



Review

# Focusing in on structural genomics: The University of Queensland structural biology pipeline

Munish Puri <sup>a,1,\*</sup>, Gautier Robin <sup>a</sup>, Nathan Cowieson <sup>a,e</sup>, Jade K. Forwood <sup>b</sup>, Pawel Listwan <sup>b,c</sup>,  
Shu-Hong Hu <sup>a</sup>, Gregor Guncar <sup>b,2</sup>, Thomas Huber <sup>b,d</sup>, Stuart Kellie <sup>a,b,c</sup>,  
David A. Hume <sup>a,b,c,e</sup>, Bostjan Kobe <sup>a,b,e</sup>, Jennifer L. Martin <sup>a,b,e</sup>

<sup>a</sup> Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia

<sup>b</sup> School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland, Australia

<sup>c</sup> Cooperative Research Centre for Chronic Inflammatory Diseases, University of Queensland, Brisbane, Queensland, Australia

<sup>d</sup> Department of Mathematics, University of Queensland, Brisbane, Queensland, Australia

<sup>e</sup> ARC Special Research Centre for Functional and Applied Genomics, University of Queensland, Brisbane, Queensland, Australia

Received 14 July 2006; received in revised form 22 September 2006; accepted 25 September 2006

## Abstract

The flood of new genomic sequence information together with technological innovations in protein structure determination have led to worldwide structural genomics (SG) initiatives. The goals of SG initiatives are to accelerate the process of protein structure determination, to fill in protein fold space and to provide information about the function of uncharacterized proteins. In the long-term, these outcomes are likely to impact on medical biotechnology and drug discovery, leading to a better understanding of disease as well as the development of new therapeutics. Here we describe the high throughput pipeline established at the University of Queensland in Australia. In this focused pipeline, the targets for structure determination are proteins that are expressed in mouse macrophage cells and that are inferred to have a role in innate immunity. The aim is to characterize the molecular structure and the biochemical and cellular function of these targets by using a parallel processing pipeline. The pipeline is designed to work with tens to hundreds of target gene products and comprises target selection, cloning, expression, purification, crystallization and structure determination. The structures from this pipeline will provide insights into the function of previously uncharacterized macrophage proteins and could lead to the validation of new drug targets for chronic obstructive pulmonary disease and arthritis.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** High throughput crystallography; Protein expression; Macrophage proteins; Crystallization; Structural genomics

## Contents

1. Introduction . . . . .	282
2. High throughput crystallography . . . . .	284
2.1. Target selection . . . . .	284
2.2. Cloning . . . . .	284
2.3. Expression and purification . . . . .	285
2.4. Crystallization . . . . .	286
2.5. Crystal to structure . . . . .	286
2.6. Protein function in macrophages and other outcomes of the pipeline . . . . .	286
3. Impact of SG . . . . .	287
3.1. Functional assignment . . . . .	287

\* Corresponding author at: Institute for Molecular Bioscience, Queensland Bioscience Precinct, University of Queensland, 306 Carmody Road, St. Lucia, Brisbane, Queensland 4072, Australia.

E-mail address: mpuri@pbi.ac.in (M. Puri).

<sup>1</sup> Present address: Fermentation and Protein Biotechnology Laboratory, Department of Biotechnology, Punjabi University, Patiala, India.

<sup>2</sup> On leave from the Department of Biochemistry and Molecular Biology, Josef Stefan Institute, Ljubljana, Slovenia.

3.2. Other implications of SG outcomes	287
3.3. Drug and drug-lead discovery	287
4. Outputs from UQ SG pipeline	287
5. Conclusion	287
Acknowledgement	288
References	288

## 1. Introduction

Structural biology has emerged as one of the most powerful approaches for defining the functions of proteins; this capacity is based on the observation that the evolutionary constraints for three-dimensional structures of proteins are higher than for sequences (Thornton et al., 2000; Yakunin et al., 2004). The strong predictive power of structure in functional annotation has resulted in the rapid growth of the new field of structural genomics (SG) (or structural proteomics) (Burley, 2000) and to the rapid development of novel high throughput technologies (Stevens and Wilson, 2001). In addition to expediting functional characterization of gene products, SG initiatives will also provide a comprehensive view of the protein structure universe, by determining the structures of representative proteins from every protein fold family and thereby filling in

protein fold space (Todd et al., 2005). SG outcomes will also identify novel drug targets (Buchanan, 2002) and advance our understanding of protein evolution. Overall, SG research promises to have a major impact on the life sciences, biotechnology and medicine (Hol, 2000).

To meet these goals, high throughput (HT) or parallel processing approaches have been developed for producing protein samples for structural biology and functional studies (Lesley et al., 2002). X-ray crystallography is the most widely used approach for protein structure determination, accounting for ~85% of structures in the Protein Data Bank (<http://www.rcsb.org/pdb/>) (Berman et al., 2000) though most SG initiatives use both crystallography and NMR approaches.

SG consortia have been established around the world (Table 1) for the systematic high throughput determination of protein structures on a genome-wide scale (Brenner, 2001).

Table 1  
List of structural genomics consortia from TargetDB (<http://targetdb.pdb.org>)

Centre/consortium	Location	Organism/focus
Berkeley Structural Genomics Center (BSGC)	USA	Structural representation of the genomes of two pathogens, <i>Mycobacterium genitalium</i> and <i>Mycobacterium pneumoniae</i>
Midwest Center for Structural Genomics (MCSG)	USA	Proteins from all three kingdoms of life
Northeast Structural Genomics Consortium (NEGS)	USA	Small proteins from eukaryotic model organisms such as <i>Saccharomyces cerevisiae</i> , <i>Caenorhabditis elegans</i> and <i>Drosophila melanogaster</i>
NewYork Structural Genomics Research Consortium (NYSGXRC)	USA	Biologically interesting proteins from model organisms and humans
Southeast Collaboratory for Structural Genomics (SECSG)	USA	<i>C. elegans</i> and <i>Pyrococcus furiosus</i> and selected human proteins
TB Structural Genomics Consortium (TB)	USA	<i>Mycobacterium tuberculosis</i> proteins, particularly potential drug targets and novel folds
Joint Center for Structural Genomics (JCSG)	USA	<i>Thermotoga maritima</i> , novel folds from <i>C. elegans</i> , and human proteins implicated in cell signaling
Center for Eukaryotic Structural Genomics (CESG)	USA	<i>Arabidopsis thaliana</i> proteins; technology development for eukaryotic proteins
Structure 2 Function Project (S2F) University of California Berkeley	USA	Functional characterization of hypothetical proteins from <i>Haemophilus influenzae</i>
Structural Genomics of Pathogenic Protozoa consortium (SGPP)	USA	Proteins from major global pathogenic protozoa, <i>Leishmania major</i> , <i>Trypanosoma brucei</i> and <i>Plasmodium falciparum</i>
Montreal-Kingston Bacterial Structural Genomics Initiative (BSGI)	Canada	Structures of potential virulence factors from pathogenic bacteria
Bacterial Targets at IGS-CNRS (BIGS)	France	Discovery of new antibacterial gene targets among evolutionary conserved genes of uncharacterized function
Oxford protein production facility (OPPF)	UK	Biomedical relevance of human pathogens, in particular Herpes viruses
Structural Proteomics in Europe (SPINE)	UK	Targets include human proteins implicated in cancer and neurodegenerative diseases
The Israel Structural Proteomics Center (ISPC)	Israel	Increase the efficiency of protein structure determination
Marseilles Structural Genomics Programme (MSGP)	France	Focus on bacterial, viral, and human ORFs
Mycobacterium Tuberculosis Structural Proteomics Project (XMTB)	Germany	Identify lead compounds against XMTB, using a structure based approach
Protein Structure Factory (PSF)	Germany	Human proteins to understand health and disease
Paris-Sud Yeast Structural Genomics (YSG)	France	Non-membrane proteins of unknown structure
RIKEN Structural Genomics Initiative (RSGI)	Japan	Proteins of biological and medical interest from mouse, <i>A. thaliana</i> and <i>Thermus thermophilus</i>
Integrated Center for Structure and Function Innovation (ISFI)	USA	Developing and applying technologies to overcome bottlenecks of production of soluble protein and protein crystallization

Some consortia target the proteomes of thermophilic organisms, which offer the advantage of working on small stable proteins, on which high-throughput methodologies can be tested (Yee et al., 2003). Although specialist membrane protein SG projects have recently emerged (Walian et al., 2004) the focus of SG consortia has been, for the most part, on soluble proteins. TargetDB (<http://targetdb.pdb.org/statistics/Target-Statistics.html>) reveals that 21 SG consortia have contributed over 2500 crystal structures since September 2000. The quality of the structures and size of the proteins are comparable to those solved by traditional structural biology approaches (Todd et al., 2005). In terms of contributing to fold space, SG consortia have been estimated to contribute around half of all novel structures reported over a recent 1-year period (Chandonia and Brenner, 2006). Furthermore, TargetDB reports that 50% of structures determined by SG consortia have less than 30% identity to known protein structures, though only around 16% of structures solved by SG were deemed to represent a new fold or superfamily (Chandonia and Brenner, 2006). The average cost of determining a protein structure by SG consortia in the US was estimated at US\$138,000; this figure increases with the novelty of the protein to \$1M for a representative of a structurally uncharacterized family and US\$2.2M for a representative structure of a new superfamily or fold (Chandonia and Brenner, 2006). These statistics show that the costs of novel protein structure determination within structural genomics consortia are still relatively high. These cost estimates make a strong case for the development of smaller, more cost-efficient research programs to tackle medically significant proteins.

At the University of Queensland (UQ) in Australia such a pipeline has been established. The UQ pipeline takes advantage of developments in high throughput processing and applies these to the parallel processing of hundreds rather than thousands of protein targets. Most importantly, the UQ pipeline has a clear focus on protein targets that are biologically or biomedically significant. Specifically, the initiative explores the structural biology of proteins that are highly expressed in macrophages and potentially involved in

the immune response. Because this population of proteins includes many of unknown structure and function, structure determination will provide insight into their mechanisms of action. In addition, it will potentially identify novel therapeutic targets and provide the all-important structural starting point for the rational design of novel drugs.

Two factors contributed to the decision to work on this specific set of proteins: (1) biomedical significance and (2) feasibility. (1) *Biomedical significance*: macrophages are specialized cells that represent the first line of defense against potential pathogens; they comprise 15–20% of cells in most organs, and are particularly abundant at the routes of pathogen entry such as lung, skin, gut and genitourinary tract (Hume et al., 2002). When a potential pathogen is recognized, the macrophage engulfs and attempts to destroy the foreign organism. The structures of proteins that regulate macrophage function therefore may form the basis for two classes of therapeutics. One option is to amplify the toxic function of macrophages to destroy foreign organisms or tumor cells more effectively. The second option is to selectively suppress the macrophage activation response as the basis for the treatment of conditions like septicaemia and toxic shock, arthritis and other chronic inflammatory diseases (Duffield, 2003). (2) *Feasibility*: full length cDNAs encoding proteins expressed in mouse macrophages, including many novel proteins, are available because of the extensive polling of the macrophage transcriptome through the FANTOM consortium (Okazaki et al., 2002; Wells et al., 2003b; Carninci et al., 2005). Furthermore, UQ has substantial in-house macrophage cell biology expertise that will add value to the structural and functional studies.

The focused HT structural biology strategy developed at UQ uses gene expression profiling for target selection so that structural information can rapidly be linked to function (Fig. 1). Through array profiling, hundreds of proteins have been identified, including many of unknown function, that are expressed in a restricted manner in the macrophage lineage or are induced by pro-inflammatory stimuli. These molecules are considered candidate regulators of innate immunity

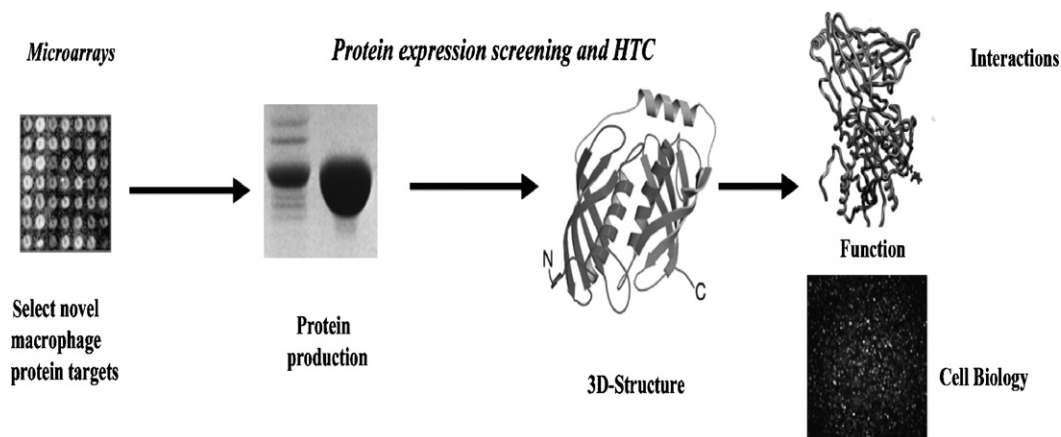


Fig. 1. Schematic diagram showing the strategy adopted by the UQ structural genomics group for parallel processing of mouse macrophage proteins. The third panel shows the crystal structure of latexin, which was solved at UQ using the pipeline. Latexin is the only known mammalian carboxypeptidase inhibitor, and is related in structure to cystatins, cysteine protease inhibitors (Aagaard et al., 2005).

(Wells et al., 2003b). In this article, we describe the parallel processing pipeline developed at UQ and aimed at determining the biological (cellular) and biochemical (molecular) functions of important uncharacterized or novel macrophage proteins that are likely to play a role in the immune response and in inflammation (Aagaard et al., 2005; Cowieson et al., 2005).

## 2. High throughput crystallography

The five key stages in the process of high throughput crystallography (HTC) (Fig. 2) are: (1) target selection; (2) cloning; (3) expression and purification; (4) crystallization screening and optimization; (5) crystal structure determination. Details of the procedures employed at UQ for each of these steps are described below. In addition a further step has been initiated at UQ to incorporate cell biology into the structural biology program; initially, this will involve the investigation of functional consequences of over-expression of the target proteins in macrophage cells and cellular localization studies.

### 2.1. Target selection

The first crucial step in any structural genomics project is the selection and prioritization of target proteins for structure determination; the selection criteria vary depending on the SG program. Targets are often selected because they are likely to have a novel fold or are representative of a large protein family for which no structure is known. These criteria ensure that resulting structures will contribute to the fundamental aim of structural genomics to fill in fold space. Proteins are also more likely to be selected if they are considered to be tractable for structure determination. Thus, for example, targets that are

predicted to be membrane proteins or that are cysteine-rich would be less likely to be selected under this criterion. The more rigorous the criteria used for selection, the greater the chances of progression through the pipeline and the more value there is in those structures that emerge from the pipeline.

At UQ, gene expression analysis is included as a criterion for target selection to maximize biological and biomedical value of selected targets. The combination of expression profiling and HT structural analysis provides a very powerful way to identify protein function, as the former suggests a cellular role of a protein (e.g. involvement in a pathway), while the latter suggests a molecular (biochemical) function (e.g. an enzymatic function). At the same time, this strategy retains the cost-effective nature of pursuing the technically more tractable structures first. The biological focus is therefore on proteins that are likely to play a role in the immune response and inflammation (Wells et al., 2003a). The genes corresponding to selected proteins are those that we have shown, using cDNA microarray experiments, to be (a) expressed selectively in macrophages and/or (b) transcriptionally regulated following stimulation of mouse macrophages and/or (c) transcriptionally regulated in mouse models of arthritis or chronic obstructive pulmonary disease. Genes that meet these selection criteria are then further filtered to remove those that do not have human orthologues (sequence identity minimum of 70%). This criterion filters out wrongly predicted open reading frames and simultaneously ensures that the selected mouse proteins will yield information that can be translated to human biology. Targets that are greater than 60 kDa are rejected because these are unlikely to express well in bacteria. Targets that are less than 10 kDa are also rejected because these are more likely to be spurious open reading frames. Sorting signals are predicted using PSORT (Nakai and Horton, 1999) and transmembrane regions are deleted from the expression constructs. Candidate transmembrane regions were annotated systematically by the FANTOM consortium (Okazaki et al., 2002). Targets that have transmembrane regions or sorting signals are not necessarily filtered out during selection of UQ targets, but these regions of sequence may be removed if the targets are selected for further processing.

Target proteins for which structures are already known are filtered out after identification by BLAST (Altschul et al., 1990) comparison with sequences in the Protein Data Bank or by threading and secondary structure prediction using 3D-PSSM (Kelley et al., 2000). The cutoff used at UQ is 30% pairwise sequence identity with a protein of known structure. Proteins with lower than 30% identity may have similar structures, but they are difficult to model reliably. The Pfam database (Bateman et al., 2004) is used at UQ to identify whether targets have individual domains that are structurally characterized; such targets may still be selected if other domains in the sequence are not structurally characterized.

### 2.2. Cloning

Modern molecular biology approaches are well-suited to automation and handling of large numbers of genes. At UQ,

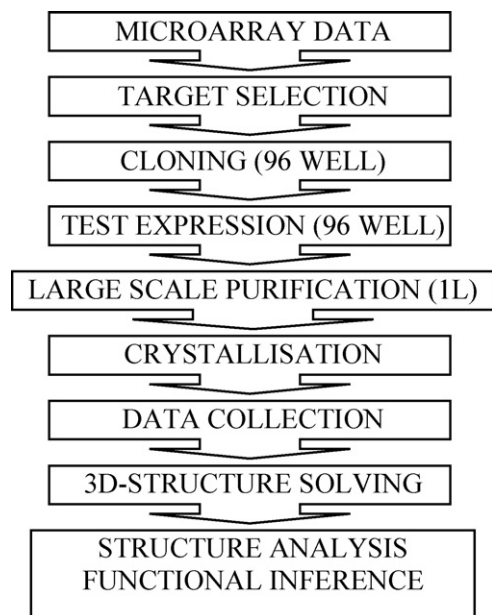


Fig. 2. Flowchart showing the major experimental steps involved in protein production at UQ. The major points of pipeline optimization and automation are cloning, expression, purification and crystallization trials.

### Modified Gateway Cloning

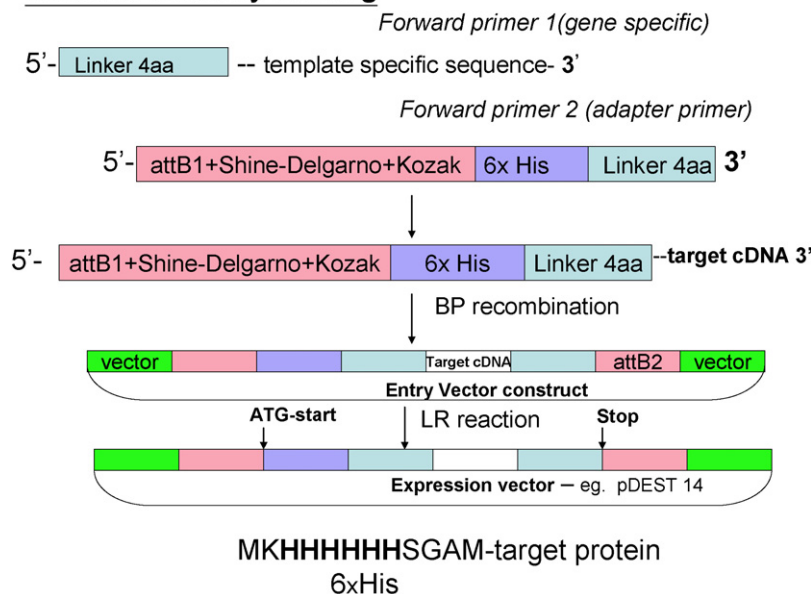


Fig. 3. Modified Gateway Cloning allowing the generation of a “minimal” hexa-histidine affinity tags; firstly, gene specific primers containing a 12-nucleotide adapter are used to amplify the target gene from a cDNA template using the PCR reaction. Adapter primers are then incorporated that encode the affinity tag during a second round of PCR. The GATEWAY cloning technology is then applied as per described in the Invitrogen manual to incorporate the amplified PCR products into a pDONr entry clone using the BP reaction, and subsequently in the pDEST14 expression vector using LR cloning (Listwan et al., 2005). BP (a reaction between an expression clone ‘attB’ flanked PCR product and a donor vector containing ‘attP’ sites to create an entry clone), LR (a recombination reaction between an entry clone ‘attL’ and a destination vector containing ‘attR’, mediated by host of recombination proteins, to create an expression clone.

standard 96-well plates are used with a Biomek 2000 liquid handling robot to perform most of the steps involved in cloning and in subsequent small scale protein expression screening. A modification of the Invitrogen Gateway<sup>®</sup> system is used to clone PCR products into an entry vector (Fig. 3). Transformation is performed in the standard manner by heat shock into chemically competent cells. Colonies are then picked manually into 1-ml cultures in a 96-well plate and grown overnight; plasmid purification from these cultures is automated by using the Biomek robot and commercial 96-well compatible kits. A novel nested PCR modification of the Gateway<sup>®</sup> cloning system was developed at UQ to adapt the resulting expression construct better for structural biology applications (Listwan et al., 2005). Using this approach, genes cloned into the donor vector are recombined into the expression vector without the need for specific AttB sites; the redesigned expression vector is also engineered to express fusion proteins with very short hexahistidine tags.

#### 2.3. Expression and purification

Bacterial expression systems represent the cheapest, quickest and easiest means of producing recombinant protein. However, the percentage of targets that proceed through the pipeline to yield soluble, purified protein from a bacterial expression system is dependent on many variables including those described above in target selection, as well as the presence of rare codons in the target gene and the particular experimental methods chosen (Braun et al., 2002). Expression

of mammalian proteins in bacteria has a much lower success rate than expression of bacterial proteins. For specific types of targets, such as membrane proteins, a range of bacterial strains are often trialled (Miroux and Walker, 1996). Expression vectors usually incorporate engineered tags to enable affinity purification of the target protein. Common examples include the hexa-histidine (6xHis), glutathione S-transferase (GST), maltose-binding protein (MBP), or thioredoxin (TRX) tags. The type of tag used can influence both expression levels and the solubility of expressed proteins (Hammarstrom et al., 2002). The 6xHis tag is small enough that the expressed proteins can often be crystallized without the need to remove the tag (Lesley et al., 2002).

At UQ, the strategy is to use bacterial expression systems to generate recombinant 6xHis-tagged mammalian proteins. The expression vectors, each containing a gene coding for a mouse macrophage protein, are transformed by heat shock into chemically competent *Escherichia coli* cells (strain BL21(DE3)pLysS). The cells are cultured in small scale (1 ml) overnight in 96-well plates and induction of expression involves autoinduction media (Studier, 2005). These media allows for fully automated growth of cultures to high optical densities without the need for monitoring growth.

For HT, the traditional means of cell lysis to release recombinant proteins (sonication or freeze-thawing cycles) are unsuitable. At UQ, the 96-well frozen cell pellets are lysed using a chemical lysis reagent. The recombinant proteins are purified in HT using the Biomek liquid handling robot and magnetic metal affinity resin. The nickel-containing resin binds

to the 6xHis-tagged proteins; unbound cellular debris is removed by washing and the 6xHis-tagged fusion proteins are eluted with a gradient of imidazole (20–250 mM). The purification of 96 ml × 1 ml cultures in a 96-well plate format, from cell culture to eluted protein, is completed within 90 min using the automated liquid-handling workstation. To analyse expression levels and solubility, we make use of the LabChip90 automated electrophoresis system for DNA and proteins ([www.caliperls.com](http://www.caliperls.com)). Samples of whole-cell extracts, soluble fractions and elutions from metal affinity resin are analyzed on a protein chip in 96-well format. The resulting information from these small-scale expressions allows identification of targets that are produced in soluble form in bacteria. These protein targets are then expressed in large scale (1–2 l) for crystallization trials and the proteins purified by metal affinity and size exclusion chromatography. Protein purity and homogeneity is further checked by MALDI-TOF mass spectrometry (Leushner, 2001), by gel filtration chromatography, circular dichroism (Kelly et al., 2005) and by dynamic light scattering (Wilson, 2003) prior to progression through to crystallization trials.

The solubility of expressed proteins is a major bottleneck, particularly when expressing mammalian proteins in bacteria. We have also developed a matrix-assisted refolding approach in which correctly folded proteins are distinguished from misfolded proteins by their elution from affinity resin under non-denaturing conditions. The assay can be applied to insoluble proteins on an individual basis but is also suited to automation (Cowieson et al., 2006).

#### 2.4. Crystallization

The production of diffraction-quality crystals by screening variables such as precipitants, pH and temperature often represents another bottleneck in SG pipelines. The number of parameters to be evaluated results in a large multi-dimensional sampling space, but in most cases the amount of protein available is small, so that the number of crystallization experiments needs to be minimised. A recent advance has been the introduction of methodology that allows the set-up of crystallization trials using nanolitre amounts of protein (Stevens and Wilson, 2001). The most common method used for crystallization screening is vapor diffusion, either by hanging drop or sitting drop; both are amenable to HT application in 96-well plate format. Free interface diffusion methods have also been developed (Hansen et al., 2002) that are suited to HT. Robots are now commonly used for all three steps in crystallization (Bard et al., 2004), namely preparation of crystallization formulations, small volume (50–200 nl) crystallization experiment set-up and monitoring of the crystallization experiments. Furthermore, many crystallization screens are available commercially (e.g. Hampton Research, Emerald Biostructures, Jena Bioscience, Molecular Dimensions) or can be prepared based on previously identified successful formulations (Page et al., 2003). At UQ, the crystallization pipeline involves first establishing the optimal protein concentration, by using a pre-screen from Hampton Research.

Then for each purified protein, a panel of five commercial or lab-prepared screens are set-up at two different temperatures (20 °C and 4 °C). The 100 nl + 100 nl sitting drops in the 96-well tray format are set-up using a TTP LabTech Mosquito robot and experiments are monitored using a DeCode Genetics Crystal Monitor. Hits are then optimized by setting up focused screens (Senger and Mueser, 2005).

#### 2.5. Crystal to structure

Over the past 15 years, the process of protein structure determination by crystallography has been revolutionized by developments in methodology, hardware and software. Indeed, it is now possible to determine the structure of a novel protein in a matter of hours from the time data measurement begins. The improvements are due to faster and simpler phasing methods such as multiple (MAD) and single wavelength anomalous dispersion (SAD) (Hendrickson, 1999), more intense X-ray sources (lab-based and at synchrotron beamlines) allowing quicker data collection and measurement of higher resolution data, improvements in computing hardware and more robust and automated software packages for processing and solving structures (Lamzin and Perrakis, 2000).

At UQ, the structure determination process uses the standard approaches of cryocrystallography and MAD phasing from selenomethionine (SeMet)-substituted protein crystals. Crystals of native protein are flash-cooled in a nitrogen gas stream at ~100 K after soaking in a suitable cryoprotectant. X-ray diffraction quality of protein crystals is assessed using the laboratory equipment. When diffraction-quality crystals are obtained and a suitable molecular replacement model is not available, the protein is expressed in minimal media in the presence of SeMet to produce SeMet-labelled protein crystals for use in MAD phasing at a synchrotron. Standard crystallographic packages are used to process and phase the data and to visualize and refine structures (Aagaard et al., 2005).

#### 2.6. Protein function in macrophages and other outcomes of the pipeline

In such a comparatively small program, proteins that are cloned into expression plasmids and which fail to express in bacteria, or which are not soluble, or which fail to generate crystals, represent a significant loss of productivity. The advantage of using the Gateway system is that the cDNAs can be cloned from the entry vector into mammalian expression vectors to enable us to investigate function in the absence of structural outcomes. We have now begun to systematically clone candidate targets into commercial mammalian destination vectors, and into an inducible expression system. These vectors can be transiently or stably transfected into the mouse macrophage cell line, RAW264. Epitope or His-tags are used to localize the expressed protein within the cell, and if expression is successful, to extract and purify the protein and aid in the identification of binding partners. Over-expression of the target protein may modify the biological function of the cells; we can routinely assay cell proliferation, adhesion and spreading, and

inducible production of inflammatory cytokines. Clearly, for proteins that do generate crystal structures, such information, combined with information about mRNA expression and regulation in macrophages, can reinforce and strengthen structure-based functional inferences.

In the event that a soluble protein is produced but fails to crystallize, we have initiated an alternative approach to gain structural information using chemical cross-linking. The assignment of proximity relationships, combined with molecular modeling, provides the structural information in this case. We used this method to identify the site of interaction between latexin, the first crystal structure from the UQ structural genomics program, and its partner protein carboxypeptidase A (Mouradov et al., 2006).

### 3. Impact of SG

Each new protein structure, whether it originates from an international consortium or a small focused program like that at UQ, provides a wealth of information for fundamental, applied and strategic research. For example, when a new protein fold is revealed, the database of known protein folds is enriched, and when the function of a new protein is determined, novel structure–function relationships are established (Zhang and Kim, 2003). In this section we comment briefly on the impact of SG protein structure outcomes.

#### 3.1. Functional assignment

A major goal of SG initiatives is the identification of function for uncharacterized gene products. It is estimated that over 2/3 of structures of proteins of unknown function can be used to directly infer a molecular function (Christendat et al., 2000; Kim et al., 2005), most often because the structures are remote homologues of proteins of known function (evaluated for example using DALI (Holm and Sander, 1994) and SCOP (Murzin et al., 1995)) or because ligands fortuitously co-crystallize bound to the protein. For uncharacterized proteins with a new fold, *in silico* approaches are being developed to use structure to locate binding sites and to help assign function (Laskowski et al., 2005; Pal and Eisenberg, 2005).

#### 3.2. Other implications of SG outcomes

As described in Section 1, one of the major goals of SG is to provide representative structures for every protein fold family. These representative structures can then be used as templates to generate protein models for all other members within that protein family (Chance et al., 2004), using automated methods such as MODELLER (John and Sali, 2003). In addition, the wealth of structural information will underpin fundamental improvements in knowledge including protein folding, protein structure prediction, and protein evolution. In terms of biomedical impact, structural data will facilitate the design of improved therapeutic agents by allowing comparison of functionally similar protein structures from pathogens and hosts, or proteins from diseased and normal tissues.

#### 3.3. Drug and drug-lead discovery

It is now widely accepted that protein structure is critical for the design of new drugs. However, the impact of SG and HT technology is now extending beyond structure determination and into new approaches for drug and drug target discovery (Claverie et al., 2002). In this context, X-ray crystallography has a major advantage of defining ligand-binding sites in intricate detail (Kuhn et al., 2002). Drug-like fragment libraries have been developed that sample chemical space more efficiently than ligand-based screens (Hann et al., 2001), and these are screened with a single target protein to identify fragments that bind and to define precisely the binding sites (Blundell and Patel, 2004). These approaches rely on robots for soaking the crystals with the fragment cocktails and to collect the X-ray diffraction data. Automated software is also being developed to identify the fragments that bind. Those drug-like fragments that are found to bind to a protein target (the hits) become the starting point for medicinal chemistry to develop potent and selective leads that may then further evolve into drugs.

### 4. Outputs from UQ SG pipeline

A total of 318 macrophage genes have been processed in three rounds of expression screening at UQ. The Gateway<sup>®</sup> approach was used to successfully clone 220 of these mouse protein targets into bacterial expression vector. These generated 52 soluble mouse macrophage proteins that have entered crystallization trials. Crystals have been produced for 6 of the 52 proteins and two crystal structures have been solved. One of these two structures is that of latexin, also referred to as tissue carboxypeptidase inhibitor, which is the only known mammalian carboxypeptidase inhibitor. The structure reveals that latexin comprises two cystatin folds in tandem (Aagaard et al., 2005) (Fig. 1). The cystatin fold is commonly found in inhibitors of cysteine proteases, but has not before been observed in tandem as a pair. The second structure solved is that of long chain acyl-CoA thioesterase (Serek et al., 2006). The success rates from the UQ pipeline to date for the 318 mouse macrophage proteins (16% soluble per selected target; 12% crystals per soluble protein; 33% structures per crystal) is comparable to the success rates reported in TargetDB for eukaryotic proteins (8% purified per selected target; 24% crystallized per purified protein; 33% structures per crystal, for 57,153 eukaryotic targets, 12 September 2006).

### 5. Conclusion

Parallel processing of thousands of protein targets is now a reality for SG consortia. However, these approaches are not limited to genome-wide studies of protein structure. On the contrary, the established protocols and automated infrastructure open up enormous possibilities for other protein science programs, because it provides the speed and throughput necessary to rapidly identify, from hundreds or thousands of starting constructs, those proteins, expression systems, crystallization conditions or crystals, which structural studies

should focus on. We have employed exactly that strategy at UQ, by using techniques and infrastructure suitable for an academic laboratory and focusing on proteins with strong biological importance. We note that several other academic labs have developed structural genomics technology in what has been termed “medium-throughput” structural biology projects (Claverie et al., 2002; Vincentelli et al., 2003; Segelke et al., 2004; Busso et al., 2005; Moreland et al., 2005). The next goal for academia and consortia alike will be to progress from a high-throughput to a higher-output mode, and to proceed from analyzing “low-hanging fruit” to addressing technically challenging targets such as membrane proteins and macromolecular assemblies.

## Acknowledgements

MP thanks DEST for the award of an Australia-Asia Fellowship. The work at UQ is supported by an Australian Research Council (ARC) grant to JLM and BK. BK is an ARC Federation Fellow and National Health and Medical Research Council (NHMRC) Honorary Research Fellow. JKF is an NHMRC CJ Martin Research Fellow. NC is an Australian Synchrotron Research Program Fellow.

## References

- Aagaard, A., Listwan, P., Cowieson, N., Huber, T., Ravasi, T., Wells, C.A., Flanagan, J.U., Kellie, S., Hume, D.A., Kobe, B., Martin, J.L., 2005. An inflammatory role for the mammalian carboxypeptidase inhibitor latexin: relationship to cystatins and the tumor suppressor TIG1. *Structure* 13, 309–317.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Bard, J., Ercolani, K., Svenson, K., Olland, A., Somers, W., 2004. Automated systems for protein crystallization. *Methods* 34, 329–347.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., Eddy, S.R., 2004. The Pfam protein families database. *Nucl. Acids Res.* 32, D138–D141.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. *Nucl. Acids Res.* 28, 235–242.
- Blundell, T.L., Patel, S., 2004. High-throughput X-ray crystallography for drug discovery. *Curr. Opin. Pharmacol.* 4, 490–496.
- Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., LaBaer, J., 2002. Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2654–2659.
- Brenner, S.E., 2001. A tour of structural genomics. *Nat. Rev. Genet.* 2, 801–809.
- Buchanan, S.G., 2002. Structural genomics: bridging functional genomics and structure-based drug design. *Curr. Opin. Drug Discov. Devel.* 5, 367–381.
- Burley, S.K., 2000. An overview of structural genomics. *Nat. Struct. Biol.* 7 (Suppl.), 932–934.
- Busso, D., Poussin-Courmontagne, P., Rose, D., Ripp, R., Litt, A., Thierry, J.C., Moras, D., 2005. Structural genomics of eukaryotic targets at a laboratory scale. *J. Struct. Funct. Genom.* 6, 81–88.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Huminecki, L., Iacono, M., Ikey, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusci, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., 2005. The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.
- Chance, M.R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G., Fajardo, J.E., Radhakannan, T., Marinkovic, N., 2004. High-throughput computational and experimental techniques in structural genomics. *Genome Res.* 14, 2145–2154.
- Chandonia, J.M., Brenner, S.E., 2006. The impact of structural genomics: expectations and outcomes. *Science* 311, 347–351.
- Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, L.P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, E.F., Gerstein, M., Edwards, A.M., Arrowsmith, C.H., 2000. Structural proteomics of an archaeon. *Nat. Struct. Biol.* 7, 903–909.
- Claverie, J.M., Monchois, V., Audic, S., Poirot, O., Abergel, C., 2002. In search of new anti-bacterial target genes: a comparative/structural genomics approach. *Comb. Chem. High Throughput Screen.* 5, 511–522.
- Cowieson, N.P., Listwan, P., Kurz, M., Aagaard, A., Ravasi, T., Wells, C., Huber, T., Hume, D.A., Kobe, B., Martin, J.L., 2005. Pilot studies on the parallel production of soluble mouse proteins in a bacterial expression system. *J. Struct. Funct. Genom.* 6, 13–20.
- Cowieson, N.P., Wensley, B., Listwan, P., Hume, D.A., Kobe, B., Martin, J.L., 2006. An automatable screen for the rapid identification of proteins amenable to refolding. *Proteomics* 6, 1750–1757.
- Duffield, J.S., 2003. The inflammatory macrophage: a story of Jekyll and Hyde. *Clin. Sci. (Lond.)* 104, 27–38.
- Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., Hard, T., 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* 11, 313–321.
- Hann, M.M., Leach, A.R., Harper, G., 2001. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864.
- Hansen, C.L., Skordalakes, E., Berger, J.M., Quake, S.R., 2002. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16531–16536.
- Hendrickson, W., 1999. Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Radiat.* 6, 845–851.
- Hol, W.G., 2000. Structural genomics for science and society. *Nat. Struct. Biol.* 7 (Suppl.), 964–966.
- Holm, L., Sander, C., 1994. The FSSP database of structurally aligned protein fold families. *Nucl. Acids Res.* 22, 3600–3609.



- Hume, D.A., Ross, I.L., Himes, S.R., Sasmono, R.T., Wells, C.A., Ravasi, T., 2002. The mononuclear phagocyte system revisited. *J. Leukoc. Biol.* 72, 621–627.
- John, B., Sali, A., 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucl. Acids Res.* 31, 3982–3992.
- Kelley, L.A., MacCallum, R.M., Sternberg, M.J., 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* 299, 499–520.
- Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* 1751, 119–139.
- Kim, S.H., Shin, D.H., Liu, J., Oganessian, V., Chen, S., Xu, Q.S., Kim, J.S., Das, D., Schulze-Gahmen, U., Holbrook, S.R., Holbrook, E.L., Martinez, B.A., Oganessian, N., DeGiovanni, A., Lou, Y., Henriquez, M., Huang, C., Jancarik, J., Pufan, R., Choi, I.G., Chandonia, J.M., Hou, J., Gold, B., Yokota, H., Brenner, S.E., Adams, P.D., Kim, R., 2005. Structural genomics of minimal organisms and protein fold space. *J. Struct. Funct. Genom.* 6, 63–70.
- Kuhn, P., Wilson, K., Patch, M.G., Stevens, R.C., 2002. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr. Opin. Chem. Biol.* 6, 704–710.
- Lamzin, V.S., Perrakis, A., 2000. Current state of automated crystallographic data analysis. *Nat. Struct. Biol.* 7 (Suppl.), 978–981.
- Laskowski, R.A., Watson, J.D., Thornton, J.M., 2005. ProFunc: a server for predicting protein function from 3D structure. *Nucl. Acids Res.* 33, W89–W93.
- Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H.E., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L.S., Miller, M.D., McPhillips, T.M., Miller, M.A., Scheibe, D., Canaves, J.M., Guda, C., Jaroszewski, L., Selby, T.L., Elsliger, M.A., Wooley, J., Taylor, S.S., Hodgson, K.O., Wilson, I.A., Schultz, P.G., Stevens, R.C., 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11664–11669.
- Leushner, J., 2001. MALDI TOF mass spectrometry: an emerging platform for genomics and diagnostics. *Expert Rev. Mol. Diagn.* 1, 11–18.
- Listwan, P., Cowieson, N., Kurz, M., Hume, D.A., Martin, J.L., Kobe, B., 2005. Modification of recombinatorial cloning for small affinity tag fusion protein construct generation. *Anal. Biochem.* 346, 327–329.
- Miroux, B., Walker, J.E., 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298.
- Moreland, N., Ashton, R., Baker, H.M., Ivanovic, I., Patterson, S., Arcus, V.L., Baker, E.N., Lott, J.S., 2005. A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallogr. D: Biol. Crystallogr.* 61, 1378–1385.
- Mouradov, D., Craven, A., Forwood, J.K., Flanagan, J.U., Garcia-Castellanos, R., Gomis-Ruth, F.X., Hume, D.A., Martin, J.L., Kobe, B., Huber, T., 2006. Modelling the structure of latexin-carboxypeptidase A complex based on chemical cross-linking and molecular docking. *Protein Eng. Des. Sel.* 19, 9–16.
- Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.
- Nakai, K., Horton, P., 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24, 34–36.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusica, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y., 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563–573.
- Page, R., Grzechnik, S.K., Canaves, J.M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R.C., Lesley, S.A., 2003. Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the *Thermotoga maritima* proteome. *Acta Crystallogr. D: Biol. Crystallogr.* 59, 1028–1037.
- Pal, D., Eisenberg, D., 2005. Inference of protein function from protein structure. *Structure* 13, 121–130.
- Segelke, B.W., Schafer, J., Coleman, M.A., Lakin, T.P., Toppani, D., Skowronek, K.J., Kantardjiev, K.A., Rupp, B., 2004. Laboratory scale structural genomics. *J. Struct. Funct. Genom.* 5, 147–157.
- Senger, A., Mueser, T., 2005. Rapid preparation of custom grid screens for crystal growth optimization. *J. Appl. Cryst.* 38, 847–850.
- Serek, R., Forwood, J.K., Hume, D.A., Martin, J.L., Kobe, B., 2006. Crystallization of the C-terminal domain of the mouse brain cytosolic long-chain acyl-CoA thioesterase. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun.* 62, 133–135.
- Stevens, R.C., Wilson, I.A., 2001. Tech.Sight. Industrializing structural biology. *Science* 293, 519–520.
- Studier, F.W., 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234.
- Thornton, J.M., Todd, A.E., Milburn, D., Borkakoti, N., Orengo, C.A., 2000. From structure to function: approaches and limitations. *Nat. Struct. Biol.* 7 (Suppl.), 991–994.
- Todd, A.E., Marsden, R.L., Thornton, J.M., Orengo, C.A., 2005. Progress of structural genomics initiatives: an analysis of solved target structures. *J. Mol. Biol.* 348, 1235–1260.
- Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V., Cambillau, C., 2003. Medium-scale structural genomics: strategies for protein expression and crystallization. *Acc. Chem. Res.* 36, 165–172.
- Walian, P., Cross, T.A., Jap, B.K., 2004. Structural genomics of membrane proteins. *Genome Biol.* 5, 215.
- Wells, C.A., Ravasi, T., Faulkner, G.J., Carninci, P., Okazaki, Y., Hayashizaki, Y., Sweet, M., Wainwright, B.J., Hume, D.A., 2003a. Genetic control of the innate immune response. *BMC Immunol.* 4, 5.
- Wells, C.A., Ravasi, T., Sultana, R., Yagi, K., Carninci, P., Bono, H., Faulkner, G., Okazaki, Y., Quackenbush, J., Hume, D.A., Lyons, P.A., 2003b. Continued discovery of transcriptional units expressed in cells of the mouse mononuclear phagocyte lineage. *Genome Res.* 13, 1360–1365.
- Wilson, W.W., 2003. Light scattering as a diagnostic for protein crystal growth—a practical approach. *J. Struct. Biol.* 142, 56–65.
- Yakunin, A.F., Yee, A.A., Savchenko, A., Edwards, A.M., Arrowsmith, C.H., 2004. Structural proteomics: a tool for genome annotation. *Curr. Opin. Chem. Biol.* 8, 42–48.
- Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M., Arrowsmith, C.H., 2003. Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. *Acc. Chem. Res.* 36, 183–189.
- Zhang, C., Kim, S.H., 2003. Overview of structural genomics: from structure to function. *Curr. Opin. Chem. Biol.* 7, 28–32.