. R., Bresnick, A. R., Burley, S. K., Jiang, J.-S., Lima, C. D., Sali, A., Almo, S. C., Bonanno, J. B., Buglino, J. A., Boulton, S., Chen, H., Eswar, N., He, G., Huang, R., Ilyin, V., McMahan, L., Pieper, U., Ray, S., Vidal, M. & Wang, L. K. (2002). *Protein Sci.* **11**, 723–738.
Cowieson, N. P., Listwan, P., Kurz, M., Aagaard, A., Ravasi, T., Wells, C., Huber, T., Hume, D. A, Kobe, B. & Martin, J. L. (2005). *J. Struct. Funct. Genomics*, **6**, 13–20.
Das, D., Oganesyan, N., Yokata, H., Pufan, R., Kim, R. & Kim, S.-H. (2005). *Proteins*, **58**, 504–508.

Hartley, J. L., Temple, G. F. & Brasch, M. A. (2000). *Genome Res.* **10**, 1788–1795.
Kingston, R. L., Baker, H. M. & Baker, E. N. (1994). *Acta Cryst.* D**50**, 429–440.
Lesley, S. A. (2001). *Protein Expr. Purif.* **22**, 159–164.
Lindwall, G., Chau, M., Gardner, S. & Kohlstaedt, L. (2000). *Protein Eng.* **13**, 67–71.
Luan, C. H., Qiu, S., Finley, J. B., Carson, M., Gray, R. J., Huang, W., Johnson, D., Tsao, J., Reboul, J., Vaglio, P., Hill, D. E., Vidal, M., DeLucas, L. J. & Luo, M. (2004). *Genome Res.* **14**, 2102–2110.
Majeed, S., Ofek, G., Belachew, A., Huang, C.-C., Zhou, T. & Kwong, P. D. (2003). *Structure*, **11**, 1061–1070.
Makrides, S. C. (1996). *Microbiol. Rev.* **60**, 512–538.
Marsischky, G. & LaBaer, J. (2004). *Genome Res.* **14**, 2020–2028.
Mittl, P. & Grütter, M. (2001). *Curr. Opin. Chem. Biol.* **5**, 402–408.
Mukhopadhyay, A. (1997). *Adv. Biochem. Eng. Biotech.* **56**, 61–109.
O'Toole, N., Grabowski, M., Otwinowski, Z., Minor, W. & Cygler, M. (2004). *Proteins*, **56**, 201–210.
Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003). *Acta Cryst.* D**59**, 1028–1037.
Rupp, B. (2003). *Acc. Chem. Res.* **36**, 173–181.
Schormann, N., Symersky, J. & Luo, M. (2004). *Acta Cryst.* D**60**, 1840–1845.
Sørensen, H. P. & Mortensen, K. K. (2005). *Microb. Cell Fact.* **4**, 1.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.
Sulzenbacher, G. *et al.* (2002). *Acta Cryst.* D**58**, 2109–2115.
Thornton, J. (2001). *Trends Biochem. Sci.* **26**, 88–89.
Tresaugues, L., Collinet, B., Minard, P., Henckes, G., Aufrere, R., Blondeau, K., Liger, D., Zhou, C. Z., Janin, J., Van Tilbeurgh, H. & Quevillon-Cheruel, S. (2004). *J. Struct. Funct. Genomics*, **5**, 195–204.
Todd, A. E., Marsden, R. L., Thornton, J. M. & Orengo, C. A. (2005). *J. Mol. Biol.* **348**, 1235–1260.
Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V. & Cambillau, C. (2003). *Acc. Chem. Res.* **36**, 165–172.
Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. & Bignon, C. (2004). *Protein Sci.* **13**, 2782–2792.