Frontiers of Biotechnology & Pharmaceuticals

生物技术及药物学前沿

Volume Two

Huang.Guo.Reiner.Zhao

科学出版社

Frontiers of Biotechnology & Pharmaceuticals

Volume Two

Huang . Guo . Reiner . Zhao

SCIENCE PRESS
Science Press New York Ltd.

Published by Science Press
16 Donghuangchenggen North Street, Beijing 100717,
P. R. China
and
Science Press New York Ltd.
84-04, 58th Ave. Elmhurst, NY 11373, U. S. A.

ISBN 7-03-009360-7/Q·1061
ISBN 1-880132-76-1

Preface to volume two of Frontiers of Biotechnology and Pharmaceuticals

As we are set to publish the second volume in a series engaged in the latest developments in the biotechnology and pharmaceutical sciences, with great fanfare, the fruits of the human genome project are on the verge of being disclosed. While the daunting task of deciphering our genetic code is a great accomplishment, perhaps an even more difficult but rewarding pathway lies ahead. From the genetic blueprint, it is estimated that upwards of 100,000 proteins are encoded. Unlocking the structures and functions of these proteins promises to give us a quantum leap in our understanding of normal cellular processes as well as the causes for many diseases which confound us today and new strategies for intervention in these diseases.

Developments in functional genomics begins part I of the book in the area of biotechnology. Aspects of how to translate the genome into the proteome in the most efficient manner is discussed. The role of nucleic acid sequences is not solely to encode genetic information. Both DNA and RNA have a variety of tertiary structures which are capable of binding ligands and catalyzing chemical transformations much like proteins. The utility of phage display as a combinatorial biology tool is described. Advancements in the automation of several important clinical diagnostics and in the area of a prostate specific antigen assay close this section.

In part II, drug development, aspects of the research that strives to improve the efficiency of the process of taking a lead compound to a drug are presented. The impact of organic synthesis on process research with an emphasis on incorporating chirality into molecules using chemical strategies such as enantioselective organometallic catalysts or chiral auxiliaries or with naturally occurring enzymes as well as an appreciation of efficiency, atom economy, and environmental concerns are discussed in the context of several recent examples. Although a practical synthesis is important to drug discovery, equally important are the physicochemical properties of a drug candidate. The efficient determination of these properties and advancements in the related fields of drug

formulation and drug delivery are discussed. This section closes with an update on the impact of analytical chemistry on the development of oligonucleotide drug candidates.

Part III, pharmaceuticals, begins with several chapters on potential new targets for therapeutic intervention. The biological role of protein tyrosine phosphatases and the effect of modulation of a particular member of this family, PTP1B, as a promising treatment for obesity and diabetes is presented. The involvement of the ubiquitin proteolytic pathway in the regulation of many cellular processes is beginning to be pieced together. This pathway presents several new targets for various types of cancers and autoimmune diseases. Likewise, biological methylations are utilized for the regulation of cellular events and offer other avenues for therapeutic intervention. The remainder of this section reviews recent approaches for treating lysosomal storage disorders, cocaine addiction, and viral infections with newly invented nucleoside analogs. This section concludes with a survey of the drugs approved by the FDA for the year 2000.

The contributors to this series have endeavored to provide the reader with topics that capture the cutting edge of pharmaceutical research. Future topics are anticipated to include new technologies to understand gene function, structure-based drug design, and a report card on the war against cancer.

Z. H., M. G., J. R., and K. Z.

I. **Biotechnology**

II. **Drug Development**

III. **Pharmaceuticals**

Functional Genomics

Yang Pan

Immunex Corporation
51 University Street
Seattle, WA 98101

1. Overview

Functional genomics appeared in the mid 1990's after the wide implementation of automated DNA sequencing made it possible to unlock the genetic information from all living organisms including humans. It refers to the development and implementation of global experimental approaches to assess the functions and activities of genes using genomics information (1). In contrast to traditional biological studies of a single gene's function, functional genomics offers information for a large number of genes in a high-throughput fashion. Transcriptional profiling is an excellent example of how functional genomics pushes the information frontier further by providing valuable expression information for the downstream biological studies. Capitalizing on the wealth of information generated from human genomics sequencing, the ultimate understanding of cellular function relies on the continuation of development and integration of high throughput functional approaches like transcriptional profiling. Examples of existing large-scale functional genomics tools and future directions will be reviewed in this article.

2. Transcriptional profiling

Transcriptional profiling (TP) utilizes DNA microarrays to study global gene regulation. The ability to comprehend from thousands of data points and to compile them into meaningful functional information greatly depends upon three elements: hardware, software and experimental design (2).

Several distinctive platform hardware technologies for microarrays are available either commercially or by self-assembly at individual research centers.

The most commonly used commercial DNA chip is the Genechip system marketed by Affymetrix. Genechip, also called oligomicroarray, uses photolithography to direct sequence-specific oligonucleotide synthesis on solid phase at very high density (3). RNAs from tissues of interest are labeled with either radioisotopes or specific reporter dyes and used for hybridization. In this format, sequencing information is required to construct the DNA microarray. As a result of *de novo* oligonucleotide synthesis, oligomicroarrays can also be used to identify genetic polymorisms by engineering oligonucleotides with single nucleotide mutations on the chip.

Another commonly used DNA microarray system is represented by the GEM system from Incyte Genomics (4). cDNA clones either from cDNA libraries or other sources are amplified by PCR and arrayed on solid supports made either of glass or nylon before hybridization. In this format the identities of the clones are not required for the generation of array data.

There are variations of array systems available commercially to complement the above two systems and most importantly to bypass some of the earlier patent restrictions. Among them, the most notable is the Flexjet DNA microassay system based on a modification of the ink-jet printing process by Rosetta (5). Flexjet microassay can synthesize oligonucleotides on the glass slides using traditional phosphoramidite chemistry. It offers the advantage of flexible array design and speedy chip production.

Although many efforts from both academia and the biotech industry have been devoted to the development of computational tools for effective analysis of the expression profiling data, there is still plenty of room for improvement. Because multiple platforms for TP are being used for data collection, and the inherent variability existing within each technology, one pressing technical challenge is to correlate data collected from different platforms and different experiments for direct comparison, a prerequisite to establish global gene expression reference databases.

Despite all the complex issues discussed above that affect TP, significant progress has been made by biologists using existing tools in areas that informative experimental designs can be applied. One such area is the yeast global gene expressions and pathway analysis. This was achieved due to the yeast's small genome size and our vast accumulated knowledge using yeast as a model organism (5). A second area is in human cancer classifications. By using relatively well-defined tumor cells, scientists have been able to unearth specific gene expression profiles linked to specific types of cancers (6). With more researchers focusing their attention on TP, we can expect more and better data to be generated, which will lead to the assembly of global gene expression databases in the near future.

3. Proteomics

In comparison to gene expression analysis at the mRNA level, proteome analysis provides more information directly affecting biological systems by analyzing the proteins. While 2-dimensional polyacrylamide gel electrophoresis coupled with mass spectrometry remains the main method to separate and analyze complex protein mixtures, it has the inherent disadvantages of being non-quantitative, difficult to automate and biased towards highly abundant proteins.

Recently a new type of quantitative analysis method using isotope-coded affinity tags (ICAT) was developed (7). This method works by tagging the cysteine residues in a given protein with a stable radioactive isotope containing biotin label. Enzymatically cleaved peptide fragments containing the biotin tags are then isolated and analyzed for sequence composition by mass spectrometry. The selective retrieval of cysteine containing peptides reduces the complexity of the protein mixtures and subsequently results in more accurate data analysis. One major application for this method is its ability to accurately quantify relative protein abundance from two different experimental protein mixtures (for example stimulated vs. unstimulated cells). This is achieved by measuring the ratio of peptides derived from the two cell populations, after labeling them separately with either normal or deuterated tags.

Meanwhile functional protein microarrays are on the horizon as the next wave of functional genomics tool (8). They will potentially provide information on protein post-translational modifications, alternatively spliced protein variants and protein-protein interactions. They may even one day allow us to screen for small molecule interactions, thereby bridging genomics and drug screening. However, unlike DNA microarrays, protein arrays need to overcome a slew of hurdles before they can be adopted for practical applications. For example, proteins are not easy to attach to solid surfaces and are prone to denaturation. Proteins cannot be amplified like DNA, which makes it difficult to generate them in large quantity and variety for chip production.

Currently, much work is devoted to overcoming the protein-coating problem. For example, a method developed by MacBeath and Schreiber last year may lead to low cost protein chip production (8). The miniaturized assay they developed can immobilize more than 10,000 proteins by covalently attaching them to chemically derivatized glass microscope slides at high density. These spotted proteins were able to retain their activities to interact with specific proteins and small molecules in solution.

The concept of protein chips has already begun to be applied to practical biological problems. For example, prototypes of phage antibody arrays have been reportedly used in high-throughput screening of antibody-antigen interactions (10, 11). The advantage of using protein arrays over conventional phage selection is the ability to reduce bias for immunodominant epitopes and abundant proteins, therefore increasing the antibody diversity and chances to discover highly specific antibody-antigen interactions.

Multi-tasking chips that enable the capture, separation and quantitative analysis of proteins from different cell populations are also in early stage development. The principle behind this is the combination of protein chip technology and mass spectrometry. Proteins from different cell populations can be affinity-captured on separate chips with a pre-treated surface under optimized conditions and then the captured proteins can be de-absorbed and ionized for analysis by laser excitation and identified by their distinctive molecular weights. In principle, this method would bypass the constraint of

keeping active proteins on a chip, since the molecular weight is independent of activity.

4. Structural Genomics

With the availability of genomics sequences from human and other species, structural genomics promises to provide a major impact on the study of biological function (12). Comprehensive protein structure databases based on genomics sequences are now feasible because of the recent advances in high throughput robotics sample handling, X-ray crystallography and NMR spectroscopy. The time needed for cloning a cDNA to solving the encoded protein structure has been greatly reduced, while the range of protein structures that have been solved has become more diversified and sophisticated, as in the case of membrane proteins. To capitalize on these advances, multiple structural genomic initiatives from both academia and commercial sources, reminiscent of those of the human genomics sequencing projects, have sprung up across the United States, Europe and Asia (13,14,15). For example, in 2000, the National Institute of General Medical Sciences started the funding for the "Protein Structure Initiative" which was designed to organize cooperative efforts from academia and aimed to solve 10,000 structures from all protein fold families in 10 years. Meanwhile, there are at least five biotech companies that have invested in commercial structure genomics by providing information on various drug targets (16). One can imagine that it is only a matter of time before a significant number of protein structures become available.

There is little doubt that protein structure databases will aid drug target selection, validation, and rational drug design. However, the benefit of structural genomics is not limited to the generation of 3-dimensional macromolecular structures. It is well known that proteins functions are largely dictated by their intrinsic structures. Like with transcriptional profiling and proteomics, protein structural information could be used to further define gene functions within complex biological pathways. For example, in some cases, a resemblance of the structure from an otherwise novel sequence to a known protein will provide an underlining evolutionary link, which in turn could result in similar functional predictions. However, to fully carry out the task, drastic

improvements are needed in several areas such as protein production (both in quantity and quality), the speed of structure determination, and integrated bioinformatic tools for data analysis and annotation.

5. Animal models

As a mammalian model system, the mouse has played and will continue to play an important role in our understanding of gene function. With concerted efforts from the NIH and the industrial powerhouse Celera, the complete sequencing of the mouse genome is imminent. Genetic deletion or insertion of targeted genes in mouse embryos by homologous recombination has traditionally been used to elucidate the functions for homologous human genes. While traditional targeted mutagenesis provides powerful insights into gene function, it is both time-consuming and laborious. It is not surprising that to date less than 5% of genes have been targeted in the mouse (17). In recent years, technologies to systematically generate knockout mouse clone databases have been established by genome-wide gene trapping in embryonic stem cells. For example the commercial database Ommbank provided by Lexicon was generated by random insertional mutagenesis via Moloney murine leukemia virus (18). The interrupted genes from the insertional mutagenesis could be identified through inserted tag sequences.

Alternatively the chemical mutagen N-ethyl-N-nitrosourea (ENU) has also been applied to introduce phenotype-based mutations systematically in the mouse (17). Large scale ENU screens have been successfully carried out in other model organisms such as Drosophila melanogaster, C. elegans and recently in zebrafish. ENU primarily introduces point mutations, therefore creating a variety of alleles, which is extremely useful in dissecting gene function. However, to identify global ENU mutations is not straightforward because multiple genes can be mutated at the same time and there is no specific tag sequence in the genome to serve as an anchor. A modified method, regional-specific mutagenesis, which uses chromosomal deletions as markers to screen mutations within a genetic region, has been developed to aid mutation identification in conjunction with ENU (19).

While the mouse models provide powerful *in vivo* biological evidence for the functions of the targeted genes, they do have limitations. Targeted mutagenesis and ENU are excellent at identifying genes critical to embryonic developmental processes, but not for those genes whose functions are either not critical in early development or can be compensated for by other functional homologues. For genes such as Endothelin–1, which is critical to blood pressure regulation, the function of this gene can not be studied in adult homozygous knockout animals due to the embryonic lethality as a result of pharyngeal arch malformation (20). To bypass this problem, conditional knockouts using the Cre/loxP system has been created to selectively turn off a gene of interest in adult mice (21). Gene function can also be studied by over-expressing them in targeted tissues using transgenic technology or viral-aided gene delivery.

6. Summary

With the recent advances in human genomics sequencing, we are entering into an exciting era where we can truly begin to decipher the complex biological functional codes on a grand scale. However, this endeavor requires fundamental rethinking of biological research. In addition to applying the traditional hypothesis-driven research, we may need to rely more on the approach of reverse genetics. Transcriptional profiling, proteomics, structural genomics and animal models are among the most promising functional genomics approaches to the generation of large amounts and specific information on a great number of genes. Aside from the large-scale functional genomics tools mentioned above, there are a wide range of highly specific cell-based assays designed to capture perturbations on the cellular level such as calcium influx, cell migration and proliferation. Optical devices have been and will play a significant role in transforming *in vitro* cell based assays into high throughput fare rivaled by small molecule screening. It will not be long before one can start to put together a comprehensive functional catalog for most of the human genes.

References

1. Boguski, M. *Science* **1999**, 286:453.
2. Young, R. A. *Cell* **2000**, 102:9.
3. Fodor, S. *Science* **1997**, 277:393.
4. Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, 270:467.
5. Hugh, T. R. *et al. Cell* **2000**, 1-2:109.
6. (a) Golob, T. R. *et al. Science* **1999**, 286:531. (b) Gygi. S. P. *et al. Nature Biotechnology* **1999**, 17:994.
7. Van Brunt, J. edited. Protein chip challenges, Signal magazine. Nov. **2000**
8. MacBeath, G.; Schreiber, S. T. *Science* **2000**, 289:1760.
9. De Mildt, R. M. T.; Mundy, C. R.; Gorrick, B. D.; Tomlinson, I. M. *Nature Biotechnology* **2000**, 18:989.
10. Holt, L. J.; Bussow, K.; Walter, G.; Tomlinson, I. M. *Nucleic Acids Research* **2000**, 28:72.
11. Russell, R. B.; Eggleston, D. S. *Nature Structural Biology* **2000**, 7, supp, 928.
12. Terwilliger, T. C. *Nature Structural Biology* **2000**, 7, supp, 935.
13. Heinemann, U. *Nature Structural Biology* **2000**, 7, supp, 940.
14. Yokoyama, S. *et al. Nature Structural Biology* **2000**, 7, supp, 943.
15. Dry, S.; McCarthy, S.; Harris, T. *Nature Structural Biology* **2000**, 7, supp, 946.
16. Anderson, K. V. *Trends in genetics* **2000**, 16:99.
17. Zambrowicz, B. P. *et al. Nature* **1998**, 392:608.
18. Schimenti, J.; Bucan, M. *Genome research* **1998**, 8:698.
19. Kurihara, Y. *et al. Nature* **1994**, 368:703.
20. Abremski, K.; Hoess, R. *J. Biol. Chem.* **1984**, 259:1509.

In vitro Selection of Ligand-binding and Catalytic RNA and DNA

Zhen Huang[1]* and Jun Zhao[2]

[1]*Program of Biochemistry and Chemistry, Department of Chemistry Brooklyn College, The City University of New York, New York 11210*

[2]*Department of Urology, Union Hospital, Tonji Medical School, Huazhong University of Science and Technology, Wuhan 430022, P.R. China*

* *E-mail: zhuang@brooklyn.cuny.edu*

Abstract: Nucleic acids are capable of forming functional structures through specifically arranging sequences and functional groups, which generate molecular recognition motifs and/or catalytic centers. An *in vitro* selection technique was initially developed to isolate functional RNAs, including ligand-binding RNA (aptamer) and catalytic RNA (ribozyme), to explore sequence, structure and function correlations of RNA molecules, and to study small molecule recognition and macromolecule interactions, such as between proteins and RNA. Analogously, this technique was later extended to isolate and investigate functional entities of the RNA counterpart, DNA. Using an *in vitro* selection technique, many protein aptamers have been isolated that selectively block signal transduction and inhibit viral activity, and ribozymes have also been optimized *in vitro* for ribozyme gene therapy applications. The therapeutic potentials of both functional RNA and DNA have spurred research of nucleic acid aptamers and catalysts in the clinical setting. This review covers recent progress in the emerging area of ligand-binding and catalytic nucleic acids.

1. Introduction

Macromolecules with multifunctional building units and sufficient size can form stable structures in the sea of solvent molecules by distributing their functional groups and intramolecular interactions in three-dimensional space.[1,2] Proper arrangement of the units and functional groups can generate arrays of hydrogen-bond accepting and donating functions, hydrophobic interactions,