the FEBS Journal

MINIREVIEW

Advances in functional protein microarray technology

Paul Bertone and Michael Snyder

Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA

Keywords
biochemical interactions; high-throughput proteomics; large-scale protein production; microarray analysis; protein arrays

Correspondence
P. Bertone, Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA
Fax: +1 203 432 3597
Tel: +1 203 432 6139
E-mail: paul.bertone@yale.edu

(Received 28 June 2005, revised 9 September 2005, accepted 14 September 2005)

doi:10.1111/j.1742-4658.2005.04970.x

Numerous innovations in high-throughput protein production and microarray surface technologies have enabled the development of addressable formats for proteins ordered at high spatial density. Protein array implementations have largely focused on antibody arrays for high-throughput protein profiling. However, it is also possible to construct arrays of full-length, functional proteins from a library of expression clones. The advent of protein-based microarrays allows the global observation of biochemical activities on an unprecedented scale, where hundreds or thousands of proteins can be simultaneously screened for protein–protein, protein–nucleic acid, and small molecule interactions. This technology holds great potential for basic molecular biology research, disease marker identification, toxicological reponse profiling and pharmaceutical target screening.

DNA microarrays have become ubiquitous in genomic research, evident by their widespread use in profiling gene expression patterns, mapping novel transcripts, detecting sequence mutations and deletions, and locating transcription factor binding sites. Although microarray experiments are invaluable for large-scale sequence analyses, little can be inferred from these studies about the functions of gene products. In contrast to the high-throughput (HT) experiments afforded by DNA arrays, those designed to elucidate the biochemical activities of encoded proteins have traditionally been carried out on single molecules. Recently, significant effort has been made towards adapting proteomic analytical methods for use with DNA microarray technologies to enable the elucidation of proteome-wide biochemical activities and interactions.

The large-scale characterization of protein complexes generally involves (a) the separation of complex protein samples and (b) the subsequent identification of individual proteins. Among the methods currently available for proteome analysis are 1D gel electrophoresis (1D-GE) and 2D gel electrophoresis (2D-GE), MS, affinity chromatography, N-terminal Edman protein sequencing, metal affinity shift assays, ^{15}N isotope labeling, and tandem affinity purification (TAP) tagging [1] to isolate protein complexes from cell extracts. Of these, protein separation by 2D-GE and subsequent identification using MS have remained the two core technologies for large-scale proteomics. The 2D-GE method entails the separation of complex protein mixtures by molecular charge in the first dimension and by mass in the second dimension. Although recent advances in 2D-GE have improved the resolution and reproducibility, the technique remains difficult to automate in a HT setting. For this reason, alternative approaches that obviate the need for gel separation, such as multidimensional protein identification technology [2], have gained popularity for large-scale proteomics efforts and are able to generate a comprehensive catalog of proteins present in complex cell extracts. HT protein analysis is expected to accelerate with the introduction of new robotic liquid

Abbreviations
2D-GE, 2D gel electrophoresis; CV, coefficient of variation; ESI, electrospray ionization; HT, high-throughput; RCA, rolling circle amplification; SELDI, surface-enhanced laser desorption/ionization; TAP, tandem affinity purification.

chromatography systems and high-resolution analysis methods such as top-down Fourier transform mass spectrometry [3].

The use of MS for protein identification has come into wider use with the advent of soft ionization techniques, such as electrospray ionization (ESI) [4] and MALDI [5,6]. Additionally, the emergence of hybrid methods incorporating electrospray technology [7] with quadrupole time-of-flight mass spectrometer tandem mass analysis (ESI Q-TOF MS/MS) allows more accurate identification of specific proteins through the generation of collision-induced dissociation (CID) spectra that yield accurate sequence tags from protonated peptide ions. Recently, the range of mass spectrometric applications has been extended by other tandem approaches such as as MALDI TOF–TOF MS/MS [8,9] and MALDI Q-TOF MS/MS [10].

Beyond the identification of individual proteins, quantitative analysis of complex samples can be accomplished through the use of surface-enhanced laser desorption/ionization (SELDI) [11,12]. This approach incorporates standard ionization techniques on different surfaces, comprising a solid support modified with various chemical or biological bait molecules. These may include antibodies, proteins, nucleic acids and metal ions. The differential surface capture of solubilized protein samples provides a unique signature that varies depending on protein composition. Unlike MS techniques, SELDI is not able to identify specific proteins in a complex sample. This is expected to change in the near future through the combined use of SELDI technology with tandem mass spectrometers.

Regardless of the methods used to measure and catalog an organism's proteome, the majority of detection and quantification methods result in denaturing of the protein samples and thus functional characterization is not possible. To obtain detailed functional information, proteins must be cloned and expressed in recombinant form and subjected to systematic biochemical analyses. Martzen *et al.* [13] developed a multiplexed assay to characterize protein function on a large scale through a divide-and-conquer strategy. The approach entails the generation of pooled purified protein samples, which are then assayed in parallel for various biochemical activities. Individual proteins that exhibit specific activities in a pooled sample are identified through a series of recurrent analyses of subpools. This procedure allows the rapid identification of proteins that participate in various biochemical pathways via a divisive search through subpopulations of functional proteins. An advantage of this method is that biochemically active, multimeric protein complexes may be identified *in vitro*; however, this does not represent an exhaustive combinatorial search as only proteins that happen to be present in a given pool may interact.

Development of addressable protein arrays

The ultimate goal of protein microarray development is to construct ordered arrays of individual proteins to assess biochemical activities on a single-molecule basis. One solution has entailed the use of analytical arrays for the purpose of protein profiling. These typically comprise a library of peptides or antibodies arrayed on the surface of a glass microscope slide [14]. In general, protein profiling entails the measurement of binding specificity, affinity, or abundance of proteins in a biological sample. To address this with microarrays involves the construction of an array whose elements are designed to capture, and thereby measure the binding specificity of, proteins present in a complex mixture such as a cell lysate or serum sample. Array features are typically antibodies that are mechanically printed as independent purified samples, but may also be peptides that are synthesized *in situ*. The latter technology holds promise for the rapid screening of high-affinity binding sequences and the identification of potential drug targets.

The synthesis of individual peptides *in situ* can be accomplished via photolithography, originally described by Fodor *et al.* [15] and later applied to oligonucleotide-based microarray fabrication [16]. Photolithographic peptide arrays involve the use of photolabile phosphoramitides that enable deprotection of the Boc group from an amino moiety, allowing polymer synthesis to proceed at discrete array locations when illuminated by a laser. Using this method, an array of 1024 peptides was constructed and probed with a mAb. Pellois *et al.* [17] developed an alternative technique that can make use of natural amino acids for peptide synthesis, using a photogenerated acid to chemically deprotect the growing polymer chain upon exposure to light.

A simpler approach to photolithography is described by the SPOT protocol [18], in which a series of activated amino acids is mechanically deposited onto a porous surface, thereby building the desired peptides sequentially. SPOT is based on conventional solid-phase synthesis chemistry, and may therefore be more accessible in terms of implementation.

Recently, Li *et al.* [19] described a novel approach to homogeneous *in situ* peptide synthesis based on a common cyclic peptide scaffold. The procedure involves the deprotection of NPPOC phosphoramitide

groups to affect the addition of side-chains to a universal core molecule, which is presythesized and applied to a silanized glass slide in a uniform manner. A library of individual peptides can then be synthesized *in situ* using maskless photolithography [20], in which a spatially addressable array is fabricated through successive photodeprotection using a bank of digitally controlled micromirrors.

To date, most protein microarray systems have been based on contact-printed antibody libraries that are used for profiling complex analyte mixtures. The most widely adopted strategy consists of a multiplex adaptation of the classical antibody sandwich assay, where a pair of antibodies binds two discrete recognition surfaces on each protein [21,22]. In this procedure, an array of antibodies, which have been immobilized through covalent bonding to a silanized glass surface, is probed with various analytes. A second biotinylated antibody is then applied which binds to captured analytes, forming an immune complex. The second antibody is finally detected with a universal antibiotin antibody conjugated to a fluorophore. This approach has been used to great effect for the simultaneous detection of multiple cytokine or chemokine levels in biological samples [23]. The highly specific antibody–complex recognition is ideal for detecting low-abundance cytokines and holds much potential for clinical diagnostic applications and discovery of therapeutic drug targets [24].

A variation on this experiment utilizes rolling circle amplification (RCA) to enhance the fluorescence signals emitted from the immune complex [25]. In this method, the detection antibody is not fluorescence-labeled but is instead conjugated to oligonucleotides that serve to prime the RCA reaction. Complementary circular oligonucleotides are extended with DNA polymerase, producing RCA products that consist of tandem repeats. Because these repeat sequences provide many redundant hybridization targets, an amplification in fluorescence signal is achieved when the RCA products are detected with fluorescence-labeled complementary DNA probes.

An impediment to the further development of antibody array technology lies in the availability of high-quality antibodies against the individual proteins in a complex sample. At present it remains unfeasible to obtain hundreds or thousands of different antibodies that can recognize and capture various proteins with high affinity and specificity – factors that are essential for preventing cross-reactivity. The problem is compounded when considering multiplex sandwich immunoassays, where two highly specific antibodies must be obtained for each protein captured. These must also recognize two different regions of the protein, each without masking the other binding domain.

Although this issue may eventually be addressed by new methods of HT antibody generation, several alternatives to protein capture have been developed that do not rely on antibody recognition. Among the most innovative of these involves imprinting technology to create artificial molecular recognition surfaces [26,27]. Peptides that correspond to signal sequences in various target proteins are used as a structural scaffold, around which polymerizable monomers are allowed to self-assemble. The monomers are crosslinked in place and the template molecule is stably removed from the complex. The cavity or imprint that remains is shape-complementary to the original template and will therefore bind identical structures with high affinity. This technology promises to accelerate antibody array development by increasing the throughput of artificial epitope production, although at present imprinting is unable to mimic larger functional proteins or other macromolecular structures.

The antibody array represents an excellent platform for HT protein profiling. However, the large-scale study of protein biochemistry using the microarray format requires the development of arrays of full-length, functional proteins. HT protein production, combined with technologies shared with the proven DNA microarray format, allow the simultaneous analysis of thousands of protein activities in a single experiment. In addition to enabling the inverse profiling experiment where functional proteins can be interrogated with individual antibodies [28], protein microarrays enable a wide range of biochemical assays in response to any solution-phase binder.

Array technologies for functional protein analysis

The principal challenges in functional protein array development comprise (a) creation of a comprehensive expression clone library, (b) HT protein production, including expression, isolation and purification, (c) adaptation of DNA microarray technology to accommodate protein substrates, (d) ensuring the stability of arrayed proteins, and (e) reduction of inter- and intra-slide variability of protein concentration between deposited samples.

By far the greatest obstacle in developing functional protein microarrays is the construction of a comprehensive expression clone library from which a large number of distinct protein samples can be produced (Fig. 1). In building a clone library, it is desirable to construct recombinant genes where fusion proteins can be produced for the purpose of affinity purification and/or slide surface attachment. Cloning the genes of

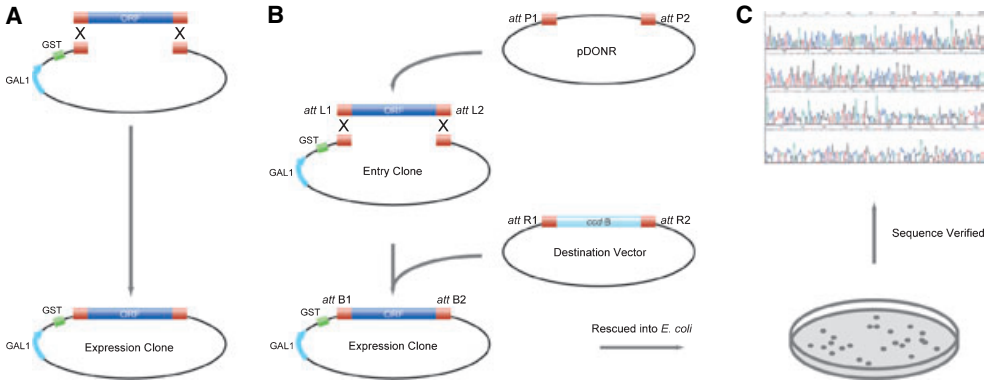

Fig. 1. Example of high-throughput cloning strategies for expression library construction. (A) The construction of in-frame gene fusions used by Zhu *et al.* [35,38] relies on gap repair-mediated recombination in yeast. Amplified ORFs are mixed with a linearized vector with ends that are identical to those of the amplified DNA, and the mixture is transformed into yeast cells where the DNA is integrated. (B) Recombination cloning based on the Gateway λ-phage integration/excision system (Invitrogen, Carlsbad, CA, USA). Recombination between the λ *att* P and host *att* B sites yields an integrated phage with ends *att* L and *att* R. During excision, recombination between *att* L and *att* R sites regenerates *att* P and *att* B. Amplified ORFs with *att* B1 and *att* B2 sites undergo recombination with a donor vector containing *att* P1 and *att* P2 sites, producing the entry vector with *att* L1 and *att* L2 sites. Subsequent recombination with a destination vector containing *att* R1 and *att* R2 sites allows transfer of the ORF into the expression vector, regenerating *att* B sites. (C) Expression clones are rescued into *Escherichia coli* and verified by DNA sequencing.

interest with an inducible promoter allows individual proteins to be expressed in high abundance. HT purification can be accomplished with the addition of C- or N-terminal tags, such as glutathione-S-transferase or the IgG-binding domain of Protein A. The incorporation of fusion tags also facilitates the verification of clone inserts by sequencing across the vector–insert junction. It is highly desirable to transform the expression vector into a homologous or related cell type, ensuring the proper delivery of the protein product to the secretory pathway and hence correct folding and post-translational modification of each recombinant protein.

Prototype formats for functional protein arrays

A critical aspect of the development of arrays of functional proteins is the selection of an experimental support and its associated method of surface attachment and immobilization of proteins. If proteins can be immobilized without disrupting their native conformations, they are likely to remain biologically active *in vitro*. The method of immobilization will also greatly influence the orientation in which proteins are attached to the support surface. Engineering a common point of attachment for all samples, typically through the inclusion of affinity tags, ensures that at least a subset of proteins maintain a uniform presentation to solution-phase binders. The array format must be compatible with appropriate detection instruments and exhibit a wide dynamic range of intensity values associated with protein binding or catalysis events. Additionally, nonspecific binding of labeled samples should be minimized in order to reduce the background and increase the signal-to-noise ratio of the experimental platform.

The materials traditionally associated with conventional protein assays are often incompatible with robotic arrayers, cannot provide the sensitivity or dynamic range expected from microarray experiments, or contribute high fluorescence background resulting in low signal-to-noise ratios. To circumvent these problems, a number of innovative platforms have been explored for prototyping the protein array format. Among these, two make use of an agarose or acrylamide gel situated on glass slides, thereby combining the utility of a solid support with the loading and binding capacity of a porous gel matrix. These methods were originally devised to increase the potential loading capacity of DNA samples on a planar microarray surface, but are generally applicable to a variety of substrates, including nucleic acids, proteins and small molecules.

Guschin *et al.* [29] developed arrays of polyacrylamide gel pads on a hydrophobic glass surface using a combination of gel photopolymerization and manual contact pin deposition. Three test proteins were loaded onto the arrays – mouse IgG1, rabbit IgG and BSA. Immunoanalysis of fluorescein isothiocyanate

(FITC)-labeled IgG against mouse IgG1 demonstrated selective binding when the arrays were imaged with a fluorescence microscope. In a related study, Mirzabekov and coworkers [30,31] experimented with a copolymerized acrylamide–bisacrylamide substrate, producing arrays of discrete gel pads of between 10 and 100 μm in diameter. In a later study, Kiyonaka *et al*. [32] developed a method of supramolecular gel formation as a spontaneous process that does not require additional polymerization steps, using the approach to develop a sensor array of fluorescent metal anion and cation receptors in a glycosylated amino acetate hydrogel matrix [33].

A study by Afanassiev *et al*. [34] explored the use of a thin, uniform layer of agarose film on a glass surface to achieve a similar effect. Activated agarose containing $NaIO_4$ was applied to silanized slides, and samples were mechanically deposited using a robotic microspotter after the gel had solidified. A limited application of this platform to protein-based assays demonstrated the binding of mAbs to immobilized recombinant human BAD protein, and reciprocally, of recombinant human (rh)BAD to immobilized antibodies in a sandwich immunoassay. Another approach involved the use of a liquid silica compound to create flexible sheets of microwell arrays in which biochemical reactions are performed en masse. Zhu *et al*. [35] devised a system of casting a silicone elastomer (polydimethylsiloxane) onto a reusable mold of laser-milled acrylic. After the microwell sheets had cured, various molecules were immobilized to the interior surface of the wells using the chemical crosslinking agent, 3-glycidoxypropyltrimethoxysilane.

Aside from demonstrating the technical feasibility of protein immobilization and binding to various molecules *in vitro*, it is essential to conduct biologically relevent experiments if protein microarrays are to become an established research platform. To explore the utility of microwell arrays for the detection of enzymatic activities, Zhu *et al*. [35] focused on 119 protein kinases from the budding yeast *Saccharomyces cerevisiae* (95 known and 24 uncharacterized). Each of the protein kinases was expressed in recombinant form and purified, then assayed for the ability to phosphorylate 17 substrate proteins. A total of 17 arrays were fabricated and one of the substrate proteins was immobilized in every well; a different protein kinase was delivered into each well in the presence of γ-ATP to determine which kinases were capable of phosphorylating the given substrate protein. The arrays were incubated and then washed such that all free enzyme was removed, and the signal from the radiolabeled proteins in each well were quantified using a phosphoimager. The possibility that measurements originated from the autophosphorylation of the kinases themselves was discounted, as in each case the substrates were bound in the wells while the enzymes remained free in solution; the wells were cleared of this reaction mixture and washed prior to imaging.

In addition to detecting expected phosphorylation activities, 27 protein kinases were found to be capable of phosphorylating tyrosine after incubation with a synthesized poly(tyrosine-glutamate) peptide. This finding was significant as yeast protein kinases are generally known to phosphorylate only serine or threonine. Phylogenetic analysis revealed sequence similarities among the tyrosine-phosphorylating kinases that were specific to several amino acids oriented in the catalytic cleft and substrate-binding domain of the enzyme. These appeared exclusively in the kinases that phosphorylated tyrosine and not in those that phosphorylated serine or threonine alone. Although it is likely that most yeast protein kinases will preferentially phosphorylate serine or threonine *in vivo*, this study demonstrated that protein arrays are sensitive enough to reveal previously uncharacterized biochemical properties in a HT assay.

On balance, each of these formats retains a number of important properties for proteomic experiments. Microwells provide the ability to preserve native protein function by carrying out reactions in an aqueous environment [35], while the hydrated gel matrix approach [29–34] also immobilizes the proteins to some degree. Any potential cross-contamination between arrayed proteins is eliminated owing to physical barriers between them [35] or the spatial separation of discrete gel pads [29–33]. Each platform affords a far greater loading capacity than the planar surface of a glass slide, allowing experiments to be performed at varying substrate concentrations. Finally, the microwell format offers the potential to perform complex, multistage experiments in solution if reaction mixtures are exchanged using microfluidic robotics.

Contact-printed functional protein microarrays

Although these approaches offer numerous technical advantages, increasing attention has been paid to adapting existing glass-slide microarray technologies for use with proteins. As a result, the practice of printing directly onto chemically treated glass surfaces, apart from their indirect function as a solid support, is now in wider use for protein arrays. The principal motivating factor for using glass slides is to take advantage of robotic microspotting arrayers and laser scanners that

Table 1. Surface chemistries for glass slide protein microarrays.

Surface chemistry	Protein attachment	Protein orientation	Modifications required
Epoxy	Covalent crosslinking	Random	None
Aldehyde	Covalent crosslinking	Random	None
Poly(L-lysine)	Adsorption	Random	None
Nitrocellulose	Adsorption, absorption	Random	None
Poly(vinylidene difluoride)	Adsorption, absorption	Random	None
Avidin	Affinity binding	Random	Biotinylation
Nickel-nitrilotriacetic acid	Affinity binding	Uniform	His_6 fusion

have become commonplace for DNA microarray fabrication and image acquisition, respectively. The standard microarray format therefore affords the opportunity for a broader range of investigators to adopt HT proteomic formats using existing DNA microarray construction and analysis equipment.

A number of different chemical treatments are suitable for the immobilization of proteins on glass, including poly(L-lysine), aldehyde, nickel, epoxy, avidin and nitrocellulose (Table 1). The choice of surface chemistry will determine the method of protein attachment to the functionalized support, which in turn will influence whether the proteins require modification prior to arraying. Amine-reactive coatings (e.g. 3-aminopropyltriethoxysilane) do not require proteins to be modified for effective immobilization, as covalent attachment is achieved through the random crosslinking of amine groups. Protein attachment to either nickel- or avidin-coated slides is made through affinity binding to histidine residues or biotin, respectively, and therefore requires a recombinant approach to construct fusion proteins [36].

The advantage of noncovalent immobilization is that because attachment takes place exclusively at the affinity tag, a significant subset of proteins will be immobilized in a presumably uniform orientation where their functional domains are exposed to solution and available to interact with the labeled sample. Other membrane-based solutions, such as fluorescent array surface technology slides (Schleicher & Schuell BioScience, Dassel, Germany), facilitate protein immobilization through passive adsorption on a nitrocellulose surface, and in many cases yield superior signal-to-noise ratios in comparison with poly(L-lysine) or aldehyde-treated slides (Fig. 2).

Robotic printing of protein microarrays

Constructing microarrays from purified proteins is more challenging than building their DNA counterparts. Aside from the complexities of preparing hundreds or thousands of individual protein samples, the proteins must remain active during the manufacture and probing of microarrays. This entails keeping protein samples in cold storage for as long as possible, and arraying them in a glycerol solution to maintain their native state on the slide surface. The procedures used for printing protein arrays are similar to those developed for DNA arrays [37], but with several important differences, discussed below.

DNA samples are arrayed in a high-humidity chamber by a robotic microspotter. Although the same equipment can be used to print protein arrays, individual samples are typically handled in a 30–35% glycerol solution to prevent denaturing, and arrayed at concentrations in the range of 0.3–1 $mg{\cdot}mL^{-1}$. As glycerol is hygroscopic, care must be taken to reduce the ambient humidity to 28–30%; this differs from the 45% humidity environment in which DNA arrays are usually printed. Printed spots of protein–glycerol solution will absorb moisture from the air in high-humidity environments and begin to expand on the slide surface. This can lead to cross-contamination of samples if the proteins are arrayed at high spatial density.

The diameter and pitch (the spacing between samples) of spots can remain similar to the parameters used for printing DNA microarrays, although this will depend on the type of slides used and the viscosity of the buffer. For example, epoxy-coated slides provide more efficient protein attachment, which generally allows for higher-density printing. Pilot studies should be carried out to determine the appropriate printing density for a given sample preparation and surface chemistry, adjusting the pitch from a more conservative spacing of 1000 μm down to a minimum of 300 μm. Depending on the number of pins used (typically between 16 and 48), a 300 μm pitch allows up to 20 000 individual protein samples to be printed on a standard microscope slide, yielding $\approx$1600 features per cm^2 where each feature is 150 μm in diameter. (Fig. 3C).

As with DNA arrays, fabrication precision should be assessed relative to the printing consistency across

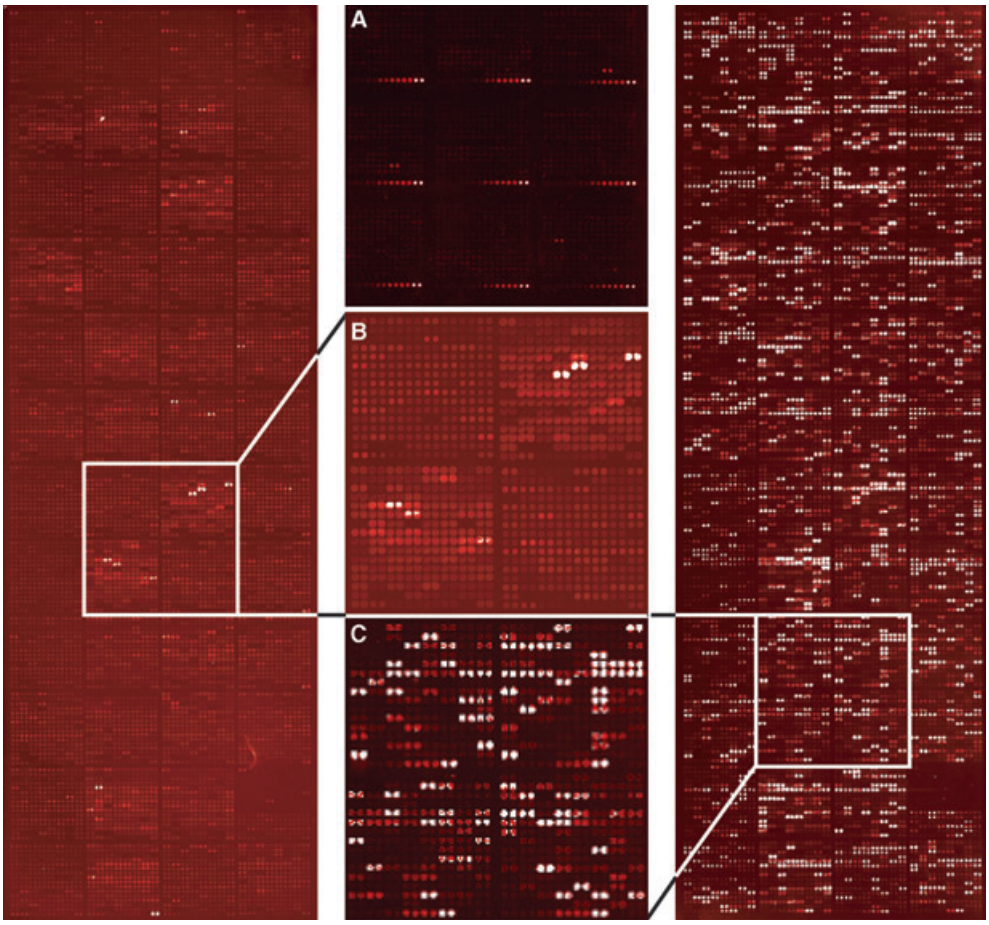

Fig. 2. Contact-printed functional protein microarrays. (A) Control proteins are spotted in a dilution series and illuminated with fluorescein-labeled antibodies. (B) Protein–protein interaction assays reveal specific binding targets. (C) Detection of lectin modifications across the yeast proteome.

all of the features on the array, consistency between features from multiple arrays, and the uniformity of sample concentrations. This is typically measured as the coefficient of variation (CV), which estimates the reproducibility of printing as the standard deviation over the mean fluorescence signal intensity. For functional protein arrays prepared from individual purified samples, intra-array CV measurements fall in the range of 15–20%, and are a measurement of both the variability in protein concentrations and the pin-to-pin consistency of sample deposition during the printing cycle. Inter-array CVs over all protein samples from replicate slides are typically lower, in the range of 12–15%. Much of the variation in sample concentration is attributed to the expression and purification of low-abundance proteins, whereas printing variation can arise as a function of the order in which samples are arrayed during long print runs.

The incorporation of an epitope tag, such as glutathione-S-transferase, is invaluable for HT affinity purification, and also allows the global quantification of protein signals by probing the array with a fluor-conjugated antibody [38]. This measurement is used to calculate the ratio of the probe-binding signal relative to the baseline fluorescence intensity of each spot on the array. As the relative concentrations and deposited amounts of protein samples will vary slightly, computing the fluorescence values in terms of intensity ratios enables the application of statistical methods similar to those used for DNA microarray data analysis [39].

Detecting biochemical interactions

Depending on the type of assay being performed, protein–substrate binding or catalysis can be detected via fluorescence labeling, chemiluminescence, radioisotope

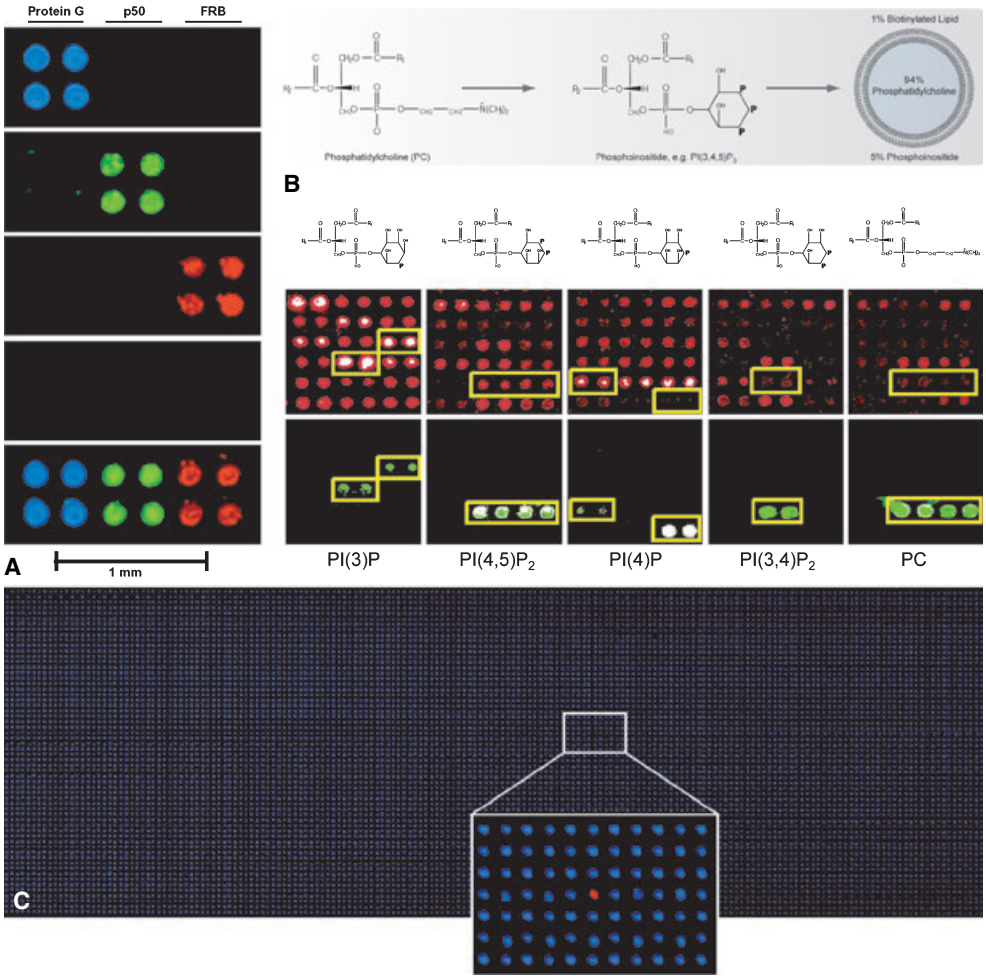

Fig. 3. Detecting biochemical interactions on protein arrays. (A and C) Protein–protein interactions were detected with fluorescein-labeled antibodies to specific proteins. In this case, BODIPYFL–IgG, Cy3–IκBα and Cy5–FKBP12 identify the presence of protein G, p50 and FKBP–rapamycin-associated protein, respectively [40]. (B) Phosphoinositide-binding specificity to functional proteins is determined in a phospholipid probing assay. The detection of protein–protein and protein–lipid binding [38] involves a two-step process. First, the sample used to probe the array is biotinylated using a simple crosslinking method. To perform the experiment, the microarray is incubated with a solution containing the biotinylated sample, washed, then incubated a second time in the presence of a fluorescent streptavidin molecule. The actual protein interactions are formed during the first incubation; subsequent binding of the streptavidin–fluor conjugate to the biotin label then allows these interactions to be detected using a microarray scanner. A simpler approach can also be employed by attaching a fluorophore directly to the solution-phase binder used to probe the array. Proteins can be labeled using amine-reactive dyes, where the succinimidyl ester moieties react with the primary amines of the protein to form stable conjugates.

labeling, or label-free methods, such as surface plasmon resonance imaging, atomic force microscopy or reflectometric interference spectroscopy. Fluorescence labeling is generally preferred as a safe and efficient method for HT analysis that is compatible with current microarray laser scanners. The detection of protein binding, as described in Zhu *et al*. [38], involves a two-step process similar to that used for multiplex sandwich immunoassays. First, the sample used to probe the array is biotinylated using a simple crosslinking method. To perform the experiment, the microarray is incubated with a solution containing the biotinylated sample, washed, then incubated a second time in the presence of a fluorescent streptavidin molecule. The actual protein interactions are formed during the first incubation; subsequent binding of the streptavidin–fluor conjugate to the biotin label then allows specific interactions to be detected using a standard microarray scanner.

A simpler approach can be employed by attaching a fluorophore directly to the solution-phase binder used to probe the array, provided that the label does not

mask an active site or binding domain of the sample molecule. Proteins can be labeled directly using amine-reactive dyes, where the succinimidyl ester moieties react with the primary amines of the protein to form stable conjugates. The labeled protein can then be purified using size-exclusion spin columns and applied to the array for the detection of protein–protein interactions. Of course, other molecules can be applied to functional protein arrays to assess different molecular interactions, such as protein–DNA binding. Fluorescence labeling of DNA can be achieved in a variety of ways, such as through direct incorporation of fluorescent-labeled nucleotides during reverse transcription, or by secondary conjugation of amine-reactive succinimidyl esters to 5-(3-aminoallyl)-dUTPs.

Functional protein arrays in use

The first study reporting the use of contact-printed, glass-slide protein arrays was described by MacBeath & Schreiber [40], who investigated the binding activities of three known pairs of interacting proteins (Fig. 3A). One protein of each pair was printed in quadruplicate onto aldehyde slides, and the arrays were probed with the labeled partners. The group also explored various feature densities of printed protein samples, successfully arraying a single protein as 10 800 discrete features on a standard microarray slide (Fig. 3C). Importantly, the researchers were able to quantify the concentrations of the bound and solution-phase proteins necessary to carry out the experiments. Sample concentrations between 100 $\mu g \cdot mL^{-1}$ and 1 $mg \cdot mL^{-1}$ were found to be suitable for protein immobilization and detection, whereas solution-phase proteins at concentrations of ≈ 12.5 pM yielded fluorescence signals that scaled linearly over four orders of magnitude. Thus, these experiments demonstrated the feasibility of arraying proteins in a standard microarray format and at feature densities comparable with those of contact-printed DNA arrays.

A subsequent study by Zhu *et al.* [38] described the development of a yeast proteome microarray containing the full-length, purified expression products of over 93% of the organism's complement of 6280 protein-coding genes. A total of 5800 ORFs were cloned as glutathione-S-transferase::His_6 fusions, and expressed in their native cells under a Gal-inducible promoter. Following HT purification, each protein sample was printed in duplicate onto glass slides using a standard robotic microspotter, at a feature density of 13 000 samples per array. This work represented the first systematic cloning and purification of an entire eukaryotic proteome, as well as the first large-scale functional protein array comprising discrete functional proteins. A number of protein attachment chemistries were evaluated, including aldehyde and nickel surface treatment. Aldehyde surfaces promote the covalent attachment of proteins by their N termini, although a concern with this method is that the random crosslinking of primary amines may disrupt the tertiary structures of some proteins. An alternative approach employed nickel-coated slides, attaching proteins via the incorporated His_6 tag. This yielded higher signal-to-noise ratios, presumably because fewer proteins were denatured after nickel attachment and their functional domains were more likely to be oriented uniformly away from the slide surface. Over 90% of the samples were found to yield significant fluorescence signals over background levels, in the range of 10–950 fg of protein.

Several different experiments were performed with the arrays, including a calmodulin-binding survey to assess protein–protein interactions and a large-scale screen for phospholipid-binding specificity. In the latter analysis, self-assembling phosphatidylcholine liposomes were incubated with five different phosphoinositide species and 1% biotin to form a series of biotinylated phospholipid probes (Fig. 3B). Proteome-wide microarray experiments identified 150 lipid-binding proteins, of which 52 were uncharacterized. In particular, the array-based interaction assays identified sets of proteins that demonstrated preferential binding to one or more phosphoinositides, apart from those that bound all of the phospholipids with equal affinity. As phosphoinositides are important components of membrane structures and second-messenger pathways, these types of protein–lipid interactions constitute a significant class of biochemical functionality.

Further applications of protein arrays

In many cases, complex proteomic experiments may benefit from the combined application of multiple analytical techniques. In a study by Huang *et al.* [41], a chemical genetic approach was used for the development of a screen for small-molecule inhibitors and enhancers of rapamycin, a peptide exhibiting a variety of functional roles in nutrient metabolism and cell cycle progression in eukaryotes. Following the identification of candidates, yeast proteome arrays [38] were interrogated with two biotinylated rapamycin-inhibitory molecules to identify their potential target proteins. Each was observed to bind multiple target proteins in addition to phosphatidylinositides. Analysis of various phenotypes revealed which binding events represented biologically meaningful interactions.

A combined experiment involving the use of protein and DNA microarrays was recently described for the systematic identification and location analysis of transcription factor proteins. Hall *et al*. [42] assayed the yeast proteome in search of novel DNA-binding proteins by probing the protein microarrays with labeled yeast genomic DNA. A total of 200 DNA-binding proteins were identified, half of which were known, or expected, to bind DNA. Of these, eight candidates were subjected to chromatin profiling via the ChIP-chip method [43], which is designed to identify transcription factor binding sites in genomic DNA. This technique entails the immunoprecipitation of specific protein–DNA complexes using antibodies against a native transcription factor protein or epitope tag. The immunoselected DNA is then sonicated, labeled and used to interrogate a DNA microarray representing intergenic or total genomic regions, thereby revealing the locations of transcription factor binding along a chromosome. ChIP-chip analysis on yeast intergenic arrays revealed Arg5,6, a mitochondrial enzyme involved in ornithine biosynthesis, to bind DNA at specific nuclear and mitochondrial loci. Altered gene expression levels were observed in Arg5,6 deletion mutants, further indicating its role as a transcriptional regulator.

Conclusions

The advent of protein-based microarrays allows the global observation of biochemical activities on an unprecedented scale, where hundreds or thousands of proteins can be simultaneously screened for protein–protein, protein–nucleic acid, and small molecule interactions, as well as post-translational modifications. Advances in HT separation techniques offer the potential for arraying proteins directly, although these technologies are at an early stage of development. Ouyang *et al*. [44] explored the separation, deposition and analysis of individual proteins from complex samples using ion soft landing and MS. In this study, a mixture of four proteins was introduced into a mass spectrometer by ESI. The proteins were then separated by their respective mass-to-charge ratios and independently deposited onto a gold substrate via ion beam focusing. Two methods of analysis were explored. Initially, the arrays were rinsed with a methanol/water solution and the mixture analyzed by ESI-MS. It was later found that the deposited proteins could be analyzed *in situ* by MALDI-MS.

Although automated mass-spectrometric separation and deposition remains a promising technology for protein microarray assembly, it is unclear whether soft-landing techniques can be applied en masse to the wide range of proteins constituting the entire proteome of an organism. At present, it seems clear that a well-curated expression clone library allows researchers to maintain the quality and identity of every protein under investigation. Additionally, MS may disrupt low-affinity biomolecular interactions, or even denature individual proteins, prior to experimental analysis. Ramachandran *et al*. [45] recently described an *in vitro* transcription and translation system that can generate proteins from a PCR product. The reactions are carried out in parallel directly on a solid support, through the use of affinity tags to anchor the products to the surface. The main drawback of this approach is that some proteins may not be properly folded and modified outside their native cellular environment. Ultimately, both of these developments offer the potential to reduce the time and complexity involved with the cloning and purification of individual proteins as a prerequisite to constructing functional protein microarrays.

In terms of assessing protein–protein interactions, protein microarray experiments can be qualitatively compared with the two-hybrid assay [46,47]. However, microarray experiments afford the ability to control the environmental parameters of an experiment, such as ion concentration, buffer pH, and the addition of reaction cofactors, in a precise and reproducible manner. Additionally, because microarrays exploit parallel interrogation to acquire many individual measurements on the same physical platform, the resulting data can be subjected to rigorous statistical analyses and tested for accuracy and reproducibility. Thus, protein-based microarrays provide the ability to characterize the biochemical functions of thousands of proteins in a parallel, quantitative format.

References

1 Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M & Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**, 1030–1032.
2 Washburn MP, Wolters D & Yates JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242–247.
3 Li Y, McIver RT Jr & Hunter RL (1994) High-accuracy molecular mass determination for peptides and proteins by Fourier transform mass spectrometry. *Anal Chem* **66**, 2077–2083.
4 Fenn JB, Mann M, Meng CK, Wong SF & Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71.

5 Karas M & Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299–2301.
6 Hillenkamp F & Karas M (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol* **193**, 280–295.
7 Shevchenko A, Keller P, Scheiffele P, Mann M & Simons K (1997) Identification of components of trans-Golgi network-derived transport vesicles and detergent-insoluble complexes by nanoelectrospray tandem mass spectrometry. *Electrophoresis* **18**, 2591–2600.
8 Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML & Burlingame AL (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* **2**, 552–558.
9 Baldwin MA, Medzihradszky KF, Lock CM, Fisher B, Settineri TA & Burlingame AL (2001) Matrix-assisted laser desorption/ionization coupled with quadrupole/orthogonal acceleration time-of-flight mass spectrometry for protein discovery, identification, and structural analysis. *Anal Chem* **73**, 1707–1720.
10 Shevchenko A, Loboda A, Shevchenko A, Ens W & Standing KG (2000) MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal Chem* **72**, 2132–2141.
11 Hutchens TW & Yip TT (1993) New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* **7**, 576–580.
12 Tang N, Tornatore P & Weinberger SR (2004) Current developments in SELDI affinity technology. *Mass Spectrom Rev* **23**, 34–44.
13 Martzen MR, McCraith SM, Spinelli SL, Torres M, Fields S, Grayhack EJ & Phizicky EM (1999) A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155.
14 Angenendt P, Glokler J, Murphy D, Lehrach H & Cahill DJ (2002) Toward optimized antibody microarrays: a comparison of current microarray support materials. *Anal Biochem* **309**, 253–260.
15 Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT & Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773.
16 Lipshutz RJ, Fodor SP, Gingeras TR & Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet* **21**, 20–24.
17 Pellois JP, Zhou X, Srivannavit O, Zhou T, Gulari E & Gao X (2002) Individually addressable parallel peptide synthesis on microchips. *Nat Biotechnol* **20**, 922–926.
18 Frank R (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports–principles and applications. *J Immunol Methods* **267**, 13–26.
19 Li S, Marthandan N, Bowerman D, Garner HR & Kodadek T (2005) Photolithographic synthesis of cyclic peptide arrays using a differential deprotection strategy. *Chem Commun* **5**, 581–583.
20 Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR & Cerrina F (1999) *Nat Biotechnol* **17**, 974–978.
21 Nielsen UB & Geierstanger BH (2004) Multiplexed sandwich assays in microarray format. *J Immunol Methods* **290**, 107–120.
22 Saviranta P, Okon R, Brinker A, Warashina M, Eppinger J & Geierstanger BH (2004) Evaluating sandwich immunoassays in microarray format in terms of the ambient analyte regime. *Clin Chem* **50**, 1907–1920.
23 Huang RP (2004) Cytokine protein arrays. *Methods Mol Biol* **264**, 215–231.
24 Huang RP, Yang W, Yang D, Flowers L, Horowitz IR, Cao X & Huang R (2005) The promise of cytokine antibody arrays in the drug discovery process. *Expert Opin Ther Targets* **9**, 601–615.
25 Schweitzer B, Roberts S, Grimwade B, Shao W, Wang M, Fu Q, Shu Q, Laroche I, Zhou Z, Tchernev VT *et al*. (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat Biotechnol* **20**, 359–365.
26 Ansell RJ, Ramstrom O & Mosbach K (1996) Towards artificial antibodies prepared by molecular imprinting. *Clin Chem* **42**, 1506–1512.
27 Titirici MM & Sellergren B (2004) Peptide recognition via hierarchical imprinting. *Anal Bioanal Chem* **378**, 1913–1921.
28 Michaud GA, Salcius M, Zhou F, Bangham R, Bonin J, Guo H, Snyder M, Predki PF & Schweitzer BI (2003) Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol* **21**, 1509–1512.
29 Guschin D, Yershov G, Zaslavsky A, Gemmell A, Shick V, Proudnikov D, Arenkov P & Mirzabekov A (1997) Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Anal Biochem* **250**, 203–211.
30 Vasiliskov AV, Timofeev EN, Surzhikov SA, Drobyshev AL, Shick VV & Mirzabekov AD (1999) Fabrication of microarray of gel-immobilized compounds on a chip by copolymerization. *Biotechniques* **27**, 592–594,596–598,600.
31 Rubina AY, Dementieva EI, Stomakhin AA, Darii EL, Pan'kov SV, Barsky VE, Ivanov SM, Konovalova EV & Mirzabekov AD (2003) Hydrogel-based protein microchips: manufacturing, properties, and applications. *Biotechniques* **34**, 1008–1014,1016–1020,1022.
32 Kiyonaka S, Sada K, Yoshimura I, Shinkai S, Kato N & Hamachi I (2004) Semi-wet peptide/protein array using supramolecular hydrogel. *Nat Mater* **3**, 58–64.
33 Yoshimura I, Miyahara Y, Kasagi N, Yamane H, Ojida A & Hamachi I (2004) Molecular recognition in a

supramolecular hydrogel to afford a semi-wet sensor chip. *J Am Chem Soc* **126**, 12204–12205.
34 Afanassiev V, Hanemann V & Wolfl S (2000) Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Res* **28**, E66.
35 Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, Smith D, Gerstein M, Reed MA & Snyder M (2000) Analysis of yeast protein kinases using protein chips. *Nat Genet* **26**, 283–289.
36 Nasoff M, Bergseid M, Hoeffler JP & Heyman JA (2000) High-throughput expression of fusion proteins. *Methods Enzymol* **328**, 515–529.
37 Shalon D, Smith SJ & Brown PO (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639–645.
38 Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T *et al*. (2001) Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105.
39 Luscombe NM, Royce TE, Bertone P, Echols N, Horak CE, Chang JT, Snyder M & Gerstein M (2003) ExpressYourself: a modular platform for processing and visualizing microarray data. *Nucleic Acids Res* **31**, 3477–3482.
40 MacBeath G & Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763.
41 Huang J, Zhu H, Haggarty SJ, Spring DR, Hwang H, Jin F, Snyder M & Schreiber SL (2004) Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc Natl Acad Sci USA* **101**, 16594–16599.
42 Hall DA, Zhu H, Zhu X, Royce T, Gerstein M & Snyder M (2004) Regulation of gene expression by a metabolic enzyme. *Science* **306**, 482–484.
43 Horak CE & Snyder M (2002) ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol* **350**, 469–483.
44 Ouyang Z, Takats Z, Blake TA, Gologan B, Guymon AJ, Wiseman JM, Oliver JC, Davisson VJ & Cooks RG (2003) Preparing protein microarrays by soft-landing of mass-selected ions. *Science* **301**, 1351–1354.
45 Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, Lau AY, Walter JC & LaBaer J (2004) Self-assembling protein microarrays. *Science* **305**, 86–90.
46 Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S & Sakaki Y (2000) Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two–hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA* **97**, 1143–1147.
47 Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al*. (2000) A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–627.