Trends in Biotechnology Vol. 20 No. 12 (Suppl.), 2002

Protein chips: from concept to practice

Young-Sam Lee and Milan Mrksich

A series of exciting reports over the past two years has established the usefulness of protein chips and made important advances in preparing protein arrays. However, several technical challenges must still be addressed to make these tools available to the wider community of researchers. Here, we discusses these challenges and survey recent opportunities for creating quantitative assays, preparing and immobilizing large numbers of proteins, using detection methods to analyze the results of chip-based experiments, and using informatics tools to interpret these results.

Protein chips have emerged as an exciting technology for the broad characterization of the activities and interactions of proteins. Like the gene-chip technology that motivated this development, protein chips will prove to be essential to researchers in biology and to commercial programs in drug discovery and diagnostics. However, and also like the gene chip, many years of research and development will be required to move from initial demonstrations of protein chips to a commercial off-the-shelf technology that all researchers can use. In this article, we outline several technical themes that will see important development and also survey recent strategies and concepts that will be important to these developments. We aim to describe the current status of protein chips for future users and to identify opportunities for researchers engaged in developing this technology.

First examples

A report by MacBeath and Schreiber two years ago established that proteins could be printed and assayed in a microarray format, and thereby had a large role in renewing the excitement for the prospect of a protein chip [1]. In this example, proteins were immobilized by reacting lysine side-chain amino groups with aldehyde-modified glass slides and the resultant chips were used to demonstrate assays for characterizing protein–protein binding interactions and kinase-mediated phosphorylation of immobilized proteins.

Shortly after this, Snyder and co-workers reported the preparation of a protein chip comprising nearly 6000 yeast gene products and used this chip to identify new classes of calmodulin- and phospholipid-binding proteins [2]. The proteins were generated by cloning the open reading frames and overproducing each of the proteins as glutathione-*S*-transferase- (GST) and His-tagged fusions [2,3]. The fusions were used to facilitate the purification of each protein and the His-tagged family were also used in the immobilization of proteins. This important work established that microarrays containing thousands of proteins could be prepared and used to discover binding interactions (Fig. 1). They also reported that proteins immobilized by way of the His tag – and therefore uniformly oriented at the surface – gave superior signals to proteins randomly attached to aldehyde surfaces [2].

Related work has addressed the construction of antibody arrays [4,5]. These arrays are intended for diagnostic applications that determine the amounts of multiple analytes in a sample rather than identifying novel protein–protein and enzyme–substrate interactions. In an early landmark report, de Wildt and Tomlinson immobilized phage libraries presenting scFv antibody fragments on filter paper to select antibodies for specific antigens in complex mixtures [4]. The use of arrays for this purpose greatly increased the throughput when evaluating antibodies, allowing nearly 20 000 unique clones to be screened in one cycle. Brown and co-workers extended this concept to create molecularly defined arrays wherein antibodies were directly attached to aldehyde-modified glass. They printed 115 commercially available antibodies and analyzed their interactions with cognate antigens with semi-quantitative results [5]. They found that many commercially available antibodies do not display sufficient affinity and specificity for their antigens. Kingsmore and co-workers used an analogous approach to prepare arrays of antibodies recognizing 75 distinct cytokines and, using the rolling-circle amplification strategy [6], could measure cytokines at femtomolar concentrations [7].

Taken together, these first examples over the past two years demonstrate the many important roles that protein chips will play, and give evidence for the widespread activity now under way to develop these tools. Yet, it is important to recognize the limitations of current technologies and to identify technical themes that will lead to protein chips that are broadly accessible and reliable. Below, we outline several technical issues that will be important to the maturation of protein-chip technologies.

Young-Sam Lee
Milan Mrksich*
Dept Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 5735 South Ellis Ave, Chicago, IL 60637, USA.
*e-mail: mmrksich@midway.uchicago.edu

0167-7799/02/$ – see front matter ©2002 Elsevier Science Ltd. All rights reserved. PII: S0167-7799(02)02091-7

Quantitative assays

Current protein chips are well suited to screening experiments for identifying new binding interactions and enzyme specificities. Although the ability to screen thousands of potential interactions in a single experiment represents a powerful and unprecedented advance over conventional methods, the chips still do not compete with conventional methods for characterizing individual interactions. In Snyder's study of calmodulin-binding proteins, for example, although several new interactions were discovered, many known interactions were not identified [2]. For all chip-based assays, the 'hits' must be characterized with subsequent experiments both to validate and to quantify the interaction. These limitations are not intrinsic to chip-based formats but appear to stem from a lack of uniformity in the activities of immobilized proteins [8]. The nonuniformities arise because proteins are often immobilized in a range of orientations and usually undergo partial denaturation at the surface (Fig. 2). Furthermore, the non-specific adsorption of proteins to a chip – whether it is caused by the intentional blocking of surfaces with BSA or by uncontrolled adsorption of soluble proteins in the course of an assay – results in the obstruction of immobilized proteins.

Recent work is addressing three themes that will make protein chips appropriate for quantitative assays of activity and ultimately reduce the need to check results with conventional assays [8]. The first theme is applying well-defined surface chemistries to ensure that immobilized ligands have a uniform environment and therefore activity. Many platforms used for the preparation of peptide and protein arrays, including glass and filter paper, are not structurally well defined and consequently do not allow surface properties to be engineered at the molecular scale. Ligands are therefore presented in a range of environments (especially with solvated hydrogels) and display a range of activities, making quantitative assays problematic. New opportunities for tailoring the environments of ligands are offered by self-assembled monolayers of alkanethiolates on gold and related organic surface chemistries, which are structurally well defined and synthetically modifiable surfaces [9]. An unresolved issue, however, is whether the extra binding capacity offered by three-dimensional matrices warrants the loss of the control provided by the two-dimensional surface chemistries.

The second, related, theme concerns the development of immobilization strategies that give excellent control over the points of attachment and the densities of ligands. Methods that rely on common reactions to bind the protein to the substrate – for example, reaction of lysine amino groups with aldehydes or simple physisorption of proteins (adsorption without chemical bonding) [1,5] – give poor control over the orientations and densities of immobilized proteins. Both factors can compromise the activity of an immobilized protein, because an active region of the protein can be obstructed by either the surface or a neighboring protein [8,10]. Snyder has recognized these limitations and immobilized proteins by way of a His tag to glass slides presenting a selective metal-chelating group [2]. Of course, the development of immobilization strategies must take into consideration the methods used to generate the proteins that are arrayed onto chips.

The third theme is developing and using surfaces that are inert (i.e. prevent both the nonspecific adsorption of proteins and the denaturation of immobilized proteins) as a platform for preparing protein chips. The inert surfaces will displace current methods, which treat the array with a blocking protein (such as serum albumin) before use [1] or conduct assays in the presence of detergents, both of which prevent unwanted adsorption. These methods have the unwanted consequences that the blocking protein can obstruct interactions with the immobilized proteins and that the presence of detergents can compromise protein activities [11].

Recent work suggests that attention to these issues can generate chips that perform as well as solution assays, thereby enabling the parallel, quantitative measurement of binding affinities or enzymatic activities on a protein array [12,13]. A first example immobilized peptides to a self-assembled monolayer of alkanethiolates on gold and used this substrate to characterize inhibitors of the Src kinase [12]. Peptides were immobilized by a Diels–Alder reaction, giving excellent control over the density and orientation of peptides. Furthermore, the monolayers were tailored with penta(ethylene glycol) groups, which are highly effective at preventing nonspecific adsorption of protein and therefore optimized the activities of immobilized proteins and reduced background signals in assays. This strategy provides routes to peptide chips – and, recently, to carbohydrate

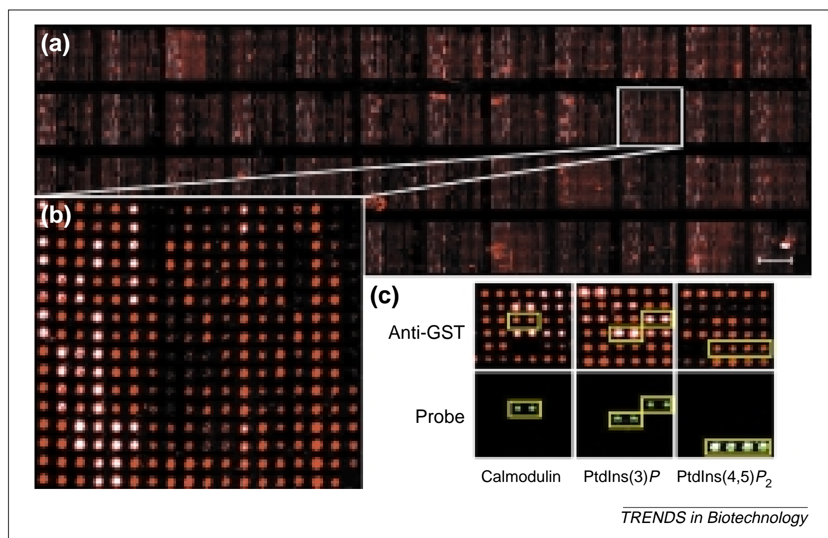

Figure 1. Protein chip used for global analysis of protein activities in yeast

(a) Several thousand glutathione-*S*-transferase (GST)-tagged proteins were spotted in duplicate on a nickel-coated microscope slide. **(b)** An enlarged image from the array shows binding of an anti-GST antibody. **(c)** Examples of assays to identify proteins that bind calmodulin and phosphatidylinositols. Abbreviations: PtdIns(3)*P*, phosphatidylinositol-3-phosphate; PtdIns(4,5)P_2, phosphatidylinositol-4,5-bisphosphate. Reprinted (abstracted/excerpted) with permission from [2] © (2001) American Association for the Advancement of Science (http://www.sciencemag.org).

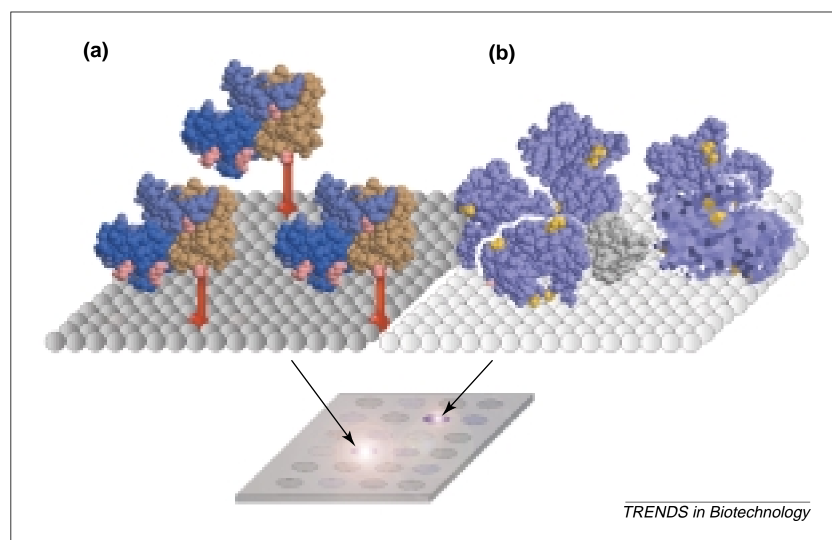

Figure 2. An illustration of the presentation of immobilized proteins in an array

(a) An idealized illustration of the presentation of immobilized proteins in an array. The proteins are all uniformly oriented, properly folded and optimally spaced to allow protein–protein interactions. (b) Current technologies present proteins in a range of orientations, with varying degrees of denaturation and with the presence of non-specifically adsorbed proteins.

chips [13] – that measure protein binding and enzymatic activities with quantitative results. We expect that an analogous surface-engineering approach will bring quantitative performance to protein chips.

Proteins and immobilization

A significant, but still underemphasized, challenge for protein chips is accessing the antibodies and proteins that are required to assemble the chip [14]. Routine applications will require the preparation of hundreds of proteins that are properly folded, carry the necessary post-translational modifications, can be obtained in pure form and can be directly immobilized with control over density and orientation, and all at an acceptable cost. The generation of antibodies is currently the most feasible, because many monoclonal antibodies are available commercially or can be generated by phage display of scFv antibody fragments. Furthermore, the structural similarity of antibodies will make it more straightforward to apply a single immobilization method for preparing antibody arrays.

The preparation of arrays with a diverse set of protein families is still difficult. The conventional approach (cloning, expressing and purifying proteins) can be used to prepare hundreds of proteins but, because it is not feasible to assay the activity of each protein, an unknown proportion of the products will not be active in the array. Furthermore, post-translational modifications including phosphorylation, proteolysis and glycosylation can substantially alter the activities of the proteins, requiring the preparation of multiple forms of each protein. This task will be especially challenging for glycosylated proteins, for which there are a multitude of glycoforms. It is clear that mammalian expression systems will be required to prepare proteins that have distinct post-translational modifications, but further work is required to develop expressions systems that have wide generality. In one clever approach, Ziauddin and Sabatini seeded mammalian cells on a surface having a patterned array of cDNAs, and found that cells were transfected by the DNA, leading to production of the encoded proteins [15]. Recent advances in high-throughput methods for the rapid cloning of PCR products to plasmids without time-consuming digestion or ligation steps will be important for producing large numbers of proteins [16,17].

There will be a strong demand for chips that contain arrays of membrane-bound proteins, which are key participants in signal-transduction events and represent the site of action for approximately half of the approved drugs [18]. Immobilization of membrane-bound proteins is still difficult, in part because the proteins require a lipid-bilayer environment to maintain their proper conformation and activity. Vogel and co-workers reported a strategy whereby G-protein-coupled receptors were immobilized within a synthetic lipid-bilayer membrane that was anchored to a self-assembled monolayer by way of streptavidin–biotin interactions [19]. Although this approach has proved to be successful for the immobilization of a single receptor type, further innovations are required to give functional microarrays containing hundreds of proteins.

The development of several new strategies will allow expressed proteins to be directly immobilized on substrates and avoid the need to purify or synthetically modify each protein. These strategies use biochemical methods to tag the protein with a moiety that selectively and irreversibly binds to a chip. In one example, an *in vitro* protein translation using a puromycin-tagged mRNA results in a covalent fusion of protein and mRNA, which can then be immobilized by way of a hybridization with the DNA of a substrate [20]. In another strategy, proteins are expressed with a peptide tag that is a substrate for biotin ligase within the cell. Hence, the biotinylated protein can be directly immobilized to substrates modified with avidin [21]. We reported a strategy that exploits the reaction of irreversible inhibitors with enzyme active sites [22]. Proteins are expressed as a cutinase fusion and the unpurified lysate is applied to a monolayer presenting a phosphonate capture ligand, resulting in covalent binding of the cutinase domain to the surface. Together, these methods will provide efficient, low-cost strategies for the site-selective immobilization of large numbers of proteins.

Detection

The utility of any microarray experiment will depend on which of a variety of options are used to analyze the chip (Table 1). Most applications have used either fluorescent tags or radiolabels to measure the binding of proteins and antibodies to arrays [1,2,6]. In an experiment that aims to identify the set of proteins that bind a protein of interest,

for example, the protein is labeled and then incubated on the array. Following a wash step, the array is scanned to reveal the presence of the tagged protein and therefore the identity of its binding partners. Experiments to identify enzymatic activities often use antibodies to bind the products of enzymes. Kinase assays, for example, rely on an anti-phosphotyrosine antibody to identify proteins that are phosphorylated by a particular kinase. Although fluorescent and radiolabel tags share excellent sensitivity and can be imaged at a micrometer resolution, they have the limitations that modification of the protein with a tag might compromise its activity and, in the case of radiolabels, that disposal is expensive. Finally, the methods are not blind: they are limited to identifying only the activities that are being probed. Hence, array-based assays using these detection methods will not identify many of the unexpected and, by extension, most exciting enzymatic activities.

Two analytical techniques now under development do not require proteins to be labeled and promise to be broadly useful for analyzing complex and undefined samples. The first, matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry, uses a laser pulse to desorb proteins from the surface followed by mass spectrometry to identify the molecular weights of the proteins [23,24]. Because this method only measures the mass of proteins at the interface, and because the desorption protocol is sufficiently mild that it does not result in fragmentation, MALDI can provide straightforward information about any enzymatic modification of a protein substrate [25]. MALDI can also be used to identify proteins that are bound to immobilized ligands of the substrate. An important technique for identifying bound proteins relies on treating the array (and the proteins that are selectively bound to the array) with proteases and then analyzing the resulting peptides to obtain sequence data.

The second technique is surface-plasmon resonance (SPR) spectroscopy and is based on an optical method that measures protein binding by way of a change in the mean refractive index near the surface [26]. In addition to the use of unlabeled proteins, this technique offers the advantage that it operates *in situ* – that is, it does not require the substrate to be rinsed and dried before analysis – and therefore provides kinetic information on binding interactions. This feature is especially important for quantifying low-affinity protein–protein interactions that would normally not be stable to the protocols for rinsing and drying before analysis. Current implementations of SPR spectroscopy are limited to observing several binding interactions simultaneously. Corn and co-workers are developing an imaging SPR system that is compatible with DNA chips, and which is expected to be applicable to protein chips [27].

Table 1. Comparison of detection methods[a]

	Quantitative analysis	Real-time analysis	Unlabeled samples	Unbiased assay	Availability
Fluorescence	Yes	No	No	No	High
Radiolabeling	Yes	No	No	No	High
MALDI–TOF	Semi	No	Yes	Yes	Medium
SPR	Yes	Yes	Yes	Partial	Limited

[a]Comparison of several different detection methods available for analyzing protein chips. Each method is qualitatively ranked for the following criteria: quantitative characterization of activity; real-time analysis of interactions; use of non-labeled proteins and complex samples; identification of unanticipated activities; availability in research laboratories. Each method has a unique pattern of strengths and limitations. Abbreviations: MALDI–TOF, matrix assisted laser desorption ionization–time-of-flight; SPR, surface plasmon resonance.

Informatics tools

The amount of data generated in a protein-chip experiment far exceeds what researchers are accustomed to analyzing. In response, there is a need for a range of informatics tools that aid the researcher in interpreting data and even go beyond the researcher's analysis to extract information from complex datasets [28]. For experiments that are designed to address a specific interaction (e.g. what are the binding partners for a soluble protein, or what are the substrates for an enzyme) the analysis of data is straightforward. The array is quantified to determine the amount of protein that binds to each array element and, with consideration of background signals, interactions are identified. Of course, protein chips that are engineered to give quantitative measurements (as described above) will always yield more robust data in these screens.

A more exciting prospect for analyzing data acquired with protein chips is the development of informatics tools that provide new or enhanced understandings of global processes underlying cellular function. These tools would provide information at a range of hierarchical levels within the cell, including identifying new components of signal-transduction pathways, connections between signaling pathways and proteomic signatures for disease, growth and differentiation. In one example, Snyder and Gerstein reported a proteome-scale analysis of protein localization in yeast [29]. This team used a high-throughput method to localize nearly 3000 tagged gene products from yeast and used these data to develop a computational algorithm for assigning the cellular locations of all yeast proteins. In another recent example, Cesareni and co-workers developed a method that combined experimental data from two-hybrid interactions and computational analysis of phage-display ligand interactions to define a protein-interaction network for peptide recognition modules [30]. These tools, and many others under development, will be invaluable to the analysis of data from protein-chip experiments.

Commercialization

Protein chips will see widespread implementation in research laboratories only when they are routinely available from commercial vendors. The first products will entail antibody arrays for the characterization of specific analyte panels. Pierce, for example, now markets an antibody array for quantifying a family of nine cytokines. Within three years, we expect that a broad selection of antibody arrays will be available from multiple vendors, along with corresponding kits that allow the user to assemble custom arrays. The arrays will harness existing instrumentation for fluorescence and radioisotope imaging to quantify results.

Chips comprising proteins for studies of selective binding interactions and enzyme activity assays will take longer to commercialize. In addition to the technical challenges outlined in this article, a substantial effort will be required to develop procedures for quality control, for improving the shelf life of chips – particularly those that present membrane-bound proteins – and for training and implementing a technical marketing and services team. We expect that the first protein chips for signal-transduction analysis and global profiling will become available in the 3–5 year timeframe, but that substantial maturation of this technology will occur during the near future. Finally, we note that the current technology companies pursuing protein chips aim to serve the drug discovery market and will not immediately affect researchers in the basic sciences.

Conclusions

Protein chips offer many exciting opportunities in both basic and applied research. A significant body of recent work has demonstrated the concept and potential of these tools, and has in turn generated a widespread interest within the biological sciences. In this article, we have sought to present a balanced view of the technical challenges that remain to developing a protein-chip technology that matches the availability and performance of today's gene chips. We recognize that many of these challenges are long-term endeavors and will not delay the introduction of the first generation of protein chips, but they will ensure continued improvement in this technology. This field will remain exciting for the multidisciplinary teams that are developing chip technologies and for the large group of users that will soon benefit from these tools.

Acknowledgements

Our work in biochip development has been supported by DARPA, NIH and NSF.

References

1 MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763
2 Zhu, H. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105
3 Zhu, H. *et al.* (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genet.* 26, 283–289
4 de Wildt, R.M. *et al.* (2000) Antibody arrays for high-throughput screening of antibody–antigen interactions. *Nat. Biotechnol.* 18, 989–994
5 Haab, B.B. *et al.* (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2, RESEARCH0004.1–RESEARCH0004.13 (http://genomebiology.com/2001/2/2/research/0004)
6 Lizardi, P. *et al.* (1998) Mutation detection and single molecule counting using isothermal rolling circle amplification. *Nat. Genet.* 19, 225–233
7 Schweitzer, B. *et al.* (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat. Biotechnol.* 20, 359–365
8 Houseman, B.T. and Mrksich, M. (2002) Towards quantitative assays with peptide chips: a surface engineering approach. *Trends Biotechnol.* 20, 279–281
9 Mrksich, M. (2000) A surface chemistry approach to studying cell adhesion. *Chem. Soc. Rev.* 29, 267–273
10 Vijayendran, R.A. and Leckband, D.E. (2001) A quantitative assessment of heterogeneity for surface-immobilized proteins. *Anal. Chem.* 73, 471–480
11 Arnebrant, T. and Wahlgren, M.C. (1995) Protein surfactant interactions at solid surfaces. *ACS Sym. Ser.* 602, 239–254
12 Houseman, B.T. *et al.* (2002) Peptide chips for the quantitative evaluation of protein kinase activity. *Nat. Biotechnol.* 20, 270–274
13 Houseman, B.T. and Mrksich, M. (2002) Carbohydrate arrays for the evaluation of protein binding and enzymatic modification. *Chem. Biol.* 9, 443–454
14 Kodadek, T. (2001) Protein microarrays: prospects and problems. *Chem. Biol.* 8, 105–115
15 Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411, 107–110
16 Heyman, J.A. *et al.* (1999) Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* 9, 383–392
17 Walhout, A.J. *et al.* (2000) GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* 328, 575–592
18 Drew, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
19 Bieri, C. *et al.* (1999) Micropatterned immobilization of a G-protein-coupled receptor and direct detection of G protein activation. *Nat. Biotechnol.* 17, 1105–1108
20 Weng, S. *et al.* (2002) Generating addressable protein microarrays with PROfusion covalent mRNA–protein fusion technology. *Proteomics* 2, 48–57
21 Chapman-Smith, A. and Cronan, J.E. (1999) The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem. Sci.* 24, 359–363
22 Hodneland, C.D. *et al.* (2002) Selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5048–5052
23 Su, J. and Mrksich, M. Using mass spectrometry to characterize self-assembled monolayers presenting peptides, proteins and carbohydrates. *Angew. Chem., Int. Ed. Engl.* (in press)
24 Gilligan, J.J. *et al.* (2002) Mass spectrometry after capture and small-volume elution of analyte from a surface plasmon resonance biosensor. *Anal. Chem.* 74, 2041–2047
25 Merchant, M. and Weinberger, S.R. (2000) Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21, 1164–1177
26 Cooper, M.A. (2002) Optical biosensors in drug discovery. *Nat. Rev. Drug Discov.* 1, 515–528
27 Nelson, B.P. *et al.* (2001) Surface plasmon resonance imaging measurements of DNA and RNA hybridization adsorption onto DNA microarrays. *Anal. Chem.* 73, 1–7
28 Gerstein, M. *et al.* (2002) Proteomics. Integrating interactomes. *Science* 295, 284–287
29 Kumar, A. *et al.* (2002) Subcellular localization of the yeast proteome. *Genes Dev.* 16, 707–719
30 Tong, A.H. *et al.* (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* 295, 321–324