

Protein Production and Molecular Tags

101

Towards High Throughput Selection of Binding Ligands: Using Flow Cytometry

Peter Pavlik, Milan Ovecka, Nileena Velappan, and Andrew Bradbury* (amb@lanl.gov)

Los Alamos National Laboratory, Los Alamos, NM

Phage display libraries represent a relatively easy way to generate binding ligands against a vast number of different targets. Although in principle, phage display selection should be amenable to automation, this has not yet been described and present selection protocols are far from high throughput. We have examined the selection process in a systematic approach and attempted to automate each individual step. Selection is carried out in the microtiter format using 24 targets as the individual selection lot size. Output is plated onto large assay trays, and a program to pick colonies in specific orders corresponding to the selection arrangement has been developed for the Qbot picking robot. This arrays clones according to the antigen they were selected against, and allows subsequent high density analysis using high density dot blots (up to 13,000 clones in the footprint of a microtiter plate). Although it proved possible to analyze such large numbers of clones using HD dot blots, it proved extremely difficult to quantify and digitize binding information. Furthermore, variations in expression levels led to non-specific binding artifacts – well-expressed clones gave binding signals which were often non-specific. Although this could be eliminated by also arraying clones on non-specific target filters, the integration of the information from the two filter types proved extremely difficult to quantify and analyze.

As an alternative we have examined the use of flow cytometry. In a model system, using bead based Luminex type assays, we have been able to carry out multiplex analyses, in which the reactivity of individual antibody clones for numerous different target parameters can be examined simultaneously. The analysis of each individual clone can be carried out in approximately 60 seconds, and all information is easily exportable to LIMS type systems, as well as being readily analyzed. In first experiments we were able to obtain information on the binding of individual antibody clones to specific and non-specific targets, as well as obtaining indications of expression levels. This was carried out by coupling different colored beads with: 1) specific antigen; 2) irrelevant antigen; 3) anti-tag antibodies (to determine expression level). Preliminary experiments with true selections will be presented.

102

Efficient Chemical Methods for the Total Synthesis of Small Proteins: The First Crystallographic Structure of a Protein Diastereomer, [D-Gln35]-ubiquitin

Duhee Bang^{1*} (duhee@uchicago.edu), George I. Makhatadze², and Stephen B. Kent¹ (skent@uchicago.edu)

¹University of Chicago, Chicago, IL and ²Pennsylvania State University, Hershey, PA

Our goal is to understand the molecular basis of the biological function of proteins using chemistry. To that end, we are developing more practical methods for the total chemical synthesis of proteins by the ligation of unprotected peptide building blocks. Our recent progress includes a *'one pot' total synthesis of proteins* [1], a *tag assisted chemical protein synthesis* [2], and an *efficient approach to the total synthesis of cysteine free proteins* [3]. Here we will describe a case study to understand the molecular basis of protein stability using a small model protein, ubiquitin.

We questioned how a natural protein would adopt a D-amino acid into its overall architecture. To definitively explore changes of local and global conformations of proteins by D-amino acid incorporation, we decided to crystallize a D-amino acid incorporated protein molecule. We used a cysteine-free globular protein, ubiquitin (76 amino acids) to chemically engineer a protein α -helix. In particular, we targeted a glycine residue of the C-cap region of a protein α -helix. The conformational space of the Gly residue is only allowed for left handed α -helix and D-amino acid residues. We present (i) an efficient strategy for total chemical syntheses of ubiquitins, (ii) direct observation of the conservation of L-configuration from protein Raney-Ni reduction (Cys \rightarrow Ala), (iii) the highest resolution {1.5 Å} crystal structure for known ubiquitin wild type, and (iv) high-resolution {1.3 Å} crystal structure of a ubiquitin diastereomer, UBQ[D-Gln35].

Our syntheses made use of the native chemical ligation [4] of three unprotected peptide segments; (1-27)-thioester; (Thz28-45)-thioester; and (Cys46-76). Native Ala28 and Ala46 were replaced by Cys28 and Cys46 to enable the use of native chemical ligation at Cys. A desulfurization reaction of the product polypeptide using Raney nickel [5] was performed to convert the cysteines to alanines. Syntheses of analogue ubiquitins were performed in the same manner. The synthetic ubiquitins were crystallized and X-ray diffraction data was col-

Figure 1. Overall fold highlighted with D-Gln35 mutation (left), 1.3 Å resolution map near the mutation (-- Glu34 – D-Gln35 – Ile36 --)

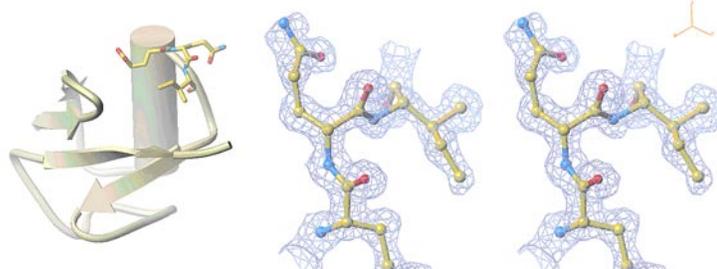
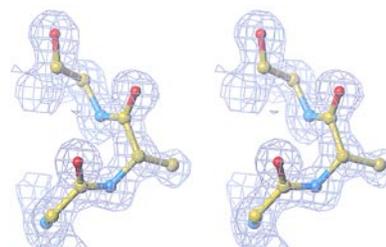


Figure 2. Conservation of L-configuration after Cys \rightarrow Ala desulfurization reaction



lected using the advanced photon source at ANL. We are currently refining ubiquitin wild-type and its diastereomer structures (partially refined structures are shown in Figure). The structures and ongoing efforts for the understanding of protein stability will highlight the power of the total chemical synthesis of proteins.

References

1. Bang, D. & Kent, S. B. (2004) *Angew. Chem. Int. Ed.* 43, 2534-2538
2. Bang, D. & Kent, S. B. *PNAS* Accepted.
3. Bang, D. G. I. Makhataдзе, S. B. Kent Manuscript in preparation
4. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* 1994, 266, 776-779.
5. L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* 2001, 123, 526-533.

103

Development and Application of Multipurpose Affinity Probes to Isolate Intact Protein Complexes Associated with Metal Reduction from *Shewanella oneidensis* MR-1

Liang Shi*, Thomas C. Squier* (thomas.squier@pnl.gov), M. Uljana Mayer*, Haishi Cao, Baowei Chen, Yuri A. Gorby, David F. Lowry, Jeff Mclean, Seema Verma, and Ping Yan

Pacific Northwest National Laboratory, Richland, WA

Our long-term goal is to develop high-throughput methods for the rapid isolation of intact protein complexes and validation of these complexes in living cells. This methodology utilizes a small genetically encoded protein tag with an 8 amino-acid sequence containing a tetracysteine motif, which can be captured using affinity reagents or labeled with fluorescent dyes *in vivo* to permit cellular validation of protein complexes. An important advantage of this strategy is that a single small and nonperturbing tag can be sequentially used to 1) isolate the intact protein complex for identification and structural analysis and 2) visualization of the location and abundance of the protein complex within cells (Chen et al., 2004; Mayer et al., 2005). Furthermore, by varying the architecture of the affinity reagent, multiple colors and photoactivatable cross-linkers can be incorporated into the design strategy to permit measurements of binding interactions within cells and the stabilization of transient interactions associated with signaling complexes.

Proof of principle for this approach has been achieved through the isolation of two protein complexes (i.e., RNA polymerase and the metal reductase complex) from *S. oneidensis* MR-1, whose metabolism is important in understanding both microbial energy production and environmental remediation. However, these strategies will be applicable to a wide range of microorganisms and will permit the identification of environmental conditions that affect the expression of critical proteins required for the formation of transient protein complexes that facilitate bacterial growth. Our hypothesis is that identifying dynamic changes in these adaptive protein complexes will provide important insights into the metabolic regulatory strategies used by these organisms to adapt to environmental changes.

RNA polymerase is a well studied system, which contains a core complex containing RNA polymerase $\alpha_2\beta\beta'$ subunits as well as regulatory proteins associated with the differential regulation of transcription. Following the expression of a tagged subunit of the RNA polymerase core complex in *S. oneidensis* MR-1, we have isolated this complex using the synthesized affinity reagent immobilized on a glass bead (Mayer et al., 2005). A critical advantage of this method is the ability to release the intact complex using a mild, one-step procedure with a competing dithiol. In addition to the identification of the core subunit complex, additional regulatory factors were identified, including the universal stress protein.

To investigate whether the current approach will also permit the identification of membrane protein complexes, we have tagged genes identified by the *Shewanella* Federation to be involved in metal reduction, and isolated members of this important protein complex. In one experiment, the metal reductase MtrC [a decaheme c-type cytochrome tentatively identified as an outer membrane protein whose activity is required for efficient reduction of Mn(IV) and Fe(III)] was genetically tagged and used to isolate two high-affinity heme-containing binding subunits in the complex that were not previously identified (i.e., OmcA and MtrA). The isolated MtrC complex maintained its activity to reduce Fe(III). This Fe(III)-reducing activity was enhanced by addition of purified MtrA, even though purified MtrA itself possessed no Fe(III)-reducing activity. Validation of this protein complex was achieved following purification of the individual proteins, using the affinity reagent dyes to measure the structural interactions between these proteins.

In summary, these multiuse affinity reagents have the advantage over other affinity tags for the high-throughput identification of protein binding partners, in that 1) the small tag can be rapidly cloned into the protein of interest and leads to minimal perturbations of binding interactions, 2) proteins are not denatured following elution permitting purification of the intact complex that can thus be further validated and studied by structural methods, and 3) the affinity reagents are cell permeable and can be used for imaging measurements to monitor protein-protein interactions in live cells.

References

1. Chen, B., M. U. Mayer, L. M. Markille, D. L. Stenoien, and T. C. Squier (2004) *Dynamic motion of helix A in the amino-terminal domain of calmodulin is stabilized upon calcium activation*. Biochemistry, in press.
2. Mayer, M. U., L. Shi, and T. C. Squier (2005) *One-step, non-denaturing isolation of an RNA Polymerase core enzyme complex using an improved multi-affinity probe resin*. J. Am. Chem. Soc., submitted.

104

A Combined Informatics and Experimental Strategy for Improving Protein Expression

Osnat Herzberg, John Moulton* (moulton@umbi.umd.edu), Fred Schwarz, and Harold Smith

Center for Advanced Research in Biotechnology, Rockville, MD

Improved success rates for recombinant protein expression are critical to many aspects of the Genomics:GTL program.

The project is focused on determining which factors determine whether or not soluble protein is produced in *E. coli*. We are investigating the role of protein family size, native expression level, protein stability and folding rate, and the response of the host cell to expression. The outcome of the project will be a set of informatics and experimental strategies. Informatics will provide a synopsis of all relevant information for a protein, ranking alternative strategies for optimization of production. Possible new strategies include the use of reporter fusions to monitor up or down regulation of known and newly discovered cell cellular response proteins; utilization of cellular response to control cell growth; protocols for the design of mutants to improve expression; inhibition of specific proteins shown to affect outcome; and co-expression of proteins found to enhance outcome.

In the first nine months of the project, a first scan of the different factors potentially affecting expression outcome has begun. 10 proteins with representative expression properties have been prepared and submitted for micro-calorimetric investigation of their stability properties. Results are currently available for five. Messenger RNA content in *E. coli* has been investigated under conditions of over-expression of 10 proteins, five of which produce high amounts of soluble protein, and five which produce substantial amounts of insoluble material. Increased transcription of a number of genes, several of which have been implicated in stress response or protein folding, correlates strongly with the solubility status of the recombinant protein. Efforts are underway to increase or disrupt expression of those genes prior to protein induction and ascertain the effect on protein solubility.

This project is supported by Genomics:GTL award DE-FG02-04ER63787

High-Throughput Production and Analyses of Purified Proteins

F. William Studier^{1*} (studier@bnl.gov), John C. Sutherland^{1,2}, Lisa M. Miller¹, and Lin Yang¹

¹Brookhaven National Laboratory, Upton, NY and ²East Carolina University, Greenville, NC

This work is aimed at improving the efficiency of high-throughput protein production from cloned coding sequences and high-throughput biophysical characterization of the proteins obtained. Vectors and protocols for high-throughput production of proteins in the T7 expression system in *Escherichia coli* are being developed and tested by expressing and purifying proteins of *Ralstonia metallidurans*, a bacterium that tolerates high concentrations of heavy metals and has potential for bioremediation. Auto-induction allows many clones in parallel to be screened for expression and solubility simply by inoculating the cultures and growing to saturation, without the need to monitor culture growth and add inducer to each culture at the proper time. Auto-induction protocols have been developed for both BL21(DE3), in which lactose induces T7 RNA polymerase and unblocks the T7lac promoter, and BL21-AI, in which arabinose induces T7 RNA polymerase and lactose unblocks the T7lac promoter. Progress is also being made in developing new vectors that allow inducible expression of proteins that are highly toxic to the host cells. The first set of 96 *Ralstonia metallidurans* proteins is being cloned and will be tested for expression and solubility in the new vectors.

Proteins produced from clones are often improperly folded or insoluble. Many such proteins can be solubilized and properly folded, whereas others appear soluble but remain aggregated or improperly folded. As high-throughput production of purified proteins becomes implemented in GTL projects and facilities, reliable analyses of the state of purified proteins will become increasingly important for quality assurance and to contribute functional information. Beam lines at the National Synchrotron Light Source analyze proteins by small-angle X-ray scattering (SAXS) to determine size and shape, X-ray fluorescence microprobe to identify bound metals, and Fourier transform infrared (FTIR), UV circular dichroism (CD), linear dichroism (LD) and fluorescence spectroscopy to assess secondary structure and possible intermolecular orientation. A flexible liquid-handling system for automated loading of samples from 96-well plates for analysis at each of these stations has been built and is being implemented with purified proteins. When fully functional, the system will be capable of high-throughput analyses of size, shape, secondary structure and metal content of purified proteins, which will complement analyses such as gel filtration, mass spectrometry and NMR.

This project is supported by the Office of Biological and Environmental Research of the Department of Energy. Work on auto-induction and vector development also receives support from the Protein Structure Initiative of the National Institute of General Medical Sciences of NIH, as part of the New York Structural Genomics Research Consortium.

Development of Genome-Scale Expression Methods

Sarah Fey, Elizabeth Landorf, Yuri Londer, Terese Peppler, and Frank Collart* (fcollart@anl.gov)

Argonne National Laboratory, Argonne, IL

Protein diversity suggests multiple expression strategies will be required to insure production of the highest possible proportion of cellular proteins. We are developing novel cellular and cell-free technologies to optimize the expression of cytoplasmic, periplasmic/secreted proteins and protein domains. These molecular tools contain elements that enable localization to appropriate cellular or extracellular compartments coupled with regulatory elements to permit control and coordination of protein expression. They also incorporate specific fusion components that promote protein stability and solubility or that facilitate detection, purification and/or protein characterization. Specific focus areas for *in vivo* expression in *E. coli* are as follows:

- Evaluation of various fusion tag cassettes to maximize the generation of soluble proteins or protein domains for downstream analysis.
- Development of a periplasmic expression system compatible with current standard high throughput cytoplasmic cloning strategies. This process has been implemented in a 96-well plate format and is being used for analysis of expression and solubility for *Shewanella* and *Geobacter* proteins directed to the cytoplasm or periplasm.
- Evaluation of a domain-based cloning and expression strategy for simple architecture membrane proteins. Proteins were analyzed for periplasmic signal sequences by sequence analysis using the signalP algorithm (1, 2) or for transmembrane regions by application of the TMHMM program (3). This approach is being applied to a set of two-component sensor and methyl accepting chemotaxis proteins from *Shewanella* and *Geobacter*.
- Generation of several constructs intended to facilitate cloning and expression of genes coding for c-type cytochromes. These constructs and host strains are being evaluated for implementation in a high throughput environment.

Our studies indicate a large fraction of proteins of highest interest are difficult to express using standard expression systems. Our novel expression methods extend the boundaries of current high throughput technology and provide strategies for expression of challenging proteins that can be implemented by the general scientific community. We are attempting to optimize distribution of purified proteins or clones that express soluble protein for characterization in detail and elucidation of biological function.

1. Nielsen, H., and Krogh, A. (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* 6, 122-30.
2. Nielsen, H., Brunak, S., and von Heijne, G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* 12, 3-9.
3. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6, 175-82.

107

Plate-Based Methods for Expression of Cytoplasmic Proteins from *Shewanella oneidensis*

Elizabeth Landorf¹, Terese Peppler¹, Sarah Fey¹, Alexander Iakounine³, Eugene Kolker², and Frank Collart^{1*} (fcollart@anl.gov)

¹Argonne National Laboratory, Argonne, IL; ²BIATECH, Bothell, WA ; and ³University of Toronto, Toronto, Canada

Transcriptomic and proteomic analyses by the *Shewanella* Federation suggest there are a large number (>500) of ORFs identified as hypothetical genes that are expressed in *S. oneidensis* cells both as mRNAs and proteins. We have expressed many of these “hypothetical” proteins using plate base methods to identify strategies appropriate for genome scale analysis. Expression was confirmed for 61% of the hypothetical target set by denaturing gel electrophoresis. This represents a lower bound since not all targets were analyzed for expression. For 21% of the targets, an obvious fusion protein was not observed after gel analysis. Soluble proteins were grouped into several relative categories based on qualitative assessment of band intensity (Commassie blue staining) after denaturing gel electrophoresis. Approximately two-thirds of clones expressing fusion products generated a protein that was soluble at an analytical scale. The solubility assignments at this analytical scale show a general correlation to the yield of soluble protein obtained in preparative scale cultures suggesting most of these samples would be suitable for functional interrogation. These expression results are consistent with the genomic expression analysis indicating the assigned ORFs produce an expression product and provide an opportunity for characterization of the samples expressed in soluble form. This analysis is essential as it will provide a protein resource for the larger scientific community that enables verification of functional hypothesis derived by the various functional prediction methods. Where possible, we are attempting to experimentally validate the functional assignment. In collaboration with the University of Toronto, we have submitted a number of purified protein samples for characterization using functional screen matrix for enzymatic activity. Functional assignments were validated for several proteins in the initial test set and new samples are being prepared for submission.

108

Generating scFv and Protein Scaffolds to Protein Targets

Brian K. Kay* (bkay@anl.gov), Michael Scholle, Ushma Kriplani, John Kehoe, and Frank Collart

Argonne National Laboratory, Argonne, IL

We have recently used a library of human single-chain Fragments of variable regions (scFv) to generate antibodies to a collection of *Geobacter*, *Shewanella*, and *Rhodobacter* proteins. The proteins were overexpressed in *E. coli*, chemically biotinylated, and captured on streptavidin coated magnetic beads for screening a large library of human scFv molecules, which were displayed on the surface of bacteriophage M13. After three rounds of affinity selection with liquid handling robotic workstations, the binding of isolates is confirmed by inserts are transferred *en masse* via ligation independent cloning into an alkaline phosphatase (AP) fusion vector for enzyme-linked binding assays (ELBA). The scFv-AP fusions can be used to check the specificity of the antibodies and probe western blots of cell lysates. In addition, soluble forms of the scFvs can be use to pull-down protein complexes from cell lysates and identify cellular machines. Finally, to overcome one potential limitation (i.e., presences of disulfides that may not form in reducing environments) of scFvs to perturb target functions intracellularly, we are currently exploring the use of FN3 monobodies and the villin headpiece as antibody-like scaffolds.

109

Cell Free Approaches for Protein Production

Gerald W. Becker^{2*}, Pavel Shiyonov², Yifei Wu², Sarah Fey¹, Elizabeth Landorf¹, Terese Peppler¹, and Frank Collart¹ (fcollart@anl.gov)

¹Argonne National Laboratory, Argonne, IL; and ²Roche Applied Science, Indianapolis, IN

Cell-free protein synthesis is an easily automated screening tool and is also a scalable process for the production of preparative amounts of protein. We developed a series of plate-based methods for protein expression and solubility screening using cell-free technology that can be applied to entire genomes. In a pilot study, we selected 384 ORFs from *Shewanella oneidensis* and amplified the coding regions with KOD polymerase. The same amplification products were used for parallel expression studies using cell-free technology and *Escherichia coli* as an *in vivo* expression system. Comparison of expression/solubility outcomes indicate a general correspondence of these parameters for the two expression strategies but there are a number of targets which were only expressed and/or soluble in one of the systems. These targets are being further evaluated to determine if specific classes of protein are more readily produced using the cell-free technology. These initial studies were extended to include a series of plate-based solubility screens that evaluated the effect of various additives such as detergents, cofactors, chaperones or auxiliary proteins. Application of this approach to a set of 24 samples indicates this is an effective strategy to improve expression/solubility outcomes for many challenging proteins. The solubility enhancement screen has been implemented in a 96-well plate format and the open nature of the cell-free system will facilitate the evaluation of additional additives in the transcription/translation reaction. These preliminary studies illustrate the many advantages of

cell-free expression systems such as the capability to support expression directly from PCR products without cloning, propagation and purification of plasmid DNA. This approach allows for an increase in the throughput capabilities for domain interrogation as well as enabling the screening for the suitability of different tags and fusion partners and the optimization of solubility by high throughput compatible methods.

110

Rapid Synthesis of Peptidic and Peptidomimetic Ligands for High-Throughput Protein Purification and Labeling

Jeffrey B.-H. Tok^{1*} (tok2@llnl.gov), Priscilla Chan¹, David Smithson², Ted Tarasow¹, and Rod Balhorn¹

¹Lawrence Livermore National Laboratory, Livermore, CA and ²University of California, San Francisco, CA

The ability to rapidly generate protein-specific affinity reagents is clearly a desirable capability toward the goal of GTL program. My laboratory has utilized a combination of chemical and biochemical combinatorial approaches to generate such affinity tags against bacterial protein targets. Specifically, we have recently established an extremely versatile and useful approach to synthesize peptide libraries, in which several peptide libraries containing up to 2.5 million unique peptide ligands had been synthesized in ~two weeks. These libraries had subsequently been utilized to successfully identify novel affinity binders against the ganglioside-binding protein domain of the bacterium *Clostridium tetanus*. Calorimetric assays to enable rapid screening of the peptide library have also been demonstrated and have subsequently afforded “lead” sequences to enable a second generation of peptide ligands.

In addition, we have recently successfully applied the peptide affinity tags to purify the proteins through immobilizing the peptide affinity tags onto solid resins. Attempts to render this effort in a high-throughput fashion will be discussed.