

## Chapter 4.4. High-throughput X-ray crystallography

K. H. CHOI

### 4.4.1. Introduction

Structural-genomics projects have contributed to major developments in automation, miniaturization and process integration in X-ray crystallography. In the structural-genomics approach, multiple open reading frames from a genome are separately cloned and expressed without prior knowledge of the structure or function of the encoded proteins. Essential steps in a typical project include: (1) bioinformatics analysis of genome sequences for potential targets, (2) gene amplification by PCR (polymerase chain reaction) and subcloning into an appropriate expression vector, (3) protein expression and purification, (4) protein crystallization, (5) X-ray data collection, and (6) data processing and structure determination. Structural-genomics projects typically operate under a 'lowest hanging fruit' philosophy, pursuing structural targets that prove to be the most amenable to crystallization. However, for individual investigators working on specific problems, it is often necessary to focus on a particular protein. Many of the protocols developed in high-throughput (HT) structural genomics can also be adapted for parallel production of multiple constructs of a single protein target which is difficult to crystallize. This chapter summarizes recent developments in HT X-ray crystallography approaches and their application to parallel production of a single protein (Gräslund, Nordlund *et al.*, 2008; Manjasetty *et al.*, 2008; Sharff & Jhoti, 2003).

### 4.4.2. Design of multiple constructs: bioinformatics analysis of genome sequences

Often, the two primary bottlenecks in crystal structure determination are obtaining a soluble protein and the initial crystallization. If initial crystallization trials do not lead to successful crystallization, diffraction-quality crystals or structure determination, modification of the protein itself is often more successful in producing useful crystals than exhaustive screening/optimization of protein crystallization conditions. In the HT approach, generation of stable domains having variable N or C termini can be performed in parallel, vastly reducing time and resource requirements. The clones can be screened systematically for soluble, active proteins amenable to crystallization. This approach is particularly suited to problems in which one has no estimate, *a priori*, of which constructs are likely to be most successful at the expression, purification or crystallization steps.

It may be difficult to predict which modifications will influence protein solubility and the protein's ability to crystallize. A common and often useful strategy is to crystallize a similar protein from an alternative biological source. If an alternative source is either unavailable or equally problematic, gene fusion with large affinity tags can be used for increasing protein expression and solubility. Several fusion proteins, including thioredoxin, maltose binding protein, glutathione S-transferase, intein, SUMO (small ubiquitin-like modifier) protein and calmodulin-binding protein, have been used to generate soluble proteins. However, the fused tag may prevent crystallization because of conformational heterogeneity resulting from a flexible

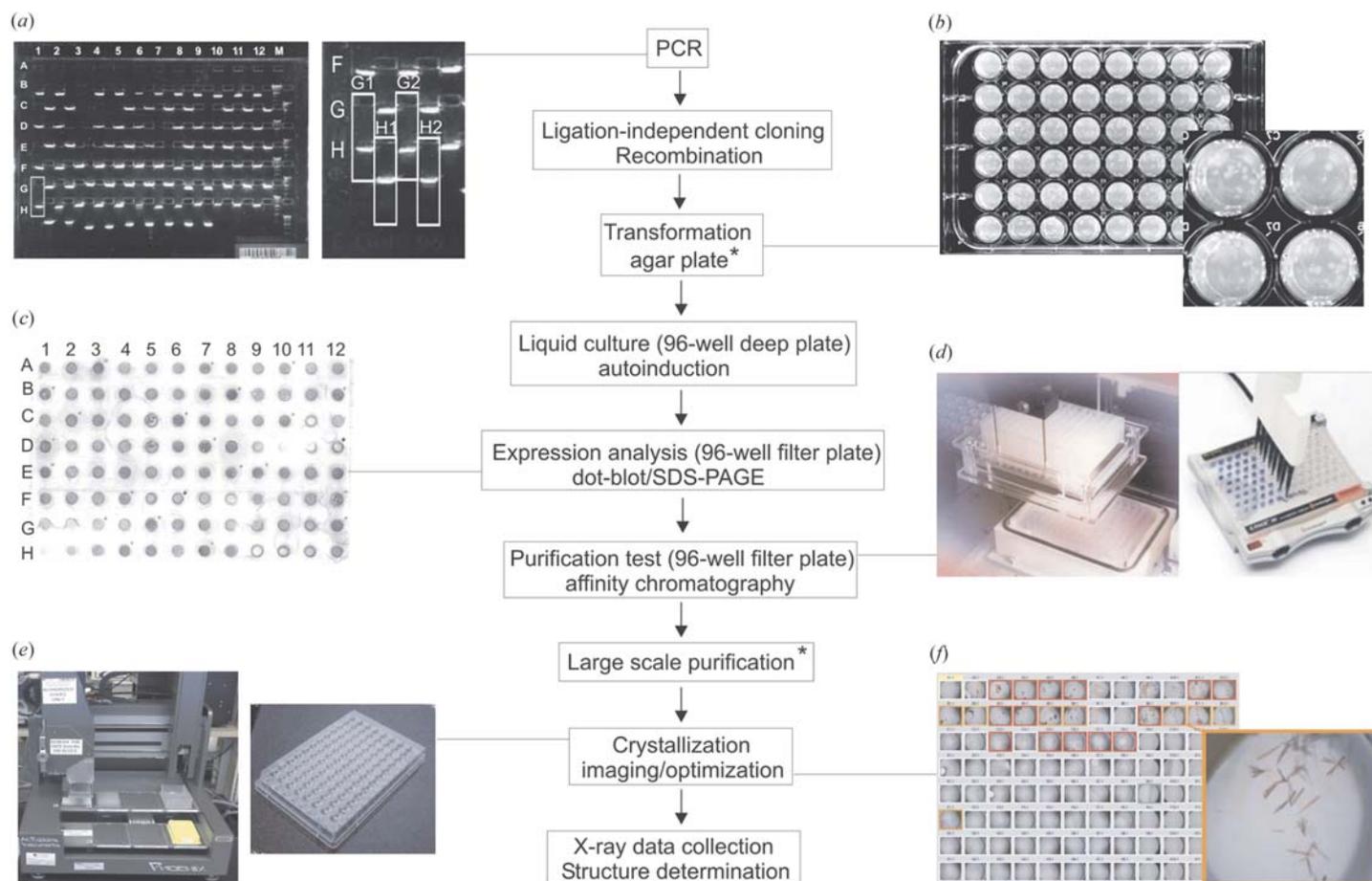
linker; hence, the tag must often be removed in an additional purification step. Protein engineering by amino-acid substitutions and chemical modifications (such as methylation of exposed lysines) have been shown to improve crystallization of some proteins (Rayment, 1997; Walter *et al.*, 2006; Wingren *et al.*, 2003). Large, flexible solvent-exposed residues (*e.g.* Lys or Glu) on the surface of proteins are substituted with smaller residues (*e.g.* Ala) to facilitate the formation of intermolecular contacts which stabilize the crystal lattice (the surface-entropy reduction approach; Cooper *et al.*, 2007; Derewenda, 2004). Random mutagenesis following selection of a desired phenotype (*e.g.* folding ability using a GFP (green fluorescent protein) reporter or solubility assays) has been used to produce soluble proteins from insoluble wild-type proteins (the directed-evolution approach; Pédelacq *et al.*, 2002). Since symmetric molecules such as homodimers may crystallize more readily than monomeric protein, new Cys residues can be introduced in a monomeric protein to induce homodimer formation *via* intermolecular disulfide bond formation (the synthetic symmetrization approach; Banatao *et al.*, 2006).

Bioinformatics tools, including multiple sequence alignment and sequence motif searches, as well as prediction of secondary structures, domain boundaries, membrane spanning and disordered regions, can be used to aid the rational design of constructs. General considerations in designing truncation mutations are to avoid truncations in the middle of predicted secondary structural elements, to avoid hydrophobic residues at the termini, and to eliminate membrane-spanning regions in the construct design. In the case of multi-domain proteins, truncation of the protein to smaller functional domains can be effective. Exact locations of the beginning and ending of a domain are still difficult to predict even with domain search programs, and thus biochemical approaches such as limited proteolysis can aid the determination of the domain boundaries. The optimal step size for truncation mutations cannot be predicted, but protein constructs varying by approximately five residues in length often show significant differences in solubility and crystallization behaviour (Choi *et al.*, 2004; Gräslund, Sagemark *et al.*, 2008).

### 4.4.3. Cloning

Development of robotics that utilize a 96-well format has changed traditional sequential cloning of individual proteins such that parallel HT cloning and protein preparation are now possible. The conventional steps in cloning are PCR amplification, restriction enzyme digestion, ligation, transformation, selection of transformers and protein expression tests. Many procedures, including amplification of target genes, cloning and screening for expression, can be achieved in parallel (96 samples at a time) using either a single liquid-handling robot or multi-channel pipettors. In 96-well parallel cloning, reaction steps for individual wells cannot be optimized, and thus care should be taken to synchronize all reactions and to minimize the number of steps to increase efficiency. For this reason, certain cloning strategies are more popular for HT cloning. Recombinant protein

#### 4.4. HIGH-THROUGHPUT X-RAY CRYSTALLOGRAPHY



**Figure 4.4.3.1**

Diagram of HT cloning, expression and purification. All steps can be performed in a 96-well format except the step labelled with an asterisk (\*). (a) PCR amplification. PCR products are loaded onto an E-gel (1% agarose) in a 96-well format with a molecular marker (lane 'M') and visualized with ethidium bromide. A close-up view of the gel outlining individual wells is also shown. (b) Transformation. Transformed cells were plated in 48-well agar plates. (c) Protein expression was tested using a western dot blot. H1 and H7 are negative and positive controls, respectively. (d) Small-scale purification using affinity resin. A 96-well filter plate in small-scale protein purification steps (left) and a 96-well pre-cast protein gel electrophoresis system (right) are shown. (e) Crystallization robot for setting up 96-well crystallization plates (right). (f) Crystal imaging and scoring system. A close-up view of an image of well B10 is shown on the right.

expression in *E. coli* is the most common, and will be discussed here.

##### 4.4.3.1. Ligation-independent and recombination cloning

Ligation-independent cloning (LIC) strategies remove the need for restriction enzyme digestion and ligation of PCR products, and are thus ideal for use in an HT cloning procedure. In LIC, PCR primers are designed to append sequences that, after treatment with T4 DNA polymerase, generate 12- to 15-base overhangs that are complementary to overhangs in the vector (Aslanidis & de Jong, 1990). These longer cohesive ends make the insert-vector complex sufficiently stable to allow the transformation of hosts without ligation of the fragments.

The recombination strategy is based on the site-specific recombination reaction involved in bacteriophage  $\lambda$  integration and excision (Hartley *et al.*, 2000). PCR primers are designed to contain the specific sequences of recombination at 5' and 3' ends of the target gene, and the resulting PCR product is subcloned into a shuttle vector *via* site-specific recombination in the presence of integrase, integration host factor proteins and excisionase (*e.g.* the Gateway cloning system from Invitrogen). This clone can then either be isolated after transformation or directly used without purification for cloning into an expression vector. The major advantage of recombination cloning is that it provides

a convenient way to shuttle an insert from one vector to another and thus is useful to test multiple expression conditions.

##### 4.4.3.2. Practical application

Since the PCR reaction must be synchronized for all 96 wells in a plate, PCR primer design is important for the success of the PCR reaction. All PCR primers should have similar melting temperatures, between 50 and 60 °C. PCR products and the prepared vector are mixed and transferred into competent cells aliquoted into a 96-well plate. The PCR products and other DNA samples during the cloning, including the restriction-enzyme digest or a plasmid preparation (if needed), can be analysed *via* gel electrophoresis using a commercially available 96-well agarose gel (E-gel 96 from Invitrogen) in less than 15 min. The gel is compatible with a 96-well format and can be loaded either with a liquid-handling robot or multi-channel pipettor (Fig. 4.4.3.1a). The PCR step usually results in success rates as high as 98%. DNA purification kits are available in a 96-well format that use a vacuum manifold. As described above, the restriction-enzyme digest and ligation steps are not necessary in some cloning strategies.

Following transformation, cells are then plated into two 48-grid agar growth plates mixed with appropriate antibiotics (Fig. 4.4.3.1b). The 48-grid agar plate is made by inserting a cloning

## 4. CRYSTALLIZATION

grill into a square Petri dish before pouring the agar, and the grill remains embedded in the media. Small glass beads can be added to each segment of the 48 grids to spread the solution. Colonies are picked by a colony picker, and are grown in 96-well deep-well blocks containing media and appropriate antibiotics. Using a liquid-handling robot and a colony picker for a single round of cloning, it is possible to achieve >80% efficiency in a 96-well format. The same robotic setup can be used for the subsequent recovery of recombinant plasmid before DNA sequencing and protein expression.

### 4.4.4. Protein expression and purification

Small-scale protein expression in an HT format enables evaluation of the expression, solubility and purification of target proteins or multiple constructs. Autoinduction allows many protein expression tests in parallel without having to monitor cell densities to optimize induction, and is thus ideal for HT approaches. Each protein in a 96-well plate will have different chemical properties and sizes, and thus affinity tags are the method of choice for parallel purification of multiple proteins on a single platform. Dot blots can be used to test expression or solubility, and miniaturized resin deposited in a 96-well plate can be used for a purification test. Although up to 90 µg of protein have been purified from 1 ml of *E. coli* cell culture (Scheich *et al.*, 2003), the amount of protein obtained from small-scale purification is generally not enough for initial crystallization trials, even with low-volume crystallization, and thus scaled-up protein production and purification is needed. The purification tags can be removed upon large-scale purification.

#### 4.4.4.1. Autoinduction

Protein expression under control of the T7 *lac* promoter system can be induced either with the chemical inducer isopropyl-β-galactoside (IPTG) or by autoinduction using a mixture of glucose, glycerol and lactose during *E. coli* growth. In autoinduction, cells grow to relatively high density in a defined ratio of glucose-to-lactose media (Studier, 2005). Initially, glucose prevents induction by lactose. When available glucose is used up, lactose then induces target protein expression switching to lactose metabolism.

Cell cultures (1 ml) are grown in 2.2 ml deep-well 96-well blocks. Protein expression is autoinduced using commercially available glucose/lactose media (*e.g.* Overnight Express from EMD Biosciences). The following day, cells are lysed either by repeated freezing and thawing cycles, or by addition of a lysozyme solution. A combination of lysozyme and benzonase solutions (*e.g.* PopCulture reagent from Novagen) eliminates the need for cell harvesting prior to lysis and lysate clarification following lysis. The cell lysate can then be directly used for expression and purification tests. Alternatively, whole-cell lysates can be filtered through a 96-well filter plate, allowing for separation of inclusion bodies from the soluble fraction (filtrate) so that protein expression in the soluble fraction can be assessed.

#### 4.4.4.2. Expression and solubility test: dot blot

Dot blot is a simple method that can be used to analyse either total protein expression in the cell lysate, or soluble protein expression following the separation of supernatant and cell pellet. Protein samples from the total cell lysate in a 96-well plate are dotted onto a nitrocellulose membrane by applying a vacuum. Target proteins are then probed with an antibody against the

protein or against an affinity tag, *e.g.* anti-His-tag antibody for His-tagged protein detection (Fig. 4.4.3.1c). Alternatively, the cell lysates are analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) for protein expression using a pre-cast 48- or 96-well SDS gel (E–PAGE 48 or 96 from Invitrogen). The gel consists of 48 or 96 wells for samples, and an additional four or eight wells for protein markers. The protein can be loaded either with a liquid-handling robot or with a multi-channel pipettor, and electrophoresis is completed within 15 min (Fig. 4.4.3.1d).

#### 4.4.4.3. Small-scale purification

Affinity purifications are preferred in a 96-well format because the specific interaction between the protein and affinity resin allows for a simple ‘bind–wash–elution’ procedure. Purification using a His-tag or a GST-tag (where GST = glutathione S-transferase) is popular in structure determination projects, and appropriate resins are available as magnetic beads or agarose resin in a 96-well plate, *i.e.* magnetic Ni-NTA or GST beads, or Ni- or Co-linked agarose discs, respectively.

The soluble fraction or whole-cell lysate is transferred to a 96-well plate containing an affinity resin for protein binding. The beads are separated from unbound protein by placing them in a magnetic stand designed to accommodate a 96-well format, or by filtration using a 96-well filter plate that retains the beads but allows passage of the cell lysate. Beads are washed, and bound protein is then eluted with the appropriate elution buffer (*i.e.* imidazole and reduced glutathione for His-tagged and GST-tagged proteins, respectively) into a 96-well plate by applying a vacuum (either as part of an appropriately equipped liquid-handling robot, or by using a 96-well vacuum manifold) or by centrifugation. Purified proteins are analysed by SDS–PAGE using a pre-cast 48- or 96-well protein gel. The gels can then be stained either by protein staining or western blotting. A new system for protein and DNA analysis has been developed based on a microfluidic chip (Caliper Labchip, Caliper Life Science). These chips contain a network of miniaturized channels, through which fluids and chemicals are moved to separate DNAs or proteins. The DNA and protein signal is measured by laser-induced fluorescence. The instrument is capable of separating 2 ng–2 µg amounts of protein, one well after another; it takes about 1 h to analyse a 96-well plate.

### 4.4.5. Crystallization

HT crystallization using 96-well crystallization plates greatly reduces the total amount of protein required for screening and thus structure determination. Robotic crystallization systems are capable of dispensing nanolitre droplets (<100 nl) and hence substantially increase the number of conditions that can be screened with a fixed amount of protein sample, as well as reducing the time required for setting up a series of crystallization trials. A thousand conditions with different crystallization parameters (*e.g.* pH, salts, temperature) can be screened with 100 µl of protein sample. At the time of writing, commercially available crystallization robots include the Honey Bee (Cartesian Micro-array), Phoenix (Rigaku), Mosquito (Molecular Dimensions Limited) and Oryx (Douglas Instrumentation). These crystallization robots can set up either vapour-diffusion, microbatch or hanging-drop methods in 96-well plates within 15 min. Crystallization robots can also be integrated into a larger