From Knowing to Controlling: A Path from Genomics to Drugs Using Small Molecule Probes

Robert L. Strausberg and Stuart L. Schreiber

Science **300**, 294 (2003);
DOI: 10.1126/science.1083395

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

VIEWPOINT

From Knowing to Controlling: A Path from Genomics to Drugs Using Small Molecule Probes

Robert L. Strausberg[1]* and Stuart L. Schreiber[2]

The National Cancer Institute Initiative in Chemical Genetics is designed to encourage the development of small molecular probes. The probes are useful for activating or inactivating protein functions, thereby providing resources that help discern the functions of gene products in normal and disease cells, as well as in tissues. This initiative includes "ChemBank," a suite of informatics tools and databases aimed at promoting the development and use of chemical genetics by scientists worldwide. The information generated with such tools should provide a critical link from genomic discovery to drug development.

Genomics holds great promise for transforming medicine, but the details of how the promise will become reality are still vague. The National Cancer Institute (NCI), like many of the NIH institutes, has several programs that focus on collecting or exploiting genomic information: The goal of the Cancer Genome Anatomy Project (http://cgap.nci.nih.gov/) (*1*), for example, is to determine gene expression profiles for normal, precancer, and cancer cells, and to share this information with the entire scientific community. We are confident that, over time, collecting and sharing these data will lead to better detection, diagnosis, and treatment for cancer patients. However, there is a large gap between knowing which genes are overexpressed in cancer and being able to modulate the functions of those genes with temporal and spatial control in tissue culture cells or animals, a necessary first step before drug development can begin in earnest. Genetic tools for eliminating function, such as knockout mice, are helpful (*2*) but frustratingly slow, whereas newer tools, such as RNA interference (RNAi), hold considerable promise but require additional development (*3*). Because cancer results after many changes occur in the cell, we would ideally like to be able to test the effects of perturbing many gene functions in different ways and in different combinations. This is simply not possible with the tools currently available to science. It is in this context that the NCI recently established a new Initiative in Chemical Genetics (ICG) to encourage the development of small-molecule probes that can be used to study the results of activating or inactivating protein functions in cells and organisms.

Chemical Genetics

NCI's interest in chemical genetics began in 1997, when a program in "Cancer Drug Discovery: Diversity Generation and Smart Assays" was announced that eventually funded six biology-chemistry centers (*4*). All of the centers shared the idea that modern synthetic organic chemistry, using techniques such as split-pool synthesis and combinatorial chemistry, could be exploited to increase the availability of small molecules as probes of biological systems, and all explicitly took on the challenge of increasing interactions between biologists and chemists. An early success was the discovery of monastrol, a small molecule inhibitor of the kinesin motor protein Eg5; monastrol is the first small molecule known to target a member of the kinesin family (*5*). Eg5 inhibitors are now entering Phase I clinical trials and have shown promising results in animal models of cancer. Another product that came from the biology-chemistry centers program is diversity-oriented synthesis (DOS), in which small molecules are directed to large expanses of chemical descriptor space by means of a planning algorithm for efficient syntheses of stereochemically complex and skeletally diverse small molecules (*6*).

The success of these programs reinforced the belief that giving academics access to the tools of high-throughput screening, coupled with access to collections of diverse small molecules, could be a powerful way to accelerate the discovery of potential drug targets. To expand access to this approach, the NCI created the Initiative in Chemical Genetics (ICG) and, after a national competition, Harvard's Institute for Chemistry and Cell Biology (ICCB) was chosen as the first recipient of ICG funding. The ICCB is committed to increasing general access to the

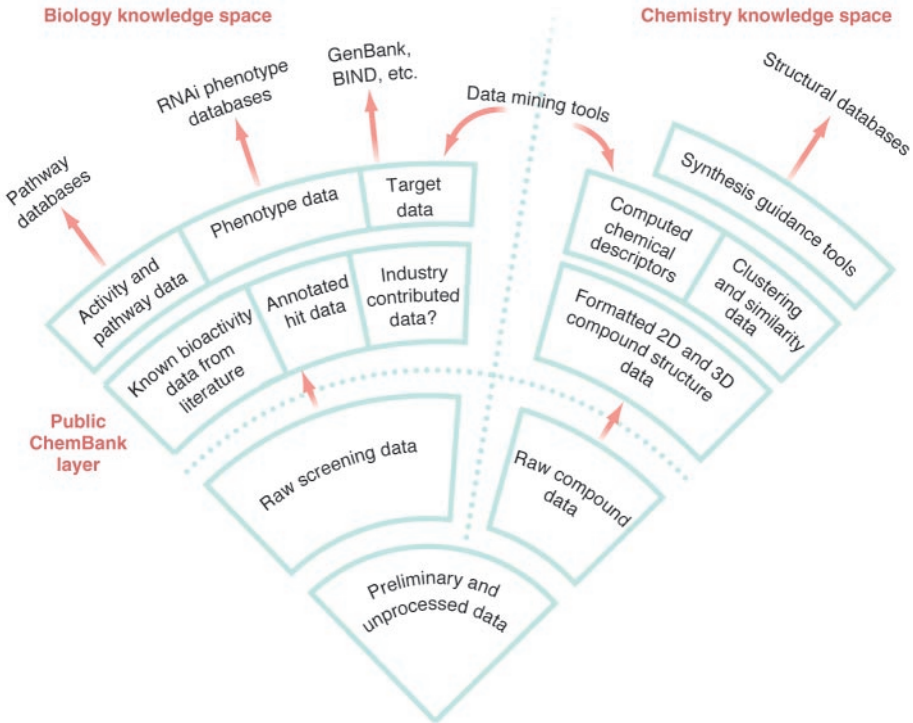

Fig. 1. The functional layers of ChemBank. The flow progresses from raw unprocessed screening and compound data at the bottom, to publicly available data and tools at the top. The upper layers of ChemBank data will be linked to outside databases such as GenBank and Protein Data Bank.

[1]Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Bethesda, MD 20892, USA. [2]Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: RLS@nih.gov

methods of chemical genetics through several mechanisms: (i) The ICCB shares expertise in high-throughput screening by offering information on how to set up a screening facility (*7*) and by providing consultation on technical issues; (ii) the DOS platform and other techniques relevant to chemical genetics developed at ICCB are now available (*8*, *9*); and (iii) the NCI and ICCB by means of the ICG are developing "ChemBank," a suite of informatic tools and federated databases that aim to promote the development and use of chemical genetics (Fig. 1).

ChemBank

One of the missions of ChemBank is to adopt common standards and language that will allow the management and sharing of chemical genetic data. The growing community of academic and industrial scientists interested in data exchange will be encouraged to submit information, using defined standards, on the structures and activities of small molecules. ChemBank is being developed to facilitate the identification of the proteins to which small molecules discovered in cellular and organismal assays bind, and to explore the underlying principles of biological networks. The ICCB will populate ChemBank's databases with baseline information generated by the ICG-funded systematic screening group, with an additional aim to incorporate the large body of data being produced outside of the ICCB. In addition, analysis tools are being developed to relate the selection of reagents, appendages, and pathways in DOS to the swaths of chemical or biological spaces of interest.

Although ChemBank is still in its infancy, selected tools and databases are becoming available; as an example, an interactive database of information on over 2000 known bioactive compounds is now available (*10*). We invite and encourage readers that have access to data on small molecules to participate in this open-access forum by contributing information to the bioactive database of ChemBank. We hope that ChemBank will be a planning and discovery tool for chemists and biologists worldwide, providing unfettered access to the data and tools of chemical genetics.

A Hopeful View of the Future

The chemical description of DNA by Watson and Crick forged an everlasting connection between chemistry and genetics. Now, in this 50th year after their landmark publication, we expect that the ICG will catalyze a new era, in which advances in chemistry are applied toward understanding the functions of DNA and its encoded products. The completion of the human genome sequence will provide definition of the precise chemical structure of the human chromosomes. Through the precise chemical perturbation of biological processes, we hope to advance our knowledge of the functions of the human genome and to intervene in disease processes as we strive toward the betterment of human health.

During the past century there have been many examples of researchers identifying and using small molecules to probe aspects of biology, generally on an ad hoc basis, and it is a long-range goal of the ICG to transform their use into a general and systematic approach to understand biology and medicine (*11*). Furthermore, in light of the new research facilities for small-molecule screening that are increasingly dotting the landscape, the ICG aims to provide tools and information to the scientific community so that the chemical genetic approach can be integrated into the fabric of day-to-day life science research. We hope that, in the future, biologists will have routine access to small-molecule probes to modulate the individual functions of their favorite protein or to dissect networks by the instantaneous modulation of combinations of proteins (*12*). Cheminformatics will no longer be largely the province of the pharmaceutical industry, and chemists will exploit their skills in organic synthesis to populate currently virgin swaths of chemical descriptor space or, even better, the swaths of chemical space that are empirically identified as optimal for overlap with a medically relevant area of biology descriptor space. A key premise of the ICG is that the use of chemistry and information science as up-front discovery tools can diminish the gap between biology and medicine in the future.

References and Notes

1. G. J. Riggins, R. L. Strausberg, *Hum. Mol. Genet.* **10**, 663 (2001).
2. T. Van Dyke, T. Jacks, *Cell* **108**, 135 (2002).
3. D. J. Shuey, D. E. McCallus, T. Giordano, *Drug Discov. Today* **7**, 1040 (2002).
4. See http://dtp.nci.nih.gov/branches/gcob/gcob_web18.html.
5. T. U. Mayer *et al*., *Science* **286**, 971 (1999).
6. S. L. Schreiber, *Science* **287**, 1964 (2000).
7. Instructions for "Building an HTS Screening Facility" are available at http://iccb.med.harvard.edu/screening.
8. H. E. Blackwell *et al*., *Chem Biol.* **8**, 1167 (2001).
9. P. A. Clemons *et al*., *Chem. Biol.* **8**, 1183 (2001).
10. See http://iccb.med.harvard.edu/chembank.
11. S. L. Schreiber, *Chem. Eng. News* **81**, 51 (2003).
12. A. F. Straight *et al*., *Science* **299**, 1743 (2003).

VIEWPOINT

Developing a Platform for Genomic Medicine in Mexico

Gerardo Jimenez-Sanchez

Mexico is preparing to develop a genomic medicine program focused on national health problems. Modern Mexicans result from an admixture of more than 65 native Indian groups with Spaniards, leading to a unique genetic makeup and a characteristic set of disease susceptibilities. Since 1999, more than 100 experts from different fields have joined efforts with government, academia, and industry to identify priorities and goals for genomic medicine in Mexico. The plan includes establishment of an Institute of Genomic Medicine with strong intramural and extramural programs. This project is expected to ease the social and financial burden of health problems in Mexico.

In Mexico, resources are limited and issues such as access to maternal and child care, provision of clean water, and proper nutrition and education continue to be a high priority. Against this backdrop of immediate needs, a large investment in developing genomic medicine might seem unreasonable. However, chronic, infectious, and degenerative diseases are major causes of mortality in Mexico today (*1*). These health problems represent a serious financial burden. Direct costs of diabetes alone account for 4 to 6% of the total annual health budget (*2*). New strategies for prevention, early diagnosis, and more effective treatment are essential to meet the mid- and long-term health care costs in Mexico. Although economic limitations often cause developing countries to postpone the implementation of novel technologies, taking advantage of the current window of opportunity to develop genomic medicine will contribute to economic growth and social welfare. The modern population of Mexico has a characteristic genetic structure, as shown by polymorphisms in blood group systems, serum proteins, major histocompatibility complex genes, and microsatellites (*3, 4*). These observations suggest that genomic medicine in

Consortium for the Institute of Genomic Medicine of Mexico, Camino al Ajusco 130-101, Col. Jardines en la Montaña, Mexico DF 14210, Mexico, and McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA. E-mail: gjimenez@inmegen.org.mx