

Review

Barry Schweitzer¹
Paul Predki¹
Michael Snyder²

¹Protometrix, Branford,
CT, USA

²Yale University,
New Haven, CT, USA

Microarrays to characterize protein interactions on a whole-proteome scale

Protein microarrays contain a defined set of proteins spotted and analyzed at high density, and can be generally classified into two categories; protein profiling arrays and functional protein arrays. Functional protein arrays can be made up of any type of protein, and therefore have a diverse set of useful applications. Advantages of these arrays include low reagent consumption, rapid interpretation of results, and the ability to easily control experimental conditions. The ultimate form of a functional protein array consists of all of the proteins encoded by the genome of an organism; such an array would be the whole proteome equivalent of the whole genome DNA arrays that are now available. While proteome microarrays may not have reached the stage of maturity of DNA microarrays, recent developments have shown that many of the barriers holding back the technology can be overcome. Arrays of this type have already been used to rapidly screen large numbers of proteins simultaneously for biochemical activities, protein-protein interactions, protein-lipid interactions, protein-nucleic acid interactions, and protein-small molecule interactions. Eventually, functional protein arrays will be used to facilitate various steps in the drug discovery and early development processes that are currently bottlenecks in the drug development continuum.

Keywords: Biomarker / Microarray / Proteome / Review

PRO 0610

Contents

1	Introduction	2190
2	Content: Producing the proteins	2191
3	Microarray production	2193
4	Detection	2194
5	Applications	2195
5.1	Protein-protein interactions	2195
5.2	Antibody-specificity profiling	2196
5.3	Immune profiling	2196
5.4	Protein-small molecule interactions	2196
5.5	Other applications	2198
6	Conclusions and future prospects	2198
7	References	2198

1 Introduction

Over the past few years, the science of genomics has revolutionized both basic research and drug discovery. It has now become routine, for instance, to use DNA micro-

arrays to probe the expression of thousands of genes and even the entire gene repertoire of a single organism. There are hundreds of examples in the literature that demonstrate the value of these whole-genome arrays for identifying genes that are differentially expressed during development, in response to environmental stimuli [1], and, more recently, in several disease states [2–4]. Similar techniques have also enabled genomics-based drug discovery to identify large numbers of new drug targets [5]. A significant drawback to such approaches is that it is quite difficult, without a great deal of further experimentation, to assess the actual biological relevance of the often hundreds of different genes that are identified in these investigations. A key factor behind this difficulty is that the variable being measured in these experiments (the relative level of RNA between two or more states) may only be indirectly related to the causative element(s) (a protein or proteins) and, in some cases, may not be related to it at all [6]. Furthermore, the data obtained from DNA microarray experiments does not give the investigator any direct information about the function of the gene products or how they might interact with one another.

Protein microarray technology may have the greatest potential for providing direct information on protein profiles and function with the level of comprehensiveness

Correspondence: Dr. Michael Snyder, Department of Molecular, cellular and Developmental Biology, Yale University, CTO 6405, New Haven, USA
E-mail: michael.snyder@yale.edu
Fax: +1-203-432-3597

Abbreviations: GST, glutathione S-transferase; His, histidine

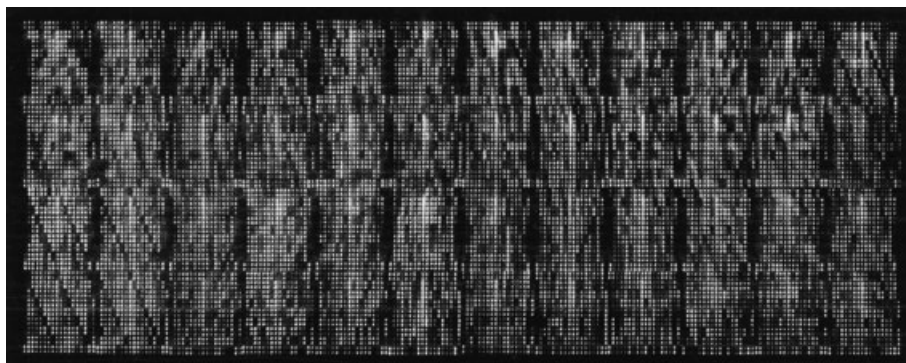


Figure 1. Fluorescent Image of Yeast ProtoArray™. Microscope slides were spotted with 5000 different proteins cloned from yeast. Each protein on the array was detected using a Cy5-labeled antibody directed against an attached epitope tag. Slides were scanned using an Axon 4000B microarray laser scanner (Axon Instruments, Union City, CA, USA) Copyright 2003 Proteomix Inc.

that scientists have come to expect with DNA microarrays [7–9]. Protein microarrays contain a defined set of proteins spotted and analyzed at high density, and can be generally classified into two categories; protein profiling arrays and functional protein arrays. Protein profiling arrays, which are not the subject of this review, usually consist of multiple antibodies printed on glass slides and are used to measure protein abundance and/or alterations (reviewed in [10]). Functional protein arrays can be made up of any type of protein, and therefore have a more diverse set of useful applications. Advantages of these arrays include low reagent consumption, rapid interpretation of results, and the ability to easily control experimental conditions. The key advantage of functional protein arrays, however, is the ability to rapidly screen large numbers of proteins simultaneously for biochemical activities, protein-protein interactions, protein-lipid interactions, protein-nucleic acid interactions, and protein-small molecule interactions. Using these arrays, one could, in a single experiment, determine all of the substrates for a protein modifying enzyme, build an entire protein interaction network, or determine all of the potential binding partners in a cell for a drug under development.

The ultimate form of a functional protein array would consist of all of the proteins encoded by the genome of an organism; such an array would be the whole proteome equivalent of the whole genome arrays that are now available. Arrays containing families of proteins with similar biological activities (e.g. protein kinases), proteins expressed in the same tissue, or proteins associated with a particular disease state, would also have obvious utility. Compared to DNA arrays, however, the manufacture and use of whole-proteome or protein family arrays present potentially significant challenges. These challenges are usually broken down into several key areas. Prohibitive costs, the length of time needed, and various technical limitations in producing the proteins that would go on these arrays have been cited as chief concerns. It has also been pointed out that immobilizing large numbers of diverse proteins on a glass slide is likely to be much more

difficult than the manufacture of arrays of DNA, which is clearly not as chemically complex as proteins. Similarly, the lack of an amplification procedure for proteins analogous to PCR for DNA has been mentioned as a problem for detection of interactions on protein arrays.

Despite the challenges, the first actual demonstration that a whole-proteome microarray can be made and used in functional assays was recently published [11]. This array consisted of approximately 5000 proteins from the budding yeast *Saccharomyces cerevisiae* deposited onto a glass microscope slide (Fig. 1). In light of this accomplishment, there is clearly a need to take a new look at the realities of producing whole-proteome arrays. It is the purpose of this review to attempt to objectively analyze the current state of the art as well as the remaining challenges regarding content, manufacture, detection, and applications of this exciting new technology.

2 Content: Producing the proteins

The fundamental advantage of using whole-proteome arrays for applications such as protein-protein interaction mapping, small molecule screening, and antibody specificity profiling is the ability to carry out the parallel analysis of thousands of proteins in a high-throughput fashion. In order to produce such arrays, however, it is necessary to convert genomic sequence information into thousands of pure, functional proteins. The basic technologies necessary for producing this content (cDNA cloning, PCR, recombinant protein expression and purification) have been in place for a number of years. Efforts to integrate each of these technologies into an automated, high-throughput system for the industrial-scale production of proteins with appropriate quality assurance procedures are only recently underway.

The initial step in the development of whole-proteome arrays is the identification of ORFs for the organism of choice. For simple organisms such as *S. cerevisiae* which

generally lack a complex intron-exon structure, this process is relatively straightforward and genomic DNA can be used as the source material. In fact, the first example of this approach was the genomic-scale PCR amplification of the ~6000 ORFs from *S. cerevisiae* for cloning into yeast plasmids [12]. A more complex strategy for identifying ORFs from genomic sequence information is required for higher organisms due to the necessity of defining translational start sites and splicing signals [13]. Once identified, the source material for cloning is usually a cDNA library or a collection of libraries. In this case, successful cloning is dependent on the presence of a full-length copy of mRNA for the genes of interest; consequently, a higher failure rate is expected for very rare and/or very large mRNAs. Nevertheless, various efforts are currently underway to generate nonredundant collections of genes suitable for use as the starting material for generating, among other things, whole-proteome arrays. These efforts range from the *Caenorhabditis elegans* ORFeome project [14] to the various public efforts to generate human genes collections, such as the Unigene set, which is a collection generated from a human fetal brain cDNA library [15], the Full-length Expression (FLEX) Gene repository under development at the Harvard Institute of Proteomics [16, 17], the Integrated Molecular Analysis of Genomes and their Expression (I.M.A.G.E.) cDNA collection that is based at the Lawrence Livermore National Laboratory [18] and the associated sequence verified Mammalian Gene Collection (MGC) full-length clone collection [19, 20]. Commercial efforts are also being undertaken to create biologically useful collections by companies such as Invitrogen and Protometrix.

A variety of strategies for insertion of the cloned ORFs into vectors can be employed (reviewed in [21]). Several recombination-based methods are now widely used that allow the facile shuttling of the cloned ORFs into various types of expression systems [22]. The use of such methods is unnecessary, however, if the choice of expression is clear or if the compromises involved in the use of a single system are acceptable. For example, expression in a homologous system is ideal for the biochemical analysis of *Escherichia coli* and yeast proteins because the proteins are in their natural environment and thus should fold properly, are subject to native modifications, and can interact with their native partners. As mentioned above, the utility of using expression in yeast to produce functional arrays of the yeast proteome has been already been demonstrated [11, 23]. For other organisms, however, heterologous expression systems are usually employed. The most common means of heterologous expression of recombinant proteins is to generate them in *E. coli*, which is low in cost and easy to use. It has been shown, however, that only about half of likely soluble recombinant proteins are actually expressed in soluble form in *E. coli* [24], probably due to the lack of proper folding resulting in the generation of insoluble inclusion bodies. Another preferred system for heterologous expression is the use of insect cells, which result in modifications that are usually similar to mammalian cells. At Protometrix, we employ a massively parallel baculovirus-based expression platform that has been adapted and optimized for high-throughput expression of mammalian proteins; we generally achieve an 80% or higher rate of success in producing soluble human proteins of the expected M_r using this system (Fig. 2), which is about two-fold higher than previous



Figure 2. Western analysis of human proteins expressed in a high-throughput baculovirus system. Soluble fractions from 92 different proteins were isolated from cultures of recombinant baculovirus-infected Sf9 cells, and were separated on SDS-PAGE gels. Detection of recombinant protein was carried out by Western analysis using an antibody directed against the GST tag on each protein. Analysis of the image indicated that 81/92 proteins (88%) were expressed at the predicted M_r .

reports [25]. It should be pointed out that this kind of parallel production of thousands of proteins is achievable with high efficiency and at a reasonable cost because of the relatively small amount of protein required for microarrays.

After expression, the next step in the process is to purify the proteins. The most efficient way to accomplish this task on a proteome-wide scale is to incorporate a fusion tag, a short peptide or protein domain that becomes linked to each member of a set of proteins, during expression vector construction. Some of the more commonly used fusion tags include glutathione S-transferase (GST), histidine (His)₆, Protein A, maltose binding protein, and the FLAG peptide [21]. These tags feature high affinity and selectivity for binding to specific resins to facilitate purification and elution under conditions that retain activity. Purification using these tags can be adapted to high-throughput protocols; this was demonstrated by Snyder and coworkers when they used glutathione-agarose to purify 5800 individual yeast GST fusion proteins, 1152 at a time in 96-well format [11, 23]. At Protometrix, we have optimized and automated these protocols so that the purification of all 5800 proteins can now be accomplished in a single day. Each step in the purification is carried out at 4°C; this feature in combination with the overall speed of the process helps to ensure that proteins are purified in a functional form. Each protein is analyzed for integrity and purity before it is passed on for protein microarray production.

In summary, the generation of content for the first generation of proteome arrays is achievable with available technologies and is, in fact, already underway. Several additional points are worth mentioning, however. First, while there is certainly value in creating a collection of expression-ready ORF clones, there is much more effort involved, and therefore greater value, in creating an integrated system of clones and expressed, purified, and quality-controlled proteins resulting from these clones. Second, it is unlikely that the efforts currently underway will generate all of the protein information in an organism. First, a single spliced mRNA is generally chosen as a template for each gene; thus, the protein diversity resulting from RNA splicing is usually not well represented. Similarly, depending on the expression system used, certain post-translational modifications may not be present in the recombinant proteins. Membrane proteins, especially ones which span the membrane multiple times, remain a challenge for efficient expression and purification. Inevitably, there will be other members of a protein set that cannot be expressed, solubilized or purified under the generic conditions used to purify an entire proteome. Finally, proteins may be functionally inactive as fusion proteins.

Nevertheless, an initial proteome collection, despite being incomplete, will certainly be extremely valuable to the life sciences community.

3 Microarray production

In order to carry out reproducible and reliable biochemical assays on a protein microarray, it is necessary to immobilize the proteins in a way that results in efficient deposition of functional protein. Due to chemical diversity of proteins, there have been questions about whether a technology currently exists or can be developed that can accomplish this kind of immobilization across an entire proteome. However, biochemists have been effectively immobilizing proteins on solid surfaces for decades, albeit usually one at a time. More recently, there have been reports on the printing of protein microarrays using a variety of surfaces and immobilization chemistries, including but not limited to polyvinylidene difluoride [26], agarose [27], and polyacrylamide gel pads [28]. Proteins arrayed on glass slides activated with aldehyde [29], polylysine [30, 31], or a homofunctional cross-linker [32] have also been described. All of the above approaches result in essentially random attachment of proteins to the surface of the array. Since most of the studies cited have been carried out with antibodies and there have been little comparative studies carried out on the different surfaces, it is not possible at this time to determine which, if any, of the above surfaces or surface chemistries are likely to be best suited for whole proteome arrays. What is clear, however, is that a number of choices are readily available for fabrication of protein microarrays, and thus claims regarding the need for highly engineered surfaces for protein arrays should be viewed with a certain degree of skepticism.

In addition to the surfaces described above, we and others have found nitrocellulose-coated slides to be particularly useful for immobilization of large numbers of different proteins [26, 33–35]. A direct comparison of these slides with amine- or aldehyde-derivatized slides for producing microarrays of proteins isolated from cancer cells clearly demonstrated that nitrocellulose-coated slides provided a nearly quantitative retention of the spotted proteins, and significantly higher detection sensitivity than the other surfaces [33]. A similar increase in sensitivity over aldehyde slides was also observed when measuring protein-protein interactions on nitrocellulose-coated slides [36]. This increase in sensitivity was shown to be the result of the higher binding capacity of the nitrocellulose-coated surface [36]. Although there is some concern about possible denaturation of proteins on the surface of such slides, their ability to retain greater amounts of pro-

tein may make them a better substrate for use in applications of whole proteome arrays that require high sensitivity, such as the detection of weak protein-protein or small molecule-protein interactions.

In principle, attachment of proteins in an oriented fashion is a preferable immobilization strategy for use on protein microarrays. In the case of the yeast proteome array described by Snyder and coworkers, this was accomplished by expressing every protein with an *N*-terminal His6 fusion tag and printing the fusion proteins on nickel-coated glass slides [11]. Assay performance on these slides was shown to be superior to random attachment using aldehyde-coated slides. Lesaichere *et al.* proposed a new strategy for the site-directed attachment of fusion proteins containing an intein-tag with a chitin-binding domain. In this strategy, fusion proteins are purified on a column of chitin beads; cleavage of the protein-intein fusion with biotinylated cysteine produces proteins that are selectively biotinylated on their C-terminal end, enabling them to be immobilized on avidin-coated slides in an oriented fashion [37].

The production of any protein microarray requires careful preparation of samples and slides prior to printing, control of printing conditions such as humidity, temperature, dust levels, and pin washing, and implementation of stringent quality control processes. Despite the fact that proteome-

scale arrays may contain 10–100 times more proteins than any antibody array that has been described to date, it has been our experience that the level of difficulty in manufacturing these arrays is comparable to that of antibody arrays. Most current applications of whole-proteome arrays are essentially qualitative in nature, whereas most antibody-based array assays require precise quantitation. Consequently, proteome-based array assays are more tolerant to variation in the amount of deposited protein, spot morphology, *etc.* For this reason, contact-type arrays are adequate for producing proteome arrays.

4 Detection

Detection of probe binding on proteome arrays should provide specific results with a high signal-to-noise ratio and relatively low instrumentation costs. Radioactivity, chemiluminescence and fluorescence are all suitable detection strategies on proteome microarrays, although fluorescence is certainly the most widely used detection method for reasons of sensitivity, stability, and availability of fluorescent scanners tailored for microarray use. Radioactivity is used much less frequently due to safety concerns and the requirement for long exposure times. Chemiluminescence, although highly sensitive, has drawbacks in terms of the potential for low feature resolution due to signal bleeding and limited dynamic range (Fig. 3).

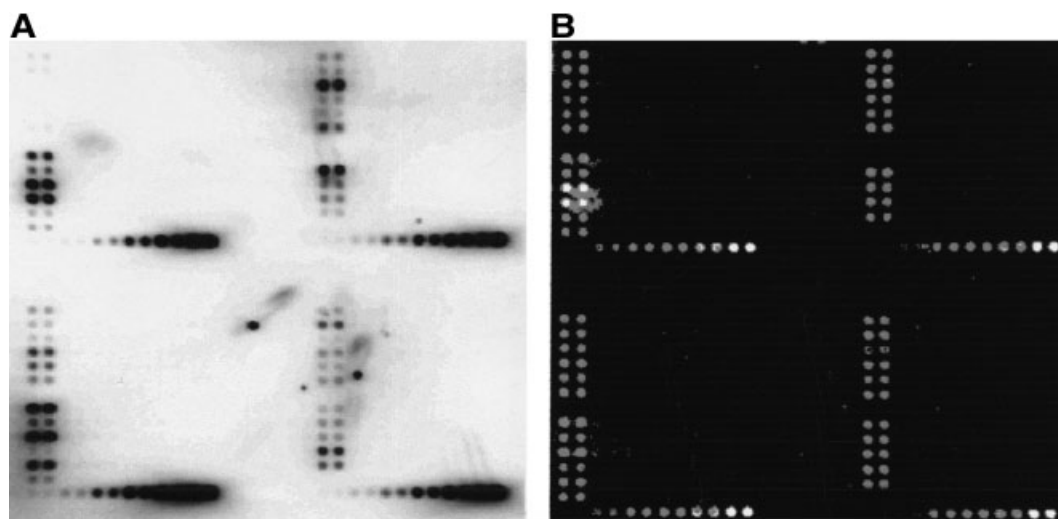


Figure 3. (A) Chemiluminescent detection on protein microarrays. The array was incubated with biotinylated calmodulin followed by horseradish peroxidase-conjugated streptavidin. Chemiluminescence signal was generated using the Pierce Femto reagent (Rockford, IL, USA) and detected on an Alpha Innotech gel documentation system (San Leandro, CA, USA). A portion of the image showing calmodulin binding proteins and biotinylated antibody positive control spots is presented. (B) Fluorescent detection on protein microarrays. The array was incubated with biotinylated calmodulin followed by Cy5-labeled streptavidin. The slide was imaged on an Axon 4000B fluorescent microarray scanner. A portion of the image showing calmodulin binding proteins and biotinylated antibody positive control spots is presented.

Furthermore, signal detection on microarrays using radioactivity and chemiluminescence can only be performed once, whereas fluorescence-based microarrays can be archived for future imaging. Surface plasmon resonance (SPR) is likely to soon emerge as an important means of detection on protein arrays. In SPR, test proteins are immobilized on a gold-coated chip, unlabeled probe is added, and the change in the angle of reflection of light caused by binding of the probe to immobilized protein is measured [38–41]. Two key advantages of SPR detection are the ability to use unlabeled probes, which results in better flexibility and possibly better preservation of probe function, and the ability to monitor the kinetics of binding, which increases the value of the data. SPR is not very sensitive, however, and thus requires more protein to be deposited on the surface of the sensor than is needed on microarrays when fluorescence, chemiluminescence, or radioactivity is used for detection. In addition, SPR-based arrays lack the ability to measure more than a few proteins at a time, and it is likely to be several years before a system able to measure binding to an entire proteome becomes available.

5 Applications

5.1 Protein-protein interactions

By identifying interacting protein partners of known function with those whose role is uncharacterized, it is often possible to infer the biological process and/or activity of the uncharacterized protein. Currently, the most popular method for detecting protein-protein interactions is the yeast two-hybrid approach. Interaction mapping using yeast two-hybrid has now been applied to several organ-

isms, including two large-scale efforts for the budding yeast *S. cerevisiae* [42, 43]. A comparison of the results of these efforts raises questions about the comprehensiveness and specificity of this approach; for example, only 6% of the 2200 reported interactions were found in both studies [44]. An alternative method for revealing protein-protein interactions involves affinity purification of proteins coupled with their identification using MS. Using this approach, Gavin *et al.* and Ho *et al.* identified a number of protein complexes from *S. cerevisiae* [45, 46]. These data sets provide crucial information regarding multiprotein assemblies within the cell that have not been obtained by two-hybrid analyses alone. Like the yeast two-hybrid data sets, a comparison of the data obtained from the two MS-based studies also showed relatively poor agreement between them; furthermore, little overlap between the yeast two-hybrid interaction datasets and the MS datasets was observed [47].

Snyder and coworkers have provided a conclusive demonstration that proteome microarrays can be used to screen for protein-protein interactions [11]. For example, 39 calmodulin-interacting proteins were identified by adding biotinylated calmodulin to a chip printed with the majority of the yeast proteome. Within this set, several expected calmodulin-binding proteins as well as 33 novel calmodulin-interacting proteins were revealed. In a single experiment, a new consensus binding site for calmodulin was defined; an achievement that would not have been possible without the comprehensive screening enabled by the proteome microarray. Several of the interactions with calmodulin were missed in the large-scale yeast two-hybrid and/or the affinity purification-MS studies, including the well-known interaction of calmodulin with calmodulin kinase. Figure 4 demonstrates the reciprocal

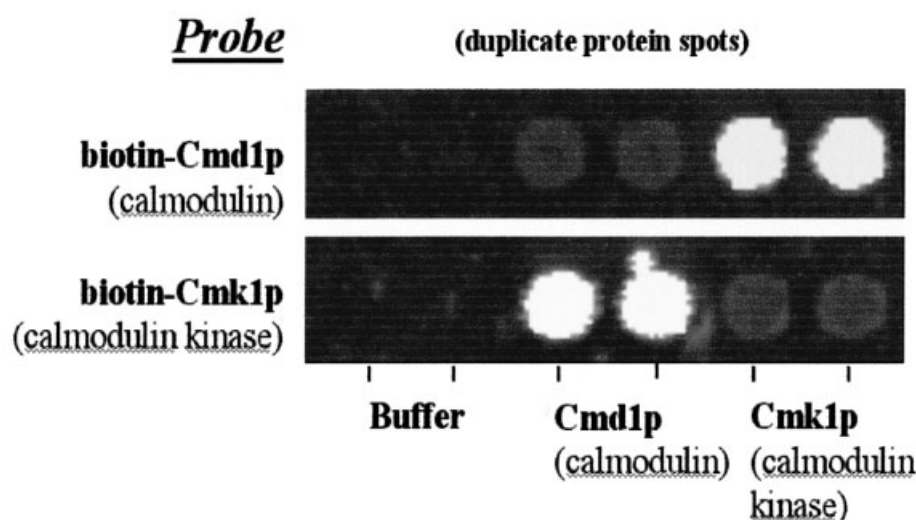


Figure 4. Demonstration of reciprocal interactions on protein arrays. Identical arrays were incubated with either biotinylated calmodulin (upper) or biotinylated calmodulin kinase (lower). Detection was carried out using Cy5-labeled streptavidin and imaged on an Axon 4000B microarray scanner. Portions of the arrays showing the interaction of the calmodulin probe with the duplicate spots of calmodulin kinase on the array (upper) and biotinylated calmodulin kinase with the duplicate spots of calmodulin on the array (lower) are presented.

nature of the interaction seen between calmodulin and calmodulin kinase; the interaction is observed independent of which protein is present in solution and which is bound on the surface of the array. Our experience has been that > 90% of protein interactions observed on our proteome microarrays exhibit such reciprocity.

In ongoing experiments with the yeast proteome microarray, we have identified a number of protein-protein interactions that to the best of our knowledge have not been revealed by yeast two-hybrid or MS-based approaches (data not shown). While we have also been able to verify many protein-protein interactions that have been reported using these methods, there have been others that we have not been able to repeat on the yeast proteome array. We recently carried out a study in which 20 pairs of proteins that had been identified as interacting by both yeast two-hybrid and MS were examined on microarrays (Michaud *et al.*, manuscript in preparation). The protein pairs were also tested for interactions by carrying out mobility shift assays in nondenaturing gels. Interestingly, 1/2 of the protein pairs did not show interactions in the mobility shift assay; these same pairs also did not show evidence of interactions on protein microarrays. On the other hand, 9 out of the 10 interactions that were seen in the mobility shift assay were also observed on microarrays. We conclude from these results that protein-protein interaction data obtained on microarrays is highly reliable in terms of sensitivity and specificity.

5.2 Antibody-specificity profiling

The use of antibodies as therapeutics, as well as the growing field of proteomics and its demand for high-throughput protein analysis, have been accompanied by an increasing demand for large numbers of antibodies with high specificities. An array containing every protein for the relevant organism represents the ideal format for an assay to test antibody specificity, since it allows the simultaneous screening of thousands of proteins in a single experiment. In a recent proof-of-principle of this approach, we screened antibodies against 11 yeast proteins with the yeast proteome microarray and found that, in addition to recognizing their cognate proteins, the antibodies cross-reacted with other yeast proteins to varying degrees (Michaud *et al.*, submitted for publication). Some of the interactions of the antibodies with noncognate proteins could be deduced by alignment of the primary amino acid sequences of the antigens and cross-reactive protein; however, these interactions could not be predicted *a priori*. We expect that similar types of results will be obtained using arrays of proteins against other species including human that are currently in development; thus,

pre-screening of antibodies on these arrays may become a critical part of the development of more specific antibodies for use in the clinic and in the laboratory.

5.3 Immune profiling

Proteome microarrays present a new and powerful approach to profiling the immune response by characterizing serum antibodies on a comprehensive scale. The process for identifying antigens which react with serum antibodies is simple; probe the array with serum, wash, and perform a secondary probing for (human) antibody. A recent demonstration of this approach was reported by Mezzasoma *et al.* who showed that arrays of microbial antigens could be used to detect human IgG and IgM directed towards these antigens in serum with good agreement with conventional ELISA assays [48]. Such studies, especially when applied to large populations, promise not only a new level of understanding of immune response to infection, but valuable information for directing the development of new and improved vaccines and serodiagnosis as a result.

Joos and coworkers initially demonstrated the utility of protein microarray technology for screening serum from patients with autoimmune disease for antibodies to 18 autoantigens [30]. More recently, Robinson *et al.* extended this work to include arrays containing 196 known autoantigens [49]. These studies used arrays of known targets of certain types of autoimmune disease. The autoantigens involved in other types of autoimmune disease (e.g. inflammatory bowel disease and psoriasis) are unknown. In principle, human proteome arrays could be used in this way to identify new autoantigens, which may help to guide the development of more efficacious antigen-specific toleration therapies [50]. Similarly, recent reports of autoantibodies against a number of antigens in sera from patients with different types of cancer suggest that protein microarrays with large numbers of tumor-expressed proteins could facilitate the identification of markers for cancer screening [51–54]. In some cases, as with HER-2/neu, the cancer biomarkers may even become therapeutic targets [55].

5.4 Protein-small molecule interactions

The ability to screen whole proteomes for drug binding has important implications for the drug discovery and development process, including helping to anticipate, reduce, and eliminate adverse drug side effects. For many drugs identified by cell-based screening, the mode of action is unknown. By screening a proteome array with drugs, candidate targets can be determined in a single

experiment. MacBeath and Schreiber were among the first to demonstrate this utility on protein arrays [29]. Specifically, they were able to show binding of the FK506 binding protein FKBP12 with several compounds with disassociation constants in the micromolar to nanomolar range. In Fig. 5, we show that binding of radiolabeled

GTP can be detected with high sensitivity on microarrays of GTP-binding proteins. Furthermore, a dose-response analysis for one of the proteins on the array revealed a close agreement with the published value for the GTP disassociation constant of this protein. Finally, a recent study by Fang *et al.* demonstrated that function-

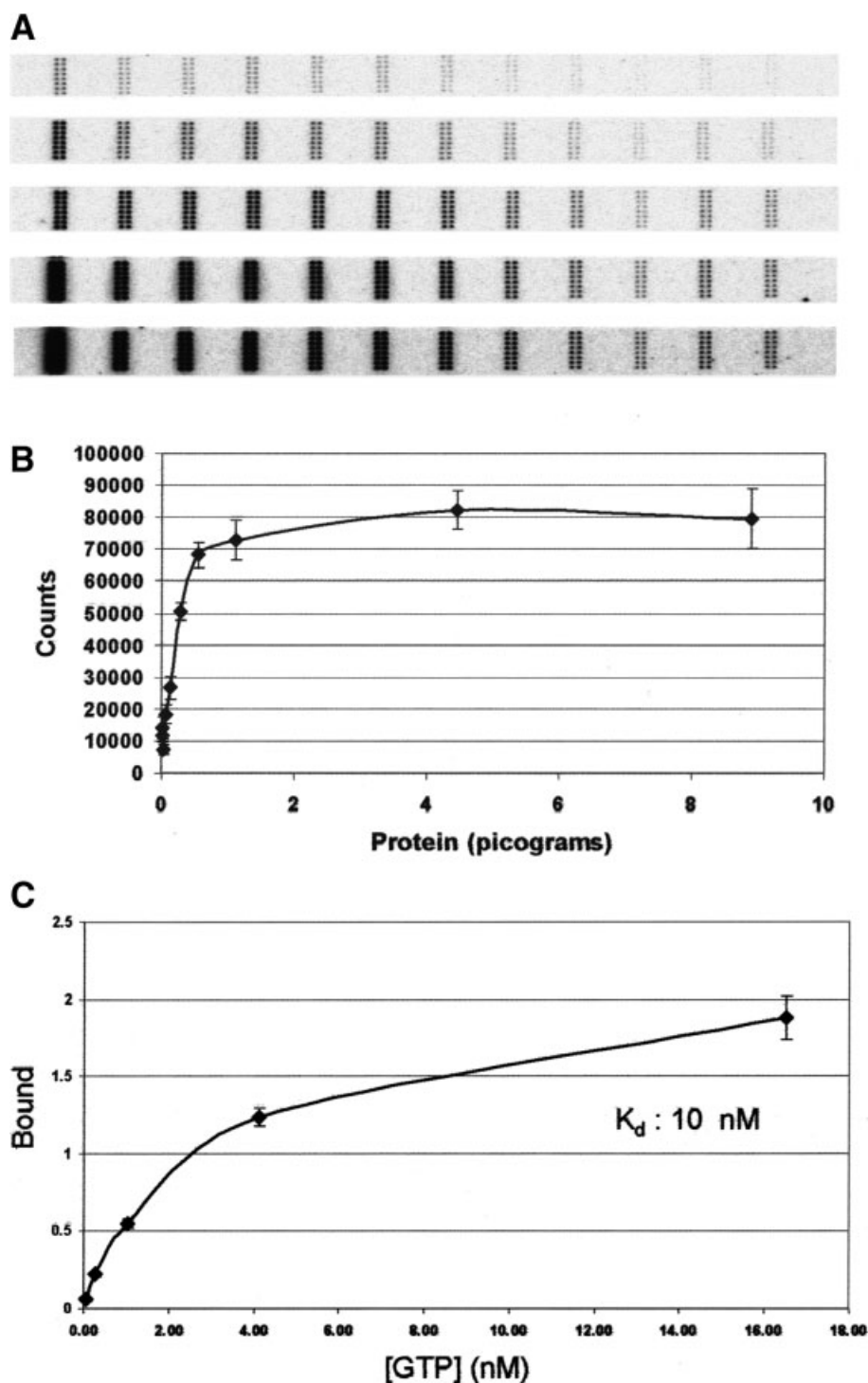


Figure 5. (A) GTP binding to human Rab3A on arrays. Serial dilutions of the human GTP binding protein Rab3A were printed on a microarray slide. Slides were incubated with [^{32} P]GTP and detection was carried out using a phosphorimager (Perkin Elmer, Boston, MA, USA). (B) Quantitation of GTP binding to Rab3A. Data obtained from the experiment shown in Fig. 5A are plotted. (C) Dose-response analysis of GTP binding to Rab3A. Microarrays printed with human Rab3A were incubated with various concentrations of [^{32} P]GTP. The amount of protein bound (in picograms) is plotted as a function of GTP concentration. Fitting of the points generated a curve corresponding to a binding constant of 10 nM.

nal G-protein coupled receptors in a lipid bilayer can be reconstituted on glass slides and shown to bind to specific ligands [56]. These results clearly show the potential of protein arrays for high-throughput drug screening for membrane-associated proteins, usually considered to be the most intractable for these types of analyses.

5.5 Other applications

The applications of proteome microarrays described above represent only a few of the types of experiments that have been carried out on these chips. For example, the yeast proteome chip was also probed with liposomes containing different phosphoinositides [11]. One hundred fifty protein targets were identified showing varying specificity dependent upon the head group; approximately 1/3 of these proteins were of unknown function, but over 50 were either membrane-associated or involved in lipid metabolism. Proteome microarrays are also an excellent means to comprehensively identify enzyme substrates. We have incubated proteome microarrays with protein kinases and radioactive ATP and been able to identify dozens of kinase substrates (data not shown). This type of screening not only provides interesting biological information, but is also a useful tool for developing assays for proteins that have only been predicted to have a particular enzymatic activity based on sequence homology.

Although whole proteome arrays may represent the ultimate in application flexibility, comprehensive subsets of the proteome also have great utility. These subsets may consist of portions of the proteome expressed in a particular tissue or subcellular compartment, or they may make up families of proteins with related function. An earlier example of this was the array of DNA-binding proteins described by Ge [57]. More recently, Bedford and coworkers reported the development of an array of 145 protein interaction domains (*e.g.* SH2, SH2, WW, PDZ) and demonstrated the specific binding of proteins from cellular extracts to elements on the array [35]. Protein kinases also represent a class of proteins that are of great interest for pharmaceutical development; a comprehensive array of these proteins would certainly have utility for development of new kinase inhibitors as well as for profiling the specificity of existing inhibitors. Zhu *et al.* purified 119 of the 122 kinases from the budding yeast *S. cerevisiae*, screened the kinases with 17 substrates in microwell-type arrays fabricated from silicone elastomer, and demonstrated kinase inhibition [23]. Clearly, the next step is to create a human kinase array in a similar manner.

6 Conclusions and future prospects

A number of interesting parallels can be drawn between the development of DNA microarrays and that of proteome microarrays. Initially, DNA microarrays were lacking in comprehensive content, required expensive equipment and reagents for manufacturing and detection, and had limited applications. As the utility of the technology for understanding biology became more obvious, the demand for comprehensive, yet less expensive, DNA arrays grew. Today, DNA microarray technology is an integral part of the field of functional genomics. While proteome microarrays have obviously not reached the stage of maturity of DNA microarrays, recent developments have shown that the technical barriers holding back the technology can be surmounted with a sufficiently focused effort. Comprehensive protein content has either been created or is in the process of being created, the production of whole proteome microarrays has been reduced to practice, and numerous useful applications have been described. The next generation of proteome microarrays, possibly consisting of large numbers of human proteins, is clearly not far off.

The most exciting prospect for the next generation of proteome microarrays lies in the utility of these arrays to address the drug discovery and development process. Human sub-proteome arrays, such as tissue-specific and disease-specific protein collections, will enable rapid characterization of disease pathways for target and biomarker identification, validation as well as antibody and drug profiling. Proteome microarrays from model organisms and animal models of interest in drug research and development will augment the target discovery, validation, selection, differentiation, and metabolic profiling traditionally pursued with these systems. Pathogen proteome arrays will also have applications in target identification and vaccine development and monitoring. The application of this technology to the steps in the drug discovery and early development processes that are the key bottlenecks in the drug development continuum will lead to more efficient and predictive selection and development of new drugs.

Received May 15, 2003

7 References

- [1] Horak, C. E., Snyder, M., *Funct. Integr. Genomics* 2002, 2, 171–180.
- [2] Nutt, C. L., Mani, D. R., Betensky, R. A., Tamayo, P. *et al.*, *Cancer Res.* 2003, 63, 1602–1607.
- [3] Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, 98, 15149–15154.

- [4] Singh, D., Febbo, P. G., Ross, K., Jackson, D. G. *et al.*, *Cancer Cell* 2002, 1, 203–209.
- [5] Butte, A., *Nat. Rev. Drug Discov.* 2002, 1, 951–960.
- [6] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, *Nat. Biotechnol.* 1999, 17, 994–999.
- [7] Mitchell, P., *Nat. Biotechnol.* 2002, 20, 225–229.
- [8] James, P., *Biotechniques* 2002, Suppl, 4–10, 12–13.
- [9] Albala, J. S., *Expert Rev. Mol. Diagn.* 2001, 1, 145–152.
- [10] Schweitzer, B., Kingsmore, S. F., *Curr. Opin. Biotechnol.* 2002, 13, 14–19.
- [11] Zhu, H., Bilgin, M., Bangham, R., Hall, D. *et al.*, *Science* 2001, 293, 2101–2105.
- [12] Hudson, J. R., Jr., Dawson, E. P., Rushing, K. L., Jackson, C. H., *et al.*, *Genome Res.* 1997, 7, 1169–1173.
- [13] Snyder, M., Gerstein, M., *Science* 2003, 300, 258–260.
- [14] Reboul, J., Vaglio, P., Rual, J. F., Lamesch, P. *et al.*, *Nat. Genet.* 2003, 34, 35–41.
- [15] Bussow, K., Nordhoff, E., Lubbert, C., Lehrach, H., Walter, G., *Genomics* 2000, 65, 1–8.
- [16] Brizuela, L., Braun, P., LaBaer, J., *Mol. Biochem. Parasitol.* 2001, 118, 155–165.
- [17] Brizuela, L., Richardson, A., Marsischky, G., Labaer, J., *Arch. Med. Res.* 2002, 33, 318–324.
- [18] Lennon, G., Auffray, C., Polymeropoulos, M., Soares, M. B., *Genomics* 1996, 33, 151–152.
- [19] Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, 99, 16899–16903.
- [20] Strausberg, R. L., Feingold, E. A., Klausner, R. D., Collins, F. S., *Science* 1999, 286, 455–457.
- [21] Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M., Fields, S., *Nature* 2003, 422, 208–215.
- [22] Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L. *et al.*, *Methods Enzymol.* 2000, 328, 575–592.
- [23] Zhu, H., Klemic, J. F., Chang, S., Bertone, P. *et al.*, *Nat. Genet.* 2000, 26, 283–289.
- [24] Braun, P., Hu, Y., Shen, B., Halleck, A. *et al.*, *Proc. Natl. Acad. Sci. USA*, 2002, 99, 2654–2659.
- [25] Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L. *et al.*, *J. Cell Biochem.* 2000, 80, 187–191.
- [26] Holt, L. J., Bussow, K., Walter, G., Tomlinson, I. M., *Nucleic Acids Res.* 2000, 28, E72.
- [27] Afanassiev, V., Hanemann, V., Wolf, S., *Nucleic Acids Res.* 2000, 28, E66.
- [28] Guschin, D., Yershov, G., Zaslavsky, A., Gemmell, A. *et al.*, *Anal. Biochem.* 1997, 250, 203–211.
- [29] MacBeath, G., Schreiber, S. L., *Science* 2000, 289, 1760–1763.
- [30] Joos, T. O., Schrenk, M., Hopfl, P., Kroger, K. *et al.*, *Electrophoresis* 2000, 21, 2641–2650.
- [31] Haab, B. B., Dunham, M. J., Brown, P. O., *Genome Biol.* 2001, 2, RESEARCH0004.
- [32] Mendoza, L. G., McQuary, P., Mongan, A., Gangadharan, R. *et al.*, *Biotechniques* 1999, 27, 778–788.
- [33] Madoz-Gurpide, J., Wang, H., Misek, D. E., Brichory, F., Hanash, S. M., *Proteomics* 2001, 1, 1279–1287.
- [34] Knezevic, V., Leethanakul, C., Bichsel, V. E., Worth, J. M. *et al.*, *Proteomics* 2001, 1, 1271–1278.
- [35] Espejo, A., Cote, J., Bednarek, A., Richard, S., Bedford, M. T., *Biochem. J.* 2002, 367, 697–702.
- [36] Kukar, T., Eckenrode, S., Gu, Y., Lian, W. *et al.*, *Anal. Biochem.* 2002, 306, 50–54.
- [37] Lesaichere, M. L., Lue, R. Y., Chen, G. Y., Zhu, Q., Yao, S. Q., *J. Am. Chem. Soc.* 2002, 124, 8768–8769.
- [38] Day, Y. S., Baird, C. L., Rich, R. L., Myszka, D. G., *Protein Sci.* 2002, 11, 1017–1025.
- [39] Day, Y. S., Myszka, D. G., *J. Pharm. Sci.* 2003, 92, 333–343.
- [40] Myszka, D. G., Rich, R. L., 2000, 3, 310–317.
- [41] Rich, R. L., Hoth, L. R., Geoghegan, K. F., Brown, T. A. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, 99, 8562–8567.
- [42] Ito, T., Chiba, T., Ozawa, R., Yoshida, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, 98, 4569–4574.
- [43] Uetz, P., Giot, L., Cagney, G., Mansfield, T. A. *et al.*, *Nature* 2000, 403, 623–627.
- [44] von Mering, C., Krause, R., Snel, B., Cornell, M. *et al.*, *Nature* 2002, 417, 399–403.
- [45] Gavin, A. C., Bosche, M., Krause, R., Grandi, P. *et al.*, *Nature* 2002, 415, 141–147.
- [46] Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D. *et al.*, *Nature* 2002, 415, 180–183.
- [47] Bader, G. D., Hogue, C. W., *Nat. Biotechnol.* 2002, 20, 991–997.
- [48] Mezzasoma, L., Bacarese-Hamilton, T., Di Cristina, M., Rossi, R. *et al.*, *Clin. Chem.* 2002, 48, 121–130.
- [49] Robinson, W. H., DiGennaro, C., Hueber, W., Haab, B. B. *et al.*, *Nat. Med.* 2002, 8, 295–301.
- [50] Robinson, W. H., Garren, H., Utz, P. J., Steinman, L., *Clin. Immunol.* 2002, 103, 7–12.
- [51] Le Naour, F., Brichory, F., Misek, D. E., Brechot, C. *et al.*, *Mol. Cell Proteomics* 2002, 1, 197–203.
- [52] Le Naour, F., Misek, D. E., Krause, M. C., Deneux, L. *et al.*, *Clin. Cancer Res.* 2001, 7, 3328–3335.
- [53] McNeel, D. G., Nguyen, L. D., Storer, B. E., Vessella, R. *et al.*, *J. Urol.* 2000, 164, 1825–1829.
- [54] Prasannan, L., Misek, D. E., Hinderer, R., Michon, J. *et al.*, *Clin. Cancer Res.* 2000, 6, 3949–3956.
- [55] Nabholz, J. M., Reese, D. M., Lindsay, M. A., Riva, A., *Clin. Breast Cancer* 2002, Suppl 2, S75–79.
- [56] Fang, Y., Frutos, A. G., Lahiri, J., *J. Am. Chem. Soc.* 2002, 124, 2394–2395.
- [57] Ge, H., *Nucleic Acids Res.* 2000, 28, e3.