DRUG DISCOVERY TODAY TARGETS

Protein biochips: the calm before the storm

Steven Bodovitz, Thomas Joos and Jutta Bachmann

The growth of protein biochip technology is on a different trajectory than other drug discovery and development technologies, such as DNA sequencing and high-throughput screening, where output per experiment has grown exponentially. By contrast, experimentation with protein biochips immediately hit barriers in output because of the limited availability of content and the challenges of running biochemical experiments on the surface of a biochip. Nevertheless, the industry has been making significant progress recently by launching new platforms with focused content and new multiplexed biochemical assays. However, this success might only represent the calm before the storm. Over the long-term, protein biochips have the potential to change the drug discovery and development process at the molecular level. The output and throughput of protein biochips could enable researchers to change from the traditional model of one target-one drug to a new model of evaluating one or more potential drugs against a panel of relevant molecular targets from a complex disease state.

The rate of advance of drug discovery and development technologies in the past decade has been exponential. If we turn the clock back ten years, GenBank (http://www.ncbi.nlm.nih.gov/) only contained 217 million base pairs of sequence (compared with 42 billion base pairs today), DNA microarrays had not been commercialized, protein identification by mass spectrometry had just become a reality [1,2], and automation was just beginning to enable high-throughput screening. Now the whole human genome has been sequenced and can be studied on a single DNA microarray, thousands of proteins can be identified in a single mass spectrometry run, and entire compound libraries can be screened in a matter of days. When protein biochips (Box 1) [3] began to emerge as the most probable candidate for the next great class of drug discovery and development technology at the beginning of this century, the common perception within the research community was that the rate of advances in protein biochip technology would match that of DNA microarrays, mass spectrometry and/or gene sequencing. This, however, has not proven to be correct. The amount of data per experiment (defined as output) that could be obtained from the earliest protein biochips [4–7] is still the standard several years later, with significant but not exponential increases in output expected within the next few years.

The primary reason for the slower-than-expected progress of the protein biochip industry has been the limited availability of content [8]. This problem is defined by a lack of high-quality capture agents that can be immobilized on the surface of a biochip – defined as a capture protein biochip – to bind to proteins from complex mixtures for the purpose of detection and quantification. In contrast to DNA microarrays, in which cDNA capture sequences can be predicted by Watson–Crick base-pairing, protein

Steven Bodovitz
BioPerspectives,
2040 Hyde Street,
San Francisco, CA 94109,
USA
e-mail:
bodovitzs@bioperspectives.com
Thomas Joos
NMI Natural and Medical Sciences Institute,
Markwiesenstr. 55,
72770 Reutlingen,
Germany
Jutta Bachmann
3 Bachmann Consulting,
Nøkkefaret 12,
1450 Nesoddtangen,
Norway

1359-6446/04/$ – see front matter PII: S1359-6446(04)03373-8

BOX 1

What is a biochip?

A biochip is a collection of miniaturized test sites – also commonly referred to as a microarray – arranged on a solid substrate that permits many tests to be performed at the same time to achieve higher throughput and speed. The miniaturized test sites can be placed onto a 2-dimensional surface, such as a glass slide, or a 3-dimensional surface, such as a bead. In the case of the 2-dimensional surface, multiple spots are arrayed, each containing a different immobilized protein or capture agent (or duplicates for control purposes). In the case of 3-dimensional surfaces, each has a unique immobilized protein or capture agent (or duplicates for control purposes), and are pooled into groups for parallel analysis.

capture agents have significantly more complex interactions with their ligands. Protein binding involves charge, hydrogen bonds and/or weak hydrophobic forces. Moreover, binding often requires specific 3-dimensional conformations, post-translational modifications and/or co-factors. Because of this complexity, each capture agent must be optimized empirically, presenting a significant challenge for large-scale production. The limited availability of content precludes experiments analogous to the comprehensive coverage of the human genome by DNA microarrays, but not high-value experiments addressing specific biological questions. Furthermore, even in the cases where more than 100 high-quality antibodies are available for capturing 100 different proteins, only approximately 30 to 40 of these can be placed together on the same biochip before the cross-reactivity between sandwich pairs becomes overwhelming [9]. Although this problem can be obviated to a certain extent by using multiple biochips, the benefits of miniaturization are diminished.

The analogy between DNA and protein biochips has fared better when applied to interaction protein biochips, where a limited number of proteins or peptides are immobilized on a surface to examine specific protein interactions. In this case, content is not a significant limitation because peptides can be synthesized and proteins can be produced using recombinant technology. The potential limitation is the question surrounding the physiological relevance of the detected interactions; in contrast to capture protein biochips, the interactions between immobilized protein or peptide and soluble protein are supposed to represent biochemical events that take place in a cell or biological fluid *in vivo*. All of these interactions take place on the surface of a biochip under a single set of conditions, which goes against the traditional dogma among biochemists that each biochemical reaction is unique and affected by a wide range of variables, including buffer conditions, protein conformation and co-factors. Moreover, in a cellular environment, a biochemical reaction will only take place if the reactants are co-localized in the same subcellular compartment; interacting biomolecules that come together on the surface of a biochip may or may not come together *in vivo*. By contrast, Michael Snyder's research group at Yale (http://.yale.edu/synder/) and several companies, such as Invitrogen Corporation (http://www.invitrogen.com), Protagen AG (http://www.protagen.com), Pepscan Systems BV (http://www.pepscan.nl) and Jerini AG (http://www.jerini.com), have shown compelling data with interaction protein biochips, in which the relevance of interactions was at least partially validated through other experimental methods [10]. The physiological relevance of the data thus remains a question that will probably only be resolved from extensive testing by researchers. However, the value of generating high-quality data from a protein biochip is readily apparent. No other technique for studying protein–protein interactions generates as many data points as quickly. If the goal is to cast the net of possible interactions far and wide, interaction protein biochips offer unparalleled speed.

Near-term prospects

The content problem initially slowed down the emergence of capture protein biochips, but developers have forged ahead by using available content and focusing on specific research areas. According to our recent analysis, at least 49 companies offer commercial capture protein biochip products and/or services, including both planar and bead-based platforms; scores of additional platforms are also available through collaborative agreements or are still under development [10]. The analytes measured include: adipokines, angiogenesis factors, apolipoproteins, cardiovascular disease markers, cell signaling molecules, chemokines, coagulation proteins, common allergens, cytokines, drugs of abuse, endocrine hormones, growth factors, HLA antigens matrix metalloproteinases, and serum cancer biomarkers. Examples for biomarker analysis are shown in Figure 1. Application areas for capture biochip assays in research include: allergies, Alzheimer's disease, apoptosis, autoimmunity, cancer, cardiovascular disease, connective tissue disorders, immunoglobulin isotyping, inflammation, infectious disease, metabolic diseases, oncology and sepsis. In addition to research and development tools, capture protein biochips are also currently used as diagnostic assays for medical conditions, such as allergies, autoimmune diseases, drug abuse, infectious diseases, heart failure and other coronary problems. In addition to diagnosing disease, capture protein biochips have the potential to identify biological warfare agents. The National Institutes of Health (http://www.nih.gov/) have allocated in excess of $1 billion to fund research on bioterrorism agents and development of vaccines and diagnostic tests.

Regarding those areas of the biochip industry dealing with interaction protein arrays, the number of companies with products and/or services on the market, according to our recent analysis, is only nine. Applications for interaction protein biochips include: characterizing antibodies and mapping epitopes; determining substrate specificities for kinases, proteases and phosphatases;

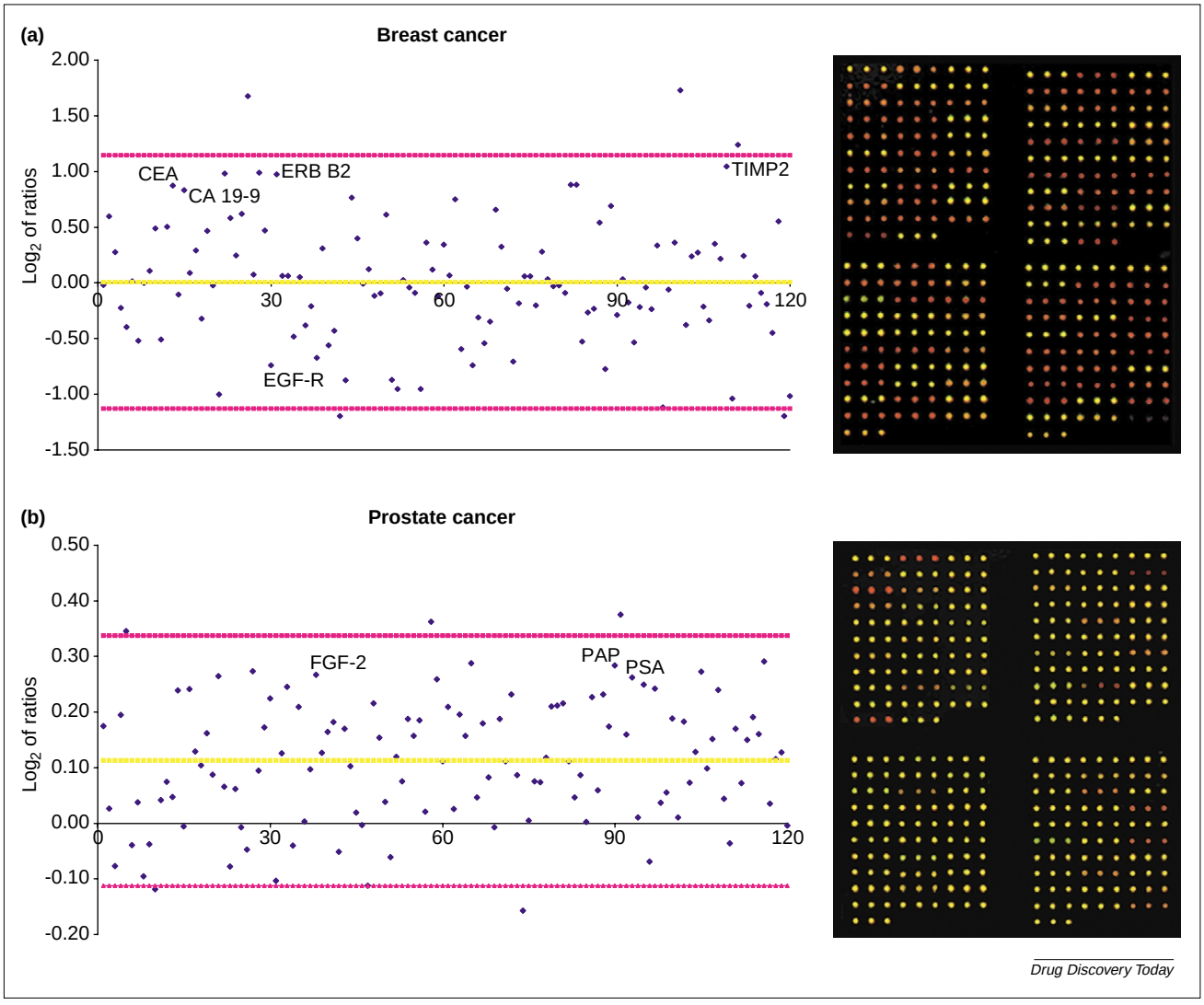

FIGURE 1

Comparative analysis of normal and cancer sera using the S&S Serum Biomarker Chip (SBC). Comparative analysis of normal and cancer sera using the S&S Serum Biomarker Chip. Scatter plots of protein abundance ratios comparing serum from breast and prostate cancer patients with serum from healthy, age- and gender-matched individuals. Serum proteins were labeled with Biotin-ULS™ and Fluorescein-ULS™, pooled and probed against the Serum Biomarker Chip. The competitive binding was detected using both streptavidin-Dy™647 and anti-fluorescein-Dy™547 conjugates with the aid of an Axon scanner. Spot-finding, background corrections, and two-channel data sampling were accomplished with ArrayVision™ FAST® software. Analysis was completed using the SBC analysis workbook, a Microsoft® Excel file that converts fluorescent data into numerical abundance values. The graphs represent the ratio of the abundance of a specific serum biomarker in Sample A relative to the same serum biomarker in Sample B (*y*-axis) for all 120 antibody specificities on the array (*x*-axis). A log2 transformation of the relative protein abundance ratios is shown. The mean of all 120 abundance ratios is shown as a yellow line and the two purple lines represent two standard deviations away from the mean abundance ratio. The pattern of elevated biomarkers identified in breast cancer differs significantly from those biomarkers elevated in the prostate cancer comparison. Data courtesy of Robert Negm, Breck Parker and Chris Zarozinski, Schleicher & Schuell BioScience (http://www.schleicher-schuell.com).

examining binding selectivity of lead compounds; and identifying pathways. Taken together, the current coverage of the proteome across all protein biochip products is approaching that of an early DNA microarray.

These technologies, although still in their early stages of development and falling well short of the early visions of proteome coverage, offer significant advantages over standard techniques and are currently in use in academia, government and industrial laboratories. The assay miniaturization increases the sensitivity of detecting low-abundance analytes [11] and reduces expenditure per data point by using less reagents and samples. The reduction of sample volume is of great importance for applications in which very small amounts of analyte are available, such as in the analysis of tumor biopsies. Multiple measurements on the same biochip (multiplexing) increase the amount of information obtained from a single experiment and enables researchers to examine patterns. This is crucial in areas such as immunology, where proteins, such as cytokines, function in concert. Multiplexing can increase the sensitivity and specificity of biomarkers by combining ones that only have limited predictive power as singlets, but high predictive power as panels [12]. Moreover, multiplexing using interaction protein biochips

BOX 2

The market model

Our market model estimates sales of protein biochips and instruments to make and use protein biochips from 70 companies, including current market leaders Biacore International AB (http://www.biacore.com), Ciphergen Biosystems, Inc. (http://www.ciphergen.com) and Luminex Corporation (http://www.luminexcorp.com). The market model, which extends from 2002 to 2008, is built on a hierarchical structure to cross-reference assumptions. At the highest level, the market is divided into three application areas: (i) drug target and biomarker identification and validation, (ii) drug development, and (iii) clinical trials and diagnostics. At the next level, the market is divided by biochip dimension: 2-dimensional planar arrays and 3-dimensional bead arrays. The next level is the type of biochip: capture and interaction. At the lowest level: hardware and biochips.

offers the possibility to identify biochemical pathways and characterize the binding selectivity of antibodies and small compounds across a representation of a cellular proteome in a single experiment. To take full advantage of its potential and to continue to grow in the near term, the protein biochip industry needs to bring robust products to the market with ever expanding content.

Long-term prospects

Once the protein biochip industry has established the reliability, utility and capabilities of its platforms, it will be ready to escape from under the shadow of the other drug discovery and development technologies that aim to study complete systems, from genome to proteome to compound library. The shadow is long in terms of numbers of data points per experiment, but short in terms of added value. Despite billions of dollars of investment in genomics, proteomics and high-throughput screening technologies over the past ten years, drug discovery and development has become less efficient [13]. The reasons for this poor return on investment are not immediately obvious. One school of thought is that it is too early to expect returns because these technologies have primarily been used to identify drug targets and potential lead compounds, both of which are early steps in a lengthy process. Although the final assessment of the return on investment will certainly require at least a few more years, it is already apparent that these technologies did not radically change the drug development process, but rather accelerated certain phases of it, such as biomarker and drug target discovery and compound screening [10]. Protein biochips, by contrast, have the potential to change the process at the molecular level.

The current process is based on the model of one target-one drug, but in reality, biological systems are complex and the majority of diseases are the result of multiple molecular changes. In this context, selecting just one target represents an arbitrary and excessive simplification. All of the molecular changes are important, and potential therapeutics should be evaluated for their abilities to reverse as many of these changes as possible. Todd Golub's group at Harvard University (http://www.broad.mit.edu/broad/toddgolub.html) has demonstrated such a multiple targets-one drug paradigm using DNA microarrays [14]. His approach, termed gene expression-based high-throughput screening, was used to identify compounds that induce the differentiation of acute myeloid leukemia cells. Dr. Golub's group screened 1,739 compounds and identified eight that reliably induced the differentiation signature and, furthermore, yielded functional evidence of *bona fide* differentiation. Caprion Pharmaceuticals (http://www.caprion.com) has demonstrated an analogous approach using mass spectrometry (C. Desjardins, personal communication). Although the details of this experiment are unfortunately proprietary, the company reports that it can distinguish between a normal and a disease state in five different animal models using sets of protein biomarkers. When the animals are treated with a drug that is known to be efficacious for the treatment of the particular disease, up to 80% of these changes in biomarkers return to normal, demonstrating the link between efficacy and multiple targets. By contrast, less efficacious drugs return only a smaller fraction of the changes to normal and often cause unrelated changes, indicating the potential for greater toxicity. These initial methodologies could be complemented or supplanted by protein biochips; in contrast to genes, proteins are directly linked to disease states, and in contrast to mass spectrometry, biochips have the throughput necessary for screening potential therapeutics. Furthermore, the ability to obtain efficacy and toxicity information from the same protein biochip assay might enable the screening process to become efficient enough to evaluate potential therapeutics in combination. Reversing a significant percentage of disease-related changes could require more than one therapeutic. The one target-one drug model might fully evolve into the multiple targets-multiple drugs model. Changing the drug discovery and development process at the molecular level might be the best medicine of all.

Summary

Even though protein biochips have fared poorly in comparisons with other high-output drug discovery and development technologies, such as DNA microarrays, the industry is now introducing a large number of protein biochip-based products for a range of applications. Based on the current level of success, we have projected that the market will grow from an estimated $122 million in 2002 to more than $500 million in 2008 (Box 2) [10]. This growth rate is based on the recent successes of the industry in providing new products, but might, in fact, be understated. Once protein biochips have established their reliability and versatility, they could be the most likely of the high-output technologies to truly revolutionize the drug discovery and development process.

References

1 Mann, M. and Wilm, M.S. (1994) Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 66, 4390–4399

2 Yates, J.R. 3rd. *et al.* (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* 214, 397–408

3 Templin, M.F. *et al.* (2004) Protein microarrays and multiplexed sandwich immunoassays: what beats the beads? *Comb. Chem. High Throughput Screen.* 7, 223–229

4 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 DOI: 10.1186/gb-2001-2-2-research0004 (http://genomebiology.com/2001/2/2/RESEARCH/0004)

5 Clark, M.D. *et al.* (1999) Construction and analysis of arrayed cDNA libraries. *Methods Enzymol.* 303, 205–233

6 MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763

7 Phizicky, E. *et al.* (2003) Protein analysis on a proteomic scale. *Nature* 422, 208–215

8 Bodovitz, S. (2003) The Protein Biochip Content Problem, *Drug Discovery World*, Fall, 50–55

9 Schweitzer, B. *et al.* (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat. Biotechnol.* 20, 359–365

10 Bodovitz, S. *et al.* (2004) Protein biochips: parallelized screening for high-output biology, *BioPerspectives and Bachmann Consulting*

11 Ekins, R.P. (1989) Multi-analyte immunoassay. *J. Pharm. Biomed. Anal.* 7, 155–168

12 Bodovitz, S. and Patterson, S. (2003) Protein biomarker strategies. *Drug Discovery World* Fall, 67-78

13 U.S. Food and Drug Administration (2004) Innovation/Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products. *FDA* (http://www.fda.gov/oc/initiatives/criticalpath/)

14 Stegmaier, K. *et al.* (2004) Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nat. Genet.* 36, 257–263