Overview of the Pipeline for Structural and Functional Characterization of Macrophage Proteins at the University of Queensland

Weining Meng[1,2], Jade K. Forwood[1,2], Gregor Guncar[1,2], Gautier Robin[1,2], Nathan P. Cowieson[2], Pawel Listwan[1,2], Dmitri Mouradov[1], Gordon King[2,3], Ian L. Ross[2,3], Jodie Robinson[2,3], Munish Puri[2], Justine M. Hill[1,2], Stuart Kellie[1,2,3], Thomas Huber[1], David A. Hume[1,2,3,4], Jennifer L. Martin[1,2,4] and Bostjan Kobe[1,2,4]

[1] School of Molecular and Microbial Sciences; [2] Institute for Molecular Bioscience; [3] Cooperative Research Centre for Chronic Inflammatory Diseases; [4] ARC Special Research Centre for Functional and Applied Genomics, University of Queensland.

Corresponding author: Bostjan Kobe, School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia; email: b.kobe@uq.edu.au; phone: +617-3365-2132; fax: +617-3365-4699.

Running title: Structural proteomics of macrophage proteins

Abstract

This chapter describes the methodology adopted in a project aimed at structural and functional characterization of proteins that potentially play an important role in mammalian macrophages. The methodology that underpins this project is applicable to both small research groups and larger structural genomics consortia. Gene products with putative roles in macrophage function are identified using gene expression information obtained via DNA microarray technology. Specific targets for structural and functional characterization are then selected based on a set of criteria aimed at maximizing insight into function. The target proteins are cloned using a modification of Gateway® cloning technology, expressed with hexa-histidine tags in *E. coli*, and purified to homogeneity using a combination of affinity and size exclusion chromatography. Purified proteins are finally subjected to crystallization trials and/or NMR-based screening to identify candidates for structure determination. Where crystallography and NMR approaches are unsuccessful, chemical cross-linking is employed to obtain structural information. This resulting structural information is used to guide cell biology experiments to further investigate the cellular and molecular function of the targets in macrophage biology. Jointly, the data sheds light on the molecular and cellular functions of macrophage proteins.

Key Words: cell biology, cross-linking, Gateway technology, high-throughput crystallography, macrophages, nuclear magnetic resonance (NMR), structural genomics, X-ray diffraction.

1. Introduction

Structural genomics initiatives aim to provide a comprehensive view of the protein structure universe by determining structures of representative proteins from every protein family (*1*). Achieving such a goal requires large, coordinated research teams and substantial funding (*2*). However, the parallel processing and high-throughput approaches adopted by structural genomics initiatives can also be applied to projects of smaller scale, and promise faster and more cost-effective results. Furthermore, a smaller team can identify a niche in the world-wide structural genomics initiative through careful protein target selection. We have applied these ideas to a project aimed at the structural characterization of proteins that play important roles in macrophage function.

Macrophages are cells representing the first line of defence against pathogens and play crucial roles in both innate and acquired immunity. 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 (*3*). Macrophages detect pathogens by receptors that recognise generic non-mammalian structures including cell wall components (e.g. lipopolysaccharide, LPS; peptidoglycans; lipotechoic acids) and microbial DNA (e.g. unmethylated CpG motifs) (*4*). Upon recognition, the macrophage engulfs and destroys the foreign organism, while at the same time activating a spectrum of genes, creating a hostile extracellular environment in the host. Additional cells are recruited to the site of invasion and an appropriate acquired immune response is primed dependent on the

class of pathogen. However, some pathogens have been able to evade these defences, and in some cases, such as *Mycobacterium tuberculosis*, take advantage of the macrophage as a portal of infection and replicate or survive within this cell type. This can lead to life-threatening conditions such as disseminated intravascular coagulation, hypotension, and pathological fever (**5**). In chronic local infections, or in response to inflammation caused by non-infectious agents that activate macrophages but cannot be cleared, macrophage products cause local tissue destruction and the wasting disease known as cachexia (**6**).

A detailed knowledge of the regulation of macrophage function will form the basis for the development of two classes of therapeutics. On one hand, it may be desirable to amplify the toxic function of macrophages to destroy microorganisms or tumor cells more effectively. Alternatively, selective suppression of components of the macrophage activation response offers approaches to treatment of acute conditions such as septicemia and toxic shock, and chronic conditions such as arthritis, atherosclerosis and obstructive lung disease (**7**).

To understand better the process of macrophage activation, we have undertaken a program to structurally and functionally characterize novel proteins involved in macrophage activation (**8**). Here we present the methodology of our pipeline (**Fig. 1**), focusing on target selection, cloning, expression, purification, and structural characterization of proteins involved in macrophage activation. The pipeline is applicable to both small research groups within academia and larger consortia.

2. Methods

2.1. Target Selection

Proteins with likely roles in macrophage function are identified based on expression profiling using DNA microarrays. The selected proteins are (a) expressed selectively in mouse macrophages and/or (b) transcriptionally regulated following stimulation of mouse macrophages *in vitro* and/or (c) transcriptionally regulated in mouse models of arthritis or chronic obstructive pulmonary disease. Only proteins with human homologues are chosen for further study to ensure medical relevance. In order to maximize the value of the determined structures, only the proteins with less than 30% amino acid sequence identity to proteins with known three-dimensional structures (Protein Data Bank; PDB (*9*)) are selected. Protein fold recognition servers such as PHYRE (*10*) and FUGUE (*11*) are used to further examine the predicted structures. Target proteins are prioritized based on their expected suitability for structural studies, using a number of properties inferred from the sequence and functional annotation (for example the presence of putative transmembrane regions predicted by TMHMM (*12*), sequence length, isoelectric point, percentage of charged residues, hydropathy index). An examination of the relevant literature is carried out on top-ranking proteins to produce the final target list. The details of the target selection procedure, the associated Web tools and customization options are presented in a separate chapter in this volume (Robin et al., A target selection method for crystallographic proteomics).

2.2. Cloning

For all high-throughput steps involving liquid handling, we use the Biomek 2000 automated liquid handling workstation and 96-well plates. The Triple Master PCR system (Eppendorf) is employed for all PCR amplifications. One shot TOP10 chemically competent *E. coli* cells (Invitrogen) are used for all transformation reactions. The primer sequences are generated automatically using an in-house Perl script that takes target nucleotide sequences as an input, generating primer sequences that start at the termini, contain additional bases to reach an annealing temperature of 55 °C, and end in a cytosine or guanine.

The target genes are amplified by PCR using a macrophage cDNA pool or cDNAs from FANTOM2 clonesets (*13*). The PCR products are purified from primers and other buffer components using the Montage 96-well PCR purification kit (Millipore) set up on the Biomek 2000 workstation.

The PCR products are cloned into expression vectors using a modification of the Gateway recombinatorial cloning methodology that allows expression of the recombinant proteins with a short hexa-histidine tag (*14, 15*). The purified PCR products are cloned into the Gateway entry vector pDONR-221 (Invitrogen) using a two-step PCR followed by a recombinatorial LR reaction (*15*). A gene-specific primer containing a linker of 12 nucleotides is used in the first PCR step, and a BP-reaction "universal adapter primer" (*15*) containing hexa-histidine tag is employed in the second PCR step. The reaction enzyme BP Clonase facilitates the recombination between a specific sequence (attB) included in the product of second PCR and the attP sequence of the donor vector, pDONR-221. Following the BP reaction, One Shot TOP10

chemically competent *E. coli* cells (Invitrogen) are used to transform the reaction mixture. The colonies are grown overnight in LB media containing kanamycin and the plasmids are then purified using the Plasmid Miniprep96 Montage kit (Millipore). The constructs are analyzed by restriction enzyme digestion with BsrG1. The genes contained within the pDONR vector are transferred into the pDEST14 expression vector using the LR reaction. The expression vectors are assayed for correct insertion of the gene by digestion with the restriction enzyme BsrG1, followed by electrophoresis. The positive clones are additionally confirmed by DNA sequencing (Australian Genome Research Facility).

2.3. Expression and Purification

As reported by the worldwide SG Centers TargetDB Statistics Report (Oct 2, 2006), only about 25% (30% for prokaryotes and viruses, 15% for eukaryotes) of clones can be successfully expressed and purified. The main hurdle is protein solubility in the non-native host. We use conventional expression and purification procedures for the soluble targets, while protein refolding (Section 2.4) is used as a salvage pathway for the insoluble proteins. An estimation of the protein expression and solubility in *E. coli* is achieved rapidly in 1-ml cultures followed by purification in 96-well format. The analysis with an automated electrophoresis instrument (Caliper 96 Bioanalyzer) enables prompt quantification of the yields and the subsequent choice of targets suitable for large-scale purification or refolding. The expression and purification protocol is outlined below followed by the protein refolding procedure in Section 2.4.

2.3.1. Small-scale Expression and Purification

Expression vectors are transformed into chemically competent *E. coli.* BL21(DE3)pLysS cells. The proteins are expressed in 1-ml autoinduction media (*16*) in 96-deep-well plates at 30 ^{0}C and 18 ^{0}C. Growth is monitored at 600 nm to determine the optimum incubation time (~20 h at 30 ^{0}C, ~70 h at 18 ^{0}C). The purification is automated on a Biomek 2000 using the Automated MagneHis^{TM} Protein Purification System Protocol #EP011 (Promega). The cell pellets are resuspended, then lysed using detergent lysis buffer (FastBreak^{TM} Cell Lysis Reagent, Promega), and MagneHis^{TM} Ni-Particles (Promega) are employed to initiate binding of the His-tagged recombinant target protein. A MagneBot 96 Magnetic Device (Promega) is used to allow the Ni-Particles to be captured by the magnet. Finally, the proteins are eluted from the resin in 100 µL of 25 mM Hepes (pH 7.4), 150 mM NaCl, 250 mM imidazole. The protein samples are analyzed on Caliper 96 Bioanalyzer to assess the size, purity and yield. Targets with accurate size are ranked according to the yield and selected for large-scale expression and purification.

2.3.2. Large-scale Expression and Purification

Target proteins selected from the small-scale expression trials are produced in large scale (2-4 litres of autoinduction media). The cultures are grown in the conditions that yielded the highest amount of purified protein in small-scale expression. Following affinity chromatography an additional step of size exclusion chromatography (SEC) is

used to purify the proteins to homogeneity. All purification steps are performed at $4\,^{0}$C. A TALON cobalt affinity resin (Scientifix) or the nickel-based affinity column, HisTrap^{TM}FF (GE Healthcare) is employed in the affinity step (the wash and elution buffers are 100 mM Hepes (pH 7.4) and 150 mM NaCl, containing 20 mM or 300 mM imidazole, respectively), and the proteins are then loaded onto the SEC column S200 HiLoad^{TM} 16/60 Superdex^{TM} (GE Healthcare). The protein collected from SEC in 100 mM Hepes (pH 7.4), 150 mM NaCl is concentrated using Amicon Ultra Centrifugal Filter Devices (Millipore) typically to ~20 mg/ml for protein characterization and crystallization.

2.4. Protein Refolding

Insoluble protein expression may be the biggest bottleneck limiting structure genomics initiatives (*17*). While 20-60% of proteins expressed in *E. coli* result in insoluble inclusion bodies (*18-20*), many of these proteins may be amenable to refolding. We have developed a matrix-assisted refolding approach, in which correctly folded proteins are distinguished from misfolded proteins by their elution from affinity resin (*21*). Proteins that are subjected to refolding while bound to metal affinity resin are often resistant to elution by imidazole. We hypothesized that misfolded proteins formed hydrophobic interactions with the surface of the resin. This difference in binding properties between folded and misfolded proteins is the basis for separating the two in this assay. Briefly, a chaotrope is used to solubilize inclusion bodies from bacterial fermentation and His-tagged protein is bound to metal affinity

chromatography resin. The chaotrope is removed by washing the resin in a renaturing buffer and correctly folded protein is subsequently eluted using imidazole. SDS-PAGE is used to compare the quantity of protein in the soluble fraction with that remaining on the resin. This represents the measure of refolding efficiency. The assay is amenable to automation on a liquid handling workstation. The details of this procedure are presented in a separate chapter in this volume (Cowieson et al., A medium or high-throughput protein refolding assay).

2.5. Protein Characterization

SDS-PAGE, size exclusion chromatography, mass spectrometry and circular dichroism (CD) spectroscopy are used to characterize the proteins after purification. Samples of each step of large-scale expression and purification are analyzed by SDS-PAGE to provide a qualitative estimate of purity and reveal any proteolytic degradation or large disparities in protein size. Calibrated size exclusion chromatography gives an estimate of the oligomerization state and the presence of aggregation. Analysis by mass spectrometry on a Voyager DE STR MALDI-TOF or Applied Biosystems QSTAR Pulsar mass spectrometers is used to examine the exact molecular weights of the purified proteins. Finally, proteins are analyzed by CD spectroscopy. We have used both a conventional CD instrument (J-810, Jasco Corporation) and synchrotron radiation CD (SRCD; Daresbury Synchrotron Laboratory, UK). SRCD is particularly suitable as a high-throughput method, because there are fewer limitations on buffer components and data collection is faster (*22*). CD

spectra yield information on protein secondary structure, and are therefore particularly useful for identifying proteins with large proportions of unstructured regions that are unlikely to yield useful high-resolution structural information.

2.6. NMR Spectroscopy

For a subset of proteins with a molecular weight less than 20 kDa, NMR spectroscopy is also used to assess their suitability for structure determination. Uniformly ^{15}N-labelled proteins are overexpressed in *E. coli* BL21(DE3) cells using modified auto-induction media containing ^{15}NH$_4$Cl (2.5 g/L) as the sole nitrogen source. Proteins are then purified using the strategy described above. Samples for NMR screening contain ~0.3 mM protein in 50 mM sodium phosphate buffer (pH 7 or pH 4), 150 mM NaCl and 1 mM DTT in H$_2$O/D$_2$O (9:1). 1D ^{1}H and 2D ^{1}H-^{15}N HSQC NMR spectra are acquired at 25 ºC on a Bruker Avance 600 MHz spectrometer equipped with a z-shielded gradient triple resonance probe, and analyzed using NMRPipe/NMRDraw (*23*). NMR spectral quality and feasibility of three-dimensional structure determination is assessed based on spectral dispersion, line widths, and number of resolved peaks observed compared to the number expected from the amino acid sequence.

2.7. Crystallography

2.7.1. Protein Concentration Optimization

The optimal protein concentration for crystallization screens is determined by

setting up a hanging drop vapor diffusion experiment under two conditions: 2.0 M ammonium sulfate, 0.1 M Tris-HCl (pH 8.5) (Hampton Crystal Screen condition 4) and 30% PEG 4000, 0.1 M Tris-HCl (pH 8.5), 0.2 M magnesium chloride (Hampton Crystal Screen condition 6) at different protein concentrations. The most suitable protein concentration is evaluated after 12 hours incubation at 18 ^{0}C as the one yielding light precipitation in at least one of these conditions.

2.7.2. Crystallization

The vapour diffusion technique (hanging and sitting drop experiments) is employed to screen the proteins using commercial (Hampton Research, Emerald Biostructures, Jena Bioscience, Molecular Dimensions) and in-house screens (*24, 25*). The Biomek 2000 robot (Beckman Coulter) is used to prepare the reservoir solutions in the trays, which are stored at 4 ^{0}C. Protein crystallization droplets are set up in 96-well plates with a Mosquito nanolitre-dispensing robot (TTP Labtech) and placed at two temperatures (18 ^{0}C and 4 ^{0}C). The experiments are monitored using a DeCode Genetics Crystal Monitor. Protein crystals identified are then optimized by setting up focused grid screens (*26*).

2.7.3. Structure determination

The X-ray diffraction quality of protein crystals is assessed using an in-house Rigaku FR-E rotating anode generator (Rigaku/MSC, Texas, USA) with a RaxisIV++ image plate detector. Crystals of native protein are flash-cooled in a nitrogen gas stream

at ~100 K after soaking in a suitable cryoprotectant.

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-labeled protein crystals for use in MAD phasing (*27*) at a synchrotron. Standard crystallographic packages such as HKL2000 (*28*), CrystalClear (Rigaku), SOLVE (*29*), Arp/Warp (*30*), the CCP4 package (*31*) and Coot (*32*) are used to process the data, obtain phase information, build the model, refine and visualize the structures.

2.8. Cross-linking

Chemical cross-linking is employed to obtain structural information on proteins for which structure determination by x-ray diffraction and NMR are not successful. The chemical cross-linkers BS^3 (bis(sulfosuccinimidyl) suberate; Sigma, S5799) and DTSSP (dithiobis(sulfosuccinimidylpropionate; Pierce, 21578) are used to obtain distance constraints. This technique allows high-throughput low-resolution structure determination, particularly when some prior structure information is available (e.g. multi-domain proteins with known domain structures). The protein of interest is allowed to react with the cross-linker, and after quenching the resulting cross-linked and non-crosslinked controls are separated by SDS-PAGE. The band of interest is excised and used for in-gel digestion with trypsin. The masses of the resulting peptides are analyzed by mass spectrometry with the help of in-house software to identify cross-linked products. Based on the identified distance constraints, models are built

using docking and modeling techniques. Details of this technique are given in a separate chapter in this volume (Mouradov et al., Structure determination using a combination of cross-linking, mass spectrometry and molecular modeling).

2.9. Functional Characterization

Our structural studies are complemented by cell biology experiments to further characterize protein function. The DNA encoding the protein is subcloned into a mammalian expression vector such as pDEST21 and transfected into the RAW264.7 murine macrophage cell line. Two types of experiments are typically carried out in the transfected cells in the first instance: subcellular localization and effect of overexpression on macrophage function. Localization experiments are carried out by a combination of confocal immunofluorescence microscopy and subcellular fractionation, followed by western blotting. The transfected protein is detected using a V5 tag expressed at the C-terminus. The functional consequences of overexpression of the protein in macrophages are assessed by monitoring the proliferative and cytokine responses to a range of stimuli such as LPS and CSF-1 (colony-stimulating factor-1), using real-time PCR and ELISA assays to detect inflammatory genes and proteins, respectively. These experiments provide information on the cellular functions of the proteins, complementing structural data that usually shed light on the molecular function, as well as providing functional data for proteins that fail to yield structural data.

3. Progress of the Pipeline

3.1. Statistics

So far, we have processed 318 macrophage proteins, of which 220 have been successfully cloned and 52 expressed in a soluble form in *E. coli,* and entered crystallization trials. Examples of structures resulting from this pipeline include latexin (1.8 Å resolution) (*33*) and long chain acyl-CoA thioesterase (2.4 Å resolution) (**Fig. 2**).

3.2. Functions Revealed

Latexin is the only known mammalian carboxypeptidase inhibitor. We have shown that latexin is expressed constitutively at high basal levels in mouse macrophages and can be further upregulated by stimulation of the cells with growth factor or proinflammatory stimuli (*33*). The crystal structure of latexin (**Fig. 2**) unexpectedly revealed structural similarities with the cysteine protease inhibitor cystatin (*33*). Together, the data suggest a role for latexin in the regulation of proteolysis during inflammation.

Acyl-CoA thioesterases (ACOTs) are a family of enzymes that are conserved through evolution from bacteria to mammals. These proteins catalyze the hydrolysis of acyl-CoA moieties to the respective fatty acid constituents and coenzyme A (*34*). Long-chain acyl-CoAs are intermediates in lipid metabolism and regulators of cellular processes including ion transport, vesicle trafficking, protein phosphorylation and gene expression (*34-36*). Mouse ACOT7 contains two thioesterase domains in tandem. We have determined the crystal structures of each domain separately, and modeled the

full-length protein using distance constraints based on chemical cross-linking (**Fig. 2**). The structure explains the requirement of the two domains for enzymatic activity and the structural basis for long-chain acyl-CoA specificity.

Acknowledgments

References

1.   Burley, S. K. (2000) An overview of structural genomics. *Nat. Struct. Biol.* **7 Suppl,** 932-934.

2.   Chandonia, J. M., and Brenner, S. E. (2006) The impact of structural genomics: expectations and outcomes. *Science* **311,** 347-351.

3.   Gordon, S., Crocker, P. R., Morris, L., Lee, S. H., Perry, V. H., and Hume, D. A. (1986) Localization and function of tissue macrophages. *Ciba Found. Symp.* **118,** 54-67.

4.   Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. (1999) Phylogenetic perspectives in innate immunity. *Science* **284,** 1313-1318.

5.   Mammen, E. F. (2000) Disseminated intravascular coagulation (DIC). *Clin. Lab. Sci.* **13,** 239-245.

6.   Morley, J. E., Thomas, D. R., and Wilson, M. M. (2006) Cachexia: pathophysiology and clinical relevance. *Am. J. Clin. Nutr.* **83,** 735-743.

7.   Duffield, J. S. (2003) The inflammatory macrophage: a story of Jekyll and Hyde. *Clin. Sci. (Lond.)* **104,** 27-38.

8.   Puri, M., Robin, G., Cowieson, N., Forwood, J. K., Listwan, P., Hu, S. H., Guncar, G., Huber, T., Kellie, S., Hume, D. A., Kobe, B., and Martin, J. L. (2006) Focusing in on structural genomics: the University of Queensland structural biology pipeline. *Biomol. Eng.* **23,** 281-289.

9.   Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H.,

Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28,** 235-242.

10.   Fleming, K., Kelley, L. A., Islam, S. A., MacCallum, R. M., Muller, A., Pazos, F., and Sternberg, M. J. (2006) The proteome: structure, function and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361,** 441-451.

11.   Shi J, B. T., Mizuguchi K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310,** 243-257.

12.   Chen, Y., Yu, P., Luo, J., and Jiang, Y. (2003) Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm. Genome* **14,** 859-865.

13.   Bono, H., Kasukawa, T., Furuno, M., Hayashizaki, Y., and Okazaki, Y. (2002) FANTOM DB: database of Functional Annotation of RIKEN Mouse cDNA Clones. *Nucleic Acids Res.* **30,** 116-118.

14.   Yokoyama, S. (2003) Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7,** 39-43.

15.   Listwan, P., Cowieson, N., Kurz, M., Hume, D. A., Martin, J. L., and Kobe, B. (2005) Modification of recombinatorial cloning for small affinity tag fusion protein construct generation. *Anal. Biochem* **346,** 327-329.

16.   Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207-234.

17.   Gilbert, M., and Albala, J. S. (2002) Accelerating code to function: sizing up the protein production line. *Curr. Opin. Chem. Biol.* **6,** 102-105.

18. Moy, S., Dieckman, L., Schiffer, M., Maltsev, N., Yu, G. X., and Collart, F. R. (2004) Genome-scale expression of proteins from Bacillus subtilis. *J. Struct. Funct. Genomics* **5,** 103-109.

19. Huang, R. Y., Boulton, S. J., Vidal, M., Almo, S. C., Bresnick, A. R., and Chance, M. R. (2003) High-throughput expression, purification, and characterization of recombinant *Caenorhabditis elegans* proteins. *Biochem. Biophys. Res. Commun.* **307,** 928-934.

20. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C. H., and Edwards, A. M. (2000) Structural proteomics: prospects for high throughput sample preparation. *Prog. Biophys. Mol. Biol.* **73,** 339-345.

21. Cowieson, N. P., Wensley, B., Listwan, P., Hume, D. A., Kobe, B., and Martin, J. L. (2006) An automatable screen for the rapid identification of proteins amenable to refolding. *Proteomics* **6,** 1750-1757.

22. Miles, A. J., and Wallace, B. A. (2006) Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chem. Soc. Rev.* **35,** 39-51.

23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 227-293.

24. Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C., and 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.

25. Majeed, S., Ofek, G., Belachew, A., Huang, C. C., Zhou, T., and Kwong, P. D. (2003) Enhancing protein crystallization through precipitant synergy. *Structure* **11,** 1061-1070.

26. Senger, A. B., and Mueser, T. C. (2005) A method for the rapid preparation of custom grid-screen crystallization trays using standardized pipetting maps is presented. *J. Appl. Cryst.* **38,** 847-850.

27. Hendrickson, W. (1999) Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron. Radiat.* **6,** 845-851.

28. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307-326.

29. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55,** 849-861.

30. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374,** 229-244.

31. CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760-763.

32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126-2132.

33. Aagaard, A., Listwan, P., Cowieson, N., Huber, T., Ravasi, T., Wells, C. A., Flanagan, J. U., Kellie, S., Hume, D. A., Kobe, B., and 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.

34. Hunt, M. C., and Alexson, S. E. (2002) The role Acyl-CoA thioesterases play in mediating intracellular lipid metabolism. *Prog. Lipid Res.* **41,** 99-130.

35. Faergeman, N. J., and Knudsen, J. (1997) Role of long-chain fatty acyl-CoA esters in the regulation of metabolism and in cell signalling. *Biochem. J.* **323 (Pt 1),** 1-12.

36. Yamada, J. (2005) Long-chain acyl-CoA hydrolase in the brain *Amino Acids* **28,** 273-278.

Figure Legends

Fig. 1. A schematic diagram describing the methodologies applied in the pipeline.

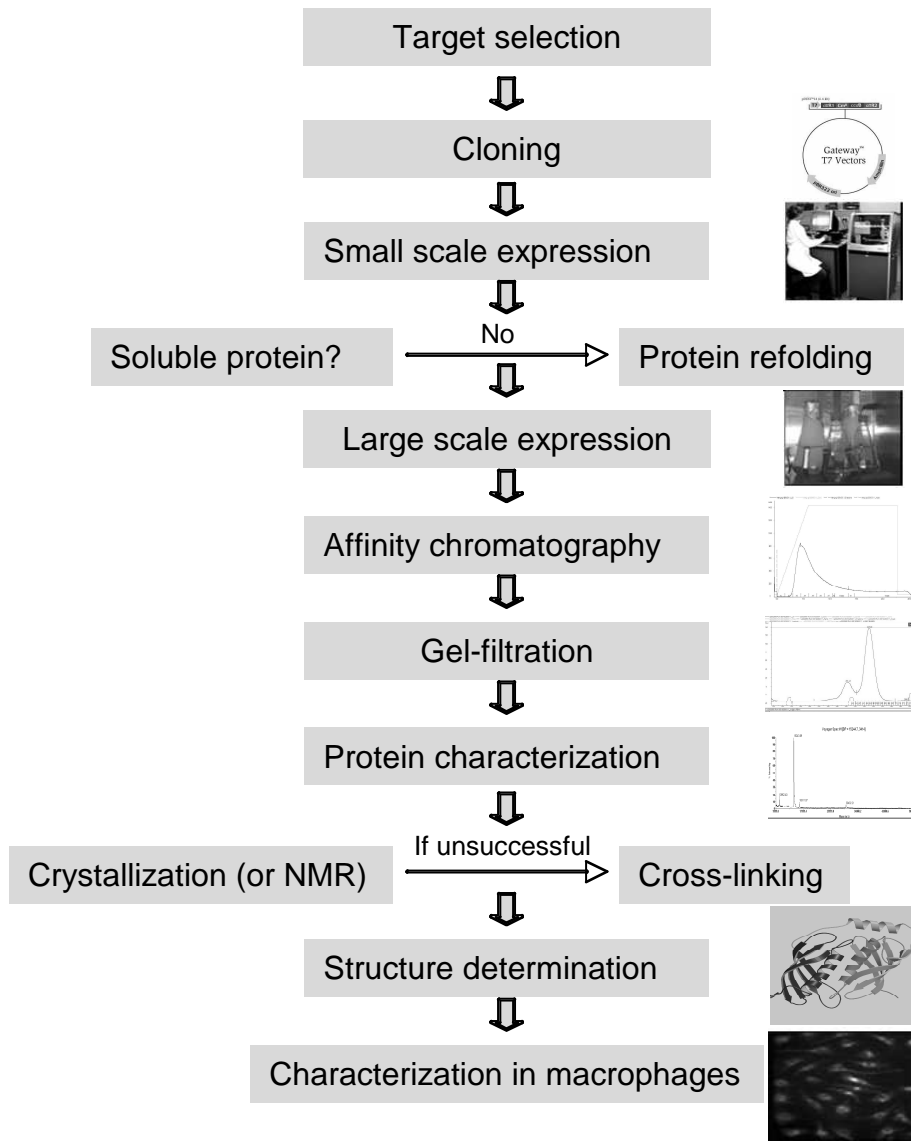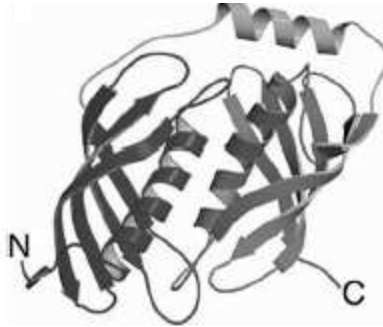Fig. 2. Structures of (**A**) latexin (*33*) and (**B**) ACOT7.

Fig. 1

Target selection
⇩
Cloning
⇩
Small scale expression
⇩
Soluble protein? —No→ Protein refolding
⇩
Large scale expression
⇩
Affinity chromatography
⇩
Gel-filtration
⇩
Protein characterization
⇩
Crystallization (or NMR) —If unsuccessful→ Cross-linking
⇩
Structure determination
⇩
Characterization in macrophages

Fig. 2

(A)

(B)