Building Biochips: A Protein Production Pipeline

Marianne G. S. de Carvalho-Kavanagh and Joanna S. Albala
Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
Livermore, CA 94550

ABSTRACT

Protein arrays are emerging as a practical format in which to study proteins in high-throughput using many of the same techniques as that of the DNA microarray. The key advantage to array-based methods for protein study is the potential for parallel analysis of thousands of samples in an automated, high-throughput fashion. Building protein arrays capable of this analysis capacity requires a robust expression and purification system capable of generating hundreds to thousands of purified recombinant proteins.

We have developed a method to utilize LLNL-I.M.A.G.E. cDNAs to generate recombinant protein libraries using a baculovirus-insect cell expression system. We have used this strategy to produce proteins for analysis of protein/DNA and protein/protein interactions using protein microarrays in order to understand the complex interactions of proteins involved in homologous recombination and DNA repair. Using protein array techniques, a novel interaction between the DNA repair protein, Rad51B, and histones has been identified.

Keywords: baculovirus, protein production, high-throughput, protein array

1. INTRODUCTION

As the final sequencing efforts of the Human Genome Project conclude, the next frontier in deciphering cellular biochemistry will be the analysis of the human proteome. The analysis of gene products *en masse* will require new strategies. Array-based protein technologies are emerging with the potential of providing parallel functional analysis of hundreds or thousands of proteins simultaneously. Array-based methods are becoming prevalent within proteomics research due to the desire to analyze proteins in an analogous format to that of the DNA microarray (1). Array-based methods provides a platform for protein study that can be automated for high-throughput analysis of protein–protein, protein-DNA, and protein-small molecule interactions to be performed against thousands of proteins simultaneously, providing important functional information for newly identified genes and gene products derived from genome sequence. Our long-term objective is to develop a system capable of expressing and purifying hundreds to thousands of human proteins in an arrayed, addressable format in quantities and purities suitable for high-throughput functional studies. A key element of our approach is the use of the collection of arrayed, human cDNAs from the LLNL-I.M.A.G.E. (Integrated Molecular Analysis of Genome Expression (2)) Consortium.

The I.M.A.G.E. Consortium is the largest public cDNA collection, and in collaboration with the National Institutes of Health Mammalian Gene Collection project, is creating a collection of full-length cDNA clones representing every human and mouse gene. The collection consists of primarily human and mouse cDNA clones; however, it also includes cDNAs from zebrafish, *Fugu*, rat, *Xenopus,* and primate species. Using high-speed robotics, the Consortium has arrayed over eight million individual cDNAs into 384-well plates, and has delivered replica plates to both sequencing centers and distributors worldwide. All available information associated with an individual clone, such as its unique identification number, the library and tissue from which it was derived, and sequence data is publicly available at http://image.llnl.gov.

Our approach to expressing the human proteome has been to use a PCR-based strategy to express human proteins using a baculovirus-based eukaryotic expression system. These proteins are His-tagged which allows for subsequent purification using well-established techniques. Our procedures allow for each component of this process, where applicable, from cDNA to protein, to be automated and optimized using robotics. This expression and purification

Microarrays and Combinatorial Techniques: Design, Fabrication, and Analysis II,
edited by Dan V. Nicolau, Ramesh Raghavachari, Proceedings of SPIE Vol. 5328
(SPIE, Bellingham, WA, 2004) · 1605-7422/04/$15 · doi: 10.1117/12.542378

system has been validated with a small number of cDNAs produced as recombinant proteins in baculovirus in a 96-well format and can be optimized to produce a diverse collection of proteins (3). We have employed our miniaturized protein production strategy for the generation of protein fragments for mapping interactions among domains of critical DNA repair proteins in the homologous recombinational repair pathway (4) as well as for the development of protein interaction microarrays (5).

2. METHODOLOGY

2.1 Cloning of I.M.A.G.E. cDNAs

Genes of interest represented by I.M.A.G.E. cDNA clones were chosen and bacterial colonies were re-arrayed from 384-well storage plates into 96-well format using an robotic arrayer (Norgren Systems; Palo Alto, CA). Most downstream steps in this process can be performed in 96-well microtiter plates and are amenable to implementation using robotic manipulations where applicable using a Tecan Genesis RSP 150 robot (Tecan US Inc., Durham, NC).

Genes were amplified by PCR from I.M.A.G.E. vectors using a gene-specific 5' forward primer containing the rare restriction enzyme site for Asc I. Each reverse PCR primer contained sequences specific for the library vector from which the cDNA was derived followed by the restriction enzyme site for Fse I. The PCR products were purified using a Wizard SV 96 Clean up kit (Promega; Madison, WI). PCR products were digested using the AscI and FseI enzymes (New England Biolabs; Beverly, MA) followed by another round of purification.

2.3 Construction of baculovirus transfer vector and generation of recombinant virus

For the creation of recombinant baculoviruses, the pBacPAK9 baculoviral transfer vector (Clontech, Palo Alto, CA) was modified to include sequence encoding an N-terminal histidine tag followed by exonuclease sites for the rare cutters AscI and FseI that were added between the BglII and PstI site of the multiple cloning site of pBacPAK9, to create pMGHis. For production of recombinant transfer vector, pMGHis was linearized with AscI and FseI and dephosphorylated with shrimp alkaline phosphatase (Fermentas; Hanover, MD). Each of the PCR-amplified and digested cDNAs were ligated into the pMGHis vector in 96-well format and transformed into *E. coli*. Two *E. coli* colonies for each cDNA clone were grown overnight and screened by PCR. The cloned genes of interest in pMGHis (200-300 ng) were co-transfected into *Sf21* insect cells in 96-well format at 4 x 10^4 cells/well with 25 ng linearized baculoviral DNA (Baculogold; Clontech, Palo Alto, CA) and 3 μg/well Superfect transfection reagent (Qiagen; Valencia, CA) and incubated at 27°C to produce recombinant virus.

2.4 High-throughput protein production

Viral amplification in the 96-well format was repeated 2-3 times to increase the viral titer without the need for plaque purification. For protein production, *Sf21* cells were infected with recombinant baculovirus in 96 deep-well plates with a 2 ml capacity (Marsh Biomedical Products, Inc.; Rochester, NY) on a Carousel Magnetic Levitation Stirrer (V & P Scientific, San Diego, CA). The instrument uses metal balls, similar in appearance to ball bearings, for mixing and aeration of the insect cells in suspension. Typically, 48-hours after infection with recombinant virus, the insect cells were pelleted by centrifugation, rinsed with PBS and lysed in 0.2 ml ice cold lysis buffer (50 mM Hepes. 150 mM NaCl, 10% glycerol and 0.5 % v/v NP-40 with added protease inhibitors). The lysates were clarified by centrifugation and supernatants were analyzed for protein concentration, used for immunoprecipitation, or analyzed by SDS-polyacrylamide gel electrophoresis. For purification of the 6x His-tagged fusion proteins, SwellGel nickel-chelated discs (Pierce; Rockford, IL) were used according to the manufacturer's recommendations. The lysate was allowed to interact with the discs for 10 minutes and then the supernatant pulled through the wells via light vacuum. The beads were washed 2-3 times with 0.2 ml per well of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 40 mM imidazole) and then the proteins of interest eluted using wash Buffer containing 250 mM imidazole. The eluted proteins were collected via light vacuum filtration into a 96-well collection plate and an aliquot of soluble protein was analyzed by SDS-polyacrylamide gel electrophoresis.

2.5 Generation of Protein Microarrays

Standard microarray spotting techniques were used to attach proteins to glass slides in a microarray format as previously described (5). Briefly, proteins (0.01 – 10 mg/mL) were resuspended in spotting buffer (50 mM HEPES pH 7.5, 5% glycerol, 50 mM KCl) and arrayed in duplicate on aminopropyl triethoxysilane and/or poly-lysine-coated glass slides (Corning Inc.; Corning, NY) using a robotic arrayer (Norgren Systems; Palo Alto, CA). All information regarding the construction and use of the protein arrays have been deposited in a LLNL microarray-related database (http://bbrp.llnl.gov/microarrays/external/index.html). Spotted arrays were covered with a hybridization chamber (Schleicher & Schuell; Keene, NH) and filled to a total volume of 300 µl with blocking buffer (3% non-fat milk, 1X PBS, 1% Tween 20, 100 µg/mL BSA). The arrays were then incubated for 15 min at 25°C with gentle shaking. Wash buffer (50 mM Tris pH 7.5, 50 mM NaCl, 2 mM DTT, 0.5% NP-40) was applied to the slides three times for 5 min at 25°C. After the final wash, 50-100 ng of target protein in 300 µL of interaction buffer (50 mM Na_2PO_4 pH 8.0, 200 mM NaCl, 5% glycerol) was added to the array for 30 min at 25°C with gentle shaking. The slides were washed with wash buffer three times for 5 min at 25°C. Primary antibody against the appropriate target protein was diluted 1:500 in 400 µL of PBS-1% Tween and incubated for 30 min at 25°C with gentle shaking, followed by three washes. As appropriate, a rhodamine-labeled goat anti-mouse or FITC-labeled anti-rabbit (Santa Cruz Biotechnology; Santa Cruz, CA) secondary antibody was diluted 1:250 in 400 µL of PBS-1% Tween and incubated for 30 min at 25°C with gentle shaking, followed by three 15 min washes with gentle shaking. Scanning and analysis of the arrays were performed on a ScanArray 5000 (488 nm laser for FITC scans and 543 nm laser for Rhodamine) using QuantArray software (Packard BioScience; Billerica, MA).

3. RESULTS

3.1 Miniaturized Co-Immunoprecipitations

The miniaturized baculoviral expression system was used to generate recombinant proteins for each of the Rad51 paralog proteins, Rad51B, Rad51C, Rad51D, Xrcc2 and Xrcc3 (6), (7), (8), (9). These proteins are homologs of the Rad51 protein, a key protein in DNA double-strand break repair (10). There are two different Rad51 paralog protein complexes that have been identified *in vitro a*nd *in vivo,* one containing Rad51C and Xrcc3, the other containing Rad51B, Rad51C, Rad51D and Xrcc2, BCDX2 (11), (12). Using computational modeling, a series of deletion mutants were generated for Rad51B, amino acids 1-75 and 76-350, Rad51C, amino acids 1-79 and 79-376, and Rad51D, amino acids 4-77 and 77-379, for domain mapping analysis to determine the regions of interaction between the proteins in the BCDX2 complex. The recombinant baculoviruses were generated from each mutant construct and co-infections and immunoprecipitations were performed to analyze for protein binding. For example, the N- and C-terminal fragments of Rad51B and Rad51C were cloned into baculoviral transfer vectors and co-expressed in insect cells. A Glu-tagged Rad51C virus (11) was co-infected with recombinant virus containing either full-length Rad51B or the Rad51B fragments, each of which contained an N-terminal His-tag. The resulting immunoprecipitant was tested for protein-protein interactions by SDS-PAGE and western blot analysis. The results from the domain mapping of the various deletion constructs demonstrated that the C-terminus of Rad51C interacts with the Rad51B, Rad51D and Xrcc3 proteins (for more details see (4)).

3.2 Protein Interaction Microarrays

We have also used protein expressed from this miniaturized baculoviral expression system to develop protein microarrays to analyze protein-protein interactions. Using proteins from baculovirus and IVT-expressed GFP fusion proteins (13), protein interaction microarrays were produced on aminopropyl triethoxysilane and/or poly-lysine coated glass slides. These protein interaction microarrays demonstrated that Rad51B has an affinity for free histones, but not histones complexed with DNA ie. nucleosomes. Recently, histones have been shown to be important protein factors in the DNA repair pathway (14). This result is interesting given that Rad51 associates with histone-related proteins

following DNA damage at foci (15). Interestingly, no interaction was observed between Rad51C, the protein complex partner of Rad51B, and histone proteins (5).

4. CONCLUSIONS

The generation of hundreds to thousands of proteins for high-throughput applications remains a challenging undertaking. We have developed a miniaturized, high-throughput baculoviral system for the production of recombinant eukaryotic proteins. An advantage of baculoviral expression is the ability to produce eukaryotic proteins with the appropriate post-translational modifications. Future work will aim to utilize protein microarrays to understand complex subproteomes and to characterize the many post-translational modification of these proteins.

Using this system, we have produced a diverse set of human proteins for application to protein arrays for protein-protein interaction mapping studies as well as for the identification of novel protein interactions. Recent strides have been made in the use of protein arrays for diagnosis of autoimmune disorders (16), (17), (18) as well as for cancer biomarker identification (19). Protein microarrays hold great promise for accelerating basic biology, biomarker identification and drug discovery to further the development of new diagnostics and prognostics for disease.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Christa Prange, Michele Gilbert, Todd Edwards, Matt Coleman, Daniel Yoshikawa, and Kristi Miller for their contributions to these studies. This work was supported by NCI award #CA81019-01 and U.C. BioSTAR grant 02-10238 awarded to J. S. A. and performed under the auspices of the U.S. DOE by the UC-LLNL under Contract No. W-7405-Eng-48.

REFERENCES

1. Albala, J.S. (2001) Array-based proteomics: the latest chip challenge. *Expert Rev Mol Diagn*, **1,** 145-152.
2. Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M.B. (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics*, **33,** 151-152.
3. Albala, J.S., Franke, K., McConnell, I.R., Pak, K.L., Folta, P.A., Rubinfeld, B., Davies, A.H., Lennon, G.G. and Clark, R. (2000) From genes to proteins: high-throughput expression and purification of the human proteome. *J Cell Biochem*, **80,** 187-191.
4. Miller, K.A., Sawicka, D., Barsky, D. and Albala, J.S. (2004) Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res*, **32,** 169-178.
5. Coleman, M.A., Miller, K.A., Beernink, P.T., Yoshikawa, D.M. and Albala, J.S. (2003) Identification of Chromatin-related Protein Interactions Using Protein Microarrays. *Proteomics*, **3,** 2101-2107.
6. Dosanjh, M.K., Collins, D.W., Fan, W., Lennon, G.G., Albala, J.S., Shen, Z. and Schild, D. (1998) Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. *Nucleic Acids Res*, **26,** 1179-1184.
7. Albala, J.S., Thelen, M.P., Prange, C., Fan, W., Christensen, M., Thompson, L.H. and Lennon, G.G. (1997) Identification of a novel human RAD51 homolog, RAD51B. *Genomics*, **46,** 476-479.
8. Liu, N., Lamerdin, J.E., Tebbs, R.S., Schild, D., Tucker, J.D., Shen, M.R., Brookman, K.W., Siciliano, M.J., Walter, C.A., Fan, W. *et al.* (1998) XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol Cell*, **1,** 783-793.
9. Pittman, D.L., Weinberg, L.R. and Schimenti, J.C. (1998) Identification, characterization, and genetic mapping of Rad51d, a new mouse and human RAD51/RecA-related gene. *Genomics*, **49,** 103-111.
10. West, S.C. (2003) Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol*, **4,** 435-445.
11. Miller, K.A., Yoshikawa, D.M., McConnell, I.R., Clark, R., Schild, D. and Albala, J.S. (2002) RAD51C Interacts with RAD51B and Is Central to a Larger Protein Complex in Vivo Exclusive of RAD51. *J Biol Chem*, **277,** 8406-8411.
12. Masson, J.Y., Tarsounas, M.C., Stasiak, A.Z., Stasiak, A., Shah, R., McIlwraith, M.J., Benson, F.E. and West, S.C. (2001) Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev*, **15,** 3296-3307.
13. Beernink, P.T., Krupka, S.S., Lao, V., Martin, G. and Coleman, M.A. (2002) Application of in vitro protein expression to human prote. *ScientificWorldJournal*, **2,** 73-74.
14. Brand, M., Moggs, J.G., Oulad-Abdelghani, M., Lejeune, F., Dilworth, F.J., Stevenin, J., Almouzni, G. and Tora, L. (2001) UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. *Embo J*, **20,** 3187-3196.
15. Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J. *et al.* (2002) Genomic instability in mice lacking histone H2AX. *Science*, **296,** 922-927.
16. Lueking, A., Possling, A., Huber, O., Beveridge, A., Horn, M., Eickhoff, H., Schuchardt, J., Lehrach, H. and Cahill, D.J. (2003) A Nonredundant Human Protein Chip for Antibody Screening and Serum Profiling. *Mol Cell Proteomics*, **2,** 1342-1349.
17. Joos, T.O., Schrenk, M., Hopfl, P., Kroger, K., Chowdhury, U., Stoll, D., Schorner, D., Durr, M., Herick, K., Rupp, S. *et al.* (2000) A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis*, **21,** 2641-2650.
18. Robinson, W.H., DiGennaro, C., Hueber, W., Haab, B.B., Kamachi, M., Dean, E.J., Fournel, S., Fong, D., Genovese, M.C., de Vegvar, H.E. *et al.* (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med*, **8,** 295-301.
19. Miller, J.C., Butler, E.B., Teh, B.S. and Haab, B.B. (2001) The application of protein microarrays to serum diagnostics: prostate cancer as a test case. *Dis Markers*, **17,** 225-234.