

Global analysis of protein function using protein microarrays

Michael G. Smith^a, Ghil Jona^a, Jason Ptacek^b, Geeta Devgan^a,
Heng Zhu^c, Xiaowei Zhu^a, Michael Snyder^{a,b,*}

^a*Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA*

^b*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA*

^c*Department of Pharmacology and Molecular Science, Johns Hopkins University, Baltimore, MD 21205, USA*

Available online 22 October 2004

Abstract

Protein microarrays containing thousands of proteins arrayed at high density can be prepared and probed for a wide variety of activities, thereby allowing the large scale analysis of many proteins simultaneously. In addition to identifying the activities of many previously uncharacterized proteins, protein microarrays can reveal new activities of well-characterized proteins, thus providing new insights about the functions of these proteins. Below, we describe the construction and use of protein microarrays and their applications using yeast as a model system.

© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Proteomics; Genome; Microarrays; High-throughput; Discovery research

1. Introduction

To date, the DNA sequences of over 200 organisms has been determined (<http://www.ncbi.nlm.nih.gov/Genomes/index.html>). Sequencing of entire genomes has identified large numbers of conserved genes, led to a more comprehensive understanding of genomic architecture, and revealed the presence of thousands of open reading frames (ORFs) whose function is either completely or largely unknown. A major challenge ahead is to determine the functions of unknown proteins and how they might work together with other known or unknown proteins to form the complex organisms in which they reside. This will involve determining the biochemical activities of each protein, how they might assemble together to carry out the biochemical reactions and cellular events in which they participate, and also how the proteins may be modified and/or regulated. Such information will aid our understanding of complex processes within these organisms such as development and ageing.

One technology that is very useful for identifying protein function, interactions and regulation is protein microarrays.

Protein microarrays are generally of two types: Antibody (or affinity) arrays and Functional Protein Arrays (Zhu et al., 2003; Jona and Snyder, 2003). Antibody arrays contain antibodies or equivalent arrays deposited at high density on the surface of a microscope slide in an addressable format (Angenendt et al., 2002). Complex mixture of proteins can be incubated on the array and antigenic peptides will bind to their corresponding antibody. The amount of antigen bound to each antibody can be determined. These arrays have been useful in profiling of cancer cells (Madoz-Gurpide et al., 2001; Knezevic et al., 2001; Sreekumar et al., 2001). Functional protein microarrays involve depositing sets of proteins onto the surface of a slide and characterizing them using a wide range of biochemical activities. Below, we review the construction and use of functional protein microarray and offer a number of examples of their utility.

2. Construction of protein microarrays: basic principles

There are three requirements for the production of functional protein microarrays (Schweitzer et al., 2003): (a) the construction of an ordered expression library, (b)

* Corresponding author. Tel.: +1 203 432 6139; fax: +1 203 432 3597.
E-mail address: michael.snyder@yale.edu (M. Snyder).

methods for purifying large number of proteins and (c) the array technology itself.

Production of an expression library involves the high-throughput cloning of open reading frames into an appropriate expression vector and validation of the quality of intact clones. Recombination cloning has emerged as the fastest method for clone production (Phizicky et al., 2003). Most current methods involve amplification of ORFs using primers that contain recombination sites at their termini and PCR. The ORFs can be cloned directly into a vector containing the appropriate recombination sites and a selectable marker. Two in vitro cloning systems that are currently popular are the GATEWAY system (Invitrogen, Carlsbad, CA) which uses the lambda Int recombinase and the Creator system which uses Cre (BD Biosciences, San Jose, CA). Additional cloning systems are available which use yeast recombination (Hudson et al., 1997) or LIC cloning (Dieckman et al., 2002).

A wide range of expression systems currently exists (Braun and LaBaer, 2003). The use of a homologous or related expression system is preferred for producing functional proteins. Presumably the proteins are more likely to be correctly folded and have proper posttranslational modification and cofactors. Thus, bacterial proteins can be expressed in bacteria, yeast proteins can be expressed in yeast, and plant proteins can be expressed in plants. Expression of mammalian proteins in mammalian cell is expensive and many researchers opt to use insect cells or other systems for the production of mammalian proteins. In order to purify proteins, the expression vectors typically encode a tag which attaches to each gene at either their amino or carboxy termini resulting in the production of fusion proteins. Glutathione-S-transferase (GST) is the most commonly used tag but others such as the IgG binding domain of Protein A or multimers of the amino acid histidine (HisX6) are also used. Tagged clones can then be verified by sequence analysis, typically by sequencing the vector-insert junction: although, for optimal verification the entire insert is sequenced. Clones deemed to have no mutations are processed further.

Once a high quality expression system is prepared and verified, the proteins are purified. Cells expressing each clone are grown in cultures in a 96-well format. Cells are concentrated, lysed and the proteins purified using the affinity tag of choice. Purified proteins are stored in glycerol and frozen at -80°C .

Protein microarrays are typically prepared by depositing the proteins onto a microscope slide using a standard contact printer (MacBeath and Schreiber, 2000; Zhu et al., 2001). A variety of slide surfaces can be used. Popular types include aldehyde surfaces for random attachment through amines, nitrocellulose coated slides for attachment presumably through hydrophobic regions, or nickel coated slides for affinity attachment of HisX6 tags. The latter was reported to provide a 10-fold better signal relative to random attachment methods (Zhu et al., 2001). However, this method does not

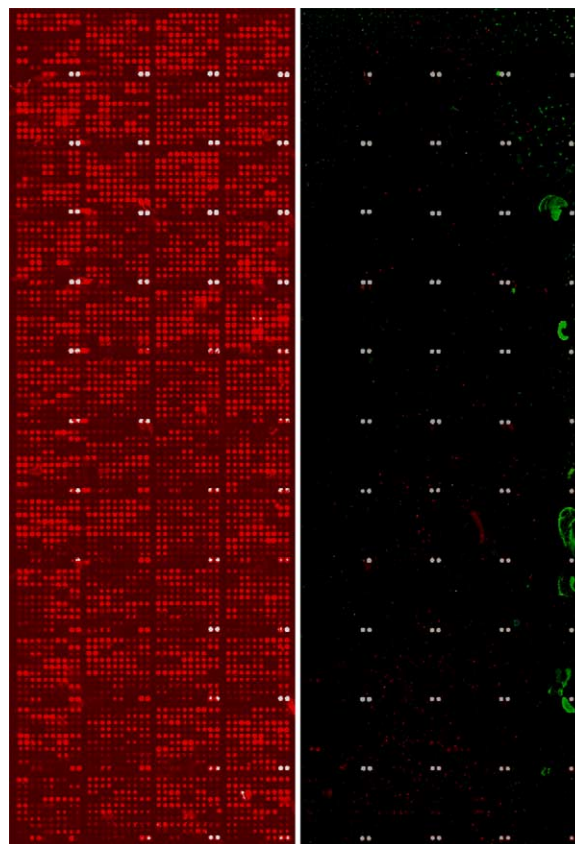


Fig. 1. Yeast protein microarray. On the left, a yeast protein microarray probed with anti-GST antibody followed by probing with Cy5-conjugated anti-rabbit antibody. The relative level of the printed GST-fused proteins can be determined. On the right, the same microarray probed Cy3-conjugated streptavidin, which recognizes the biotinylated control proteins printed on the slide.

work well for certain applications such as kinase assays. Once printed the quantities of material deposited on a slide can be determined by indirect immunofluorescence using antibodies raised against the GST tag (Fig. 1).

Once proteins are deposited, they can be tested for a variety of biochemical activities including interactions with proteins, small molecules, or nucleic acids, and kinase activities (MacBeath and Schreiber, 2000; Zhu et al., 2001; Hall et al., 2004). Binding assays are typically carried out using either biotinylated probes which are subsequently detected using streptavidin labeled with a fluorescent dye or alternatively using a probe that has been directly labeled with a fluorescent dye (Fig. 2). Examples different binding assays and other assays are described below.

Recently, a method has been described that allows the production of proteins directly from a PCR product using an in vitro transcription/translation system (Ramachandran et al., 2004). The reactions are carried out directly on slides and the proteins attach to slides through affinity tags. This procedure alleviates the laborious cloning steps and greatly accelerates the production of protein arrays. Additional

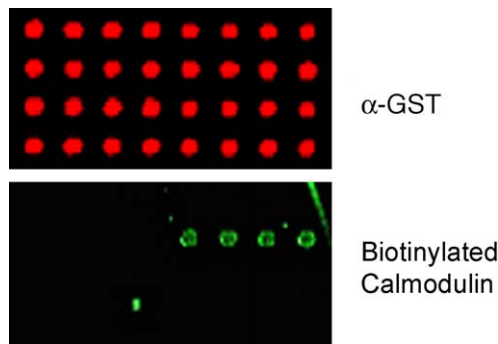


Fig. 2. Protein–protein interaction assay. Top: anti-GST probing of the microarray (as in Fig. 1). Bottom: a yeast protein microarray from the same printing probed with biotinylated calmodulin, and stained with Cy3-streptavidin.

effort is required to compare its sensitivity relative to use of purified proteins.

One limitation of the microarray approach is that arrays can only be constructed using proteins from organisms whose genome has been sequenced. This is not a significant problem in that most of the common experimental organisms have been sequenced and given the rapid rate of sequence acquisition, all relevant genomes will be known soon. One technical issue that is currently being addressed is the ability to purify full length, active proteins. Most proteins are small, soluble and easy to purify. However, some proteins are difficult to purify due to their exceptionally large size or hydrophobicity. Advances in cloning technology and membrane protein microarrays will enable the study of these more difficult proteins.

3. Construction of a yeast proteome microarray

To illustrate the construction and utility of proteins chips, our work with a yeast proteome array is described (Zhu et al., 2001). Briefly, a library of ~6000 yeast expression clones was

prepared in an expression vector that overproduces the proteins as GST fusion proteins when yeast are grown in media containing galactose. The clones were prepared using recombination in yeast, sequenced across their vector-insert junction and reintroduced into yeast to form clonal isolates. Cultures were grown in a 96-well format and expression of proteins is induced by growth in galactose. Cells were concentrated, lysed and proteins prepared in a 96-well format. Immunoblot analysis revealed that approximately 75–80% of the proteins are full length and in sufficient quantities to be detected using most assays. The proteins were then deposited onto nickel- or nitrocellulose-coated slides for binding assays. Other types of slides have been used for kinase assays.

4. Binding assay applications of the yeast proteome array

The yeast proteome array has been used for a wide variety of applications (Fig. 3). Initial studies investigated protein–protein interactions using biotinylated probes (Zhu et al., 2001). Incubation of arrays with biotinylated calmodulin revealed many known partners for calmodulin and well as 33 new binding partners. Comparison of the binding targets revealed that many share a common motif, similar, but distinct from the IQ containing motifs that had been reported previously.

In addition to protein–protein interactions, slides were also examined for interactions with phospholipids (Zhu et al., 2001). Labeled phosphatidylcholine (PC) liposomes either without phosphoinositides (PIs) or containing 5% of either PI₃P, PI₄P, PI_{3,4}P₂, PI_{4,5}P₂, and PI_{3,4,5}P₃ were incubated with the protein array. Greater than 150 phospholipid binding proteins were identified; one third of the lipid binding proteins exhibited a preference with PI as compared to PC vesicles alone. A wide range of proteins was found to bind to the lipid vesicles and over 50% were previously known to be associated with membranes.

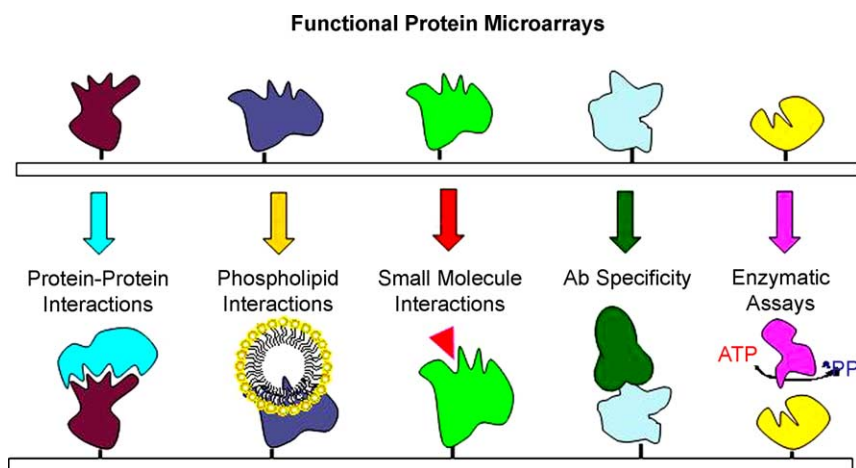


Fig. 3. Functional protein microarrays. Spotted protein arrays can be used to study many interactions including protein–protein, protein–lipid, protein–small molecule. These microarrays can also be used to verify antibody targets and identify enzymatic substrates.

Screens were also performed for protein interactions with small molecules (Huang et al., 2004). Identifying proteins that interact with small molecules is expected to be particularly important for determining the mechanism of action of drugs and their side effects. To test this idea, Huang et al. (2004) probed the yeast proteome array with two small molecules that affect the TOR signaling pathway. Surprisingly, each molecule bound to multiple proteins indicating that many small molecules are not likely to be monospecific. Examination of the phenotypes of yeast strains mutated for each target helped determine which proteins represented the physiologically relevant target in each case.

We have recently screened for novel DNA binding activities by probing the yeast proteome array with fluorescently-labeled yeast genomic DNA (Hall et al., 2004). 200 proteins that bound DNA were identified. One half were known or expected to bind DNA the other half were not. The proteins unexpected to bind DNA comprised a wide variety of functional categories and their verity as DNA binding proteins remains to be determined in vivo. We selected eight proteins and determined their association with specific chromosomal loci in vivo using chromatin immunoprecipitation and probing of DNA microarrays (chIP chip, Horak and Snyder, 2002). One metabolic enzyme, Arg5,6, which is involved in ornithine biosynthesis (a precursor to arginine) was found to associate with specific mitochondrial and nuclear loci in vivo. Follow up experiments revealed that this enzyme associates with specific chromosomal loci in vitro and we were able to use this information to define a DNA-binding motif for this protein. Thus, a novel DNA binding activity was found to be associated with a well-characterized protein thereby identifying a novel function for that protein. It is plausible that many eukaryotic proteins will have multiple functions, many of which will only be determined using unbiased proteomic approaches.

Our studies using the yeast proteome microarray have enabled us to study protein–protein interactions using a well-studied protein and redefine its preferred binding motif. Our analysis of protein–lipid interactions revealed a number of proteins previously thought to interact with membranes and that many proteins can distinguish the various types of membranes. Small molecules studies have helped to determine which target of the drug is the physiologically relevant target and will aid in minimizing the drug's side effects. Lastly, our studies on nucleic acid–protein interactions revealed that proteins or enzymes of known function may have additional or alternative functions that were previously unknown.

5. Other applications of the yeast proteome array

Microarrays are ideally suited for examining antibody specificity; in principle all potential proteins that react with an antibody can be identified using a comprehensive

proteome array. Moreover they can contain both native and denatured proteins. This idea was tested by Michaud et al. (2003) who probed the yeast proteome with 14 different commercial antibodies: 6 monoclonal antibodies, 6 anti-peptide antibodies and two antibodies raised against full length proteins. In general, the monoclonal antibodies exhibited higher specificity than their polyclonal counterparts; among the polyclonal sera, the anti-peptide antibodies were the most specific. However, even monoclonal antibodies can show crossreactivity demonstrating the importance in (a) having assays to determine the specificity of antibodies that are used and (b) multiple antibody probes may ultimately be most useful for antibody studies.

In addition to assessing antibody specificity, proteome arrays are expected to be valuable for examining post-translational modifications. We have used the arrays for examining in vitro substrates for yeast kinases (Ptacek, J. et al., unpublished). The chips were incubated with proteins kinases in the presence of ^{33}P -ATP and proteins that were phosphorylated identified. This procedure identified many known substrates as well as a number of new ones. Many of the new ones reside in the same subcellular compartments as the kinases indicating that they are likely to be bona fide substrates in vivo.

6. Conclusions and future directions

In summary, protein microarrays have shown considerable promise in determining protein–protein, protein–lipid, protein–ligand, and enzyme–substrate interactions. Protein microarrays also have great potential in drug development and clinical diagnostics. Future directions will focus on a number of areas: (a) development of such arrays for other organisms (Reboul et al., 2003), (b) preparation of functional arrays for membrane proteins, and (c) use of label-free detection systems for monitoring interactions.

References

- Angenendt, P., Glokler, J., Murphy, D., Lehrach, H., Cahill, D.J., 2002. Toward optimized antibody microarrays: a comparison of current microarray support materials. *Anal. Biochem.* 309, 253–260.
- Braun, P., LaBaer, J., 2003. High throughput protein production for functional proteomics. *Trends Biotechnol.* 9, 383–388.
- Dieckman, L., Gu, M., Stols, L., Donnelly, M.I., Collart, F.R., 2002. High throughput methods for gene cloning and expression. *Protein Expr. Purif.* 25, 1–7.
- Jona, G., Snyder, M., 2003. Recent developments in analytical and functional protein microarrays. *Curr. Opin. Mol. Ther.* 3, 271–277.
- Hall, D.A., Zhu, H., Royce, T., Gerstein, M., Snyder, M., 2004. Regulation of gene expression by a metabolic enzyme. *Science* 306, 482–484.
- Horak, C., Snyder, M., 2002. ChIP chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol.* 350, 469–484.
- Huang, J., Zhu, H., Haggarty, S.J., Spring, D.R., Snyder, M., Schreiber, S., 2004. Finding new components of the TOR signaling pathway using chemical genetics and proteome chips, submitted for publication.

- Hudson Jr., J.R., Dawson, E.P., Rushing, K.L., Jackson, C.H., Lockshon, D., Conover, D., Lanciault, C., Harris, J.R., Simmons, S.J., Rothstein, R., Fields, S., 1997. The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res.* 7, 1169–1173.
- Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J., Petricoin 3rd, E.F., Krizman, D.B., 2001. Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* 10, 1271–1278.
- MacBeath, G., Schreiber, S.L., 2000. Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763.
- Madoz-Gurpide, J., Wang, H., Misek, D.E., Brichory, F., Hanash, S.M., 2001. Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. *Proteomics* 10, 1279–1287.
- Michaud, G.A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P.F., Schweitzer, B.L., 2003. Analyzing antibody specificity with whole proteome microarrays. *Nat. Biotechnol.* 21, 1509–1512.
- Phizicky, E., Bastiaens, P.I., Zhu, H., Snyder, M., Fields, S., 2003. Protein analysis on a proteomic scale. *Nature* 422, 208–215.
- Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., LaBaer, J., 2004. Self-assembling protein microarrays. *Science* 305, 86–90.
- Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T., Hudson Jr., J.R., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S., Endress, G.A., Jenna, S., Chevet, E., Papatotiropoulos, V., Tolias, P.P., Ptacek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H., Doucette-Stamm, L., Hill, D.E., Vidal, M., 2003. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* 34, 35–41.
- Schweitzer, B., Predki, P., Snyder, M., 2003. Microarrays to characterize protein interactions on a whole-proteome scale. *Proteomics* 3, 2190–2199.
- Sreekumar, A., Nyati, M.K., Varambally, S., Barrette, T.R., Ghosh, D., Lawrence, T.S., Chinnaiyan, A.M., 2001. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.* 61, 7585–7593.
- Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, D.A., Gerstein, M., Snyder, M., 2001. Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105.
- Zhu, H., Bilgin, M., Snyder, M., 2003. *Proteomics. Ann. Rev. Biochem.* 72, 783–812.