

Review

Chemical Genomics and Medicinal Systems Biology: Chemical Control of Genomic Networks in Human Systems Biology for Innovative Medicine

Tae Kook Kim*

Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
Institute of Chemistry and Cell Biology-Initiative for Chemical Genomics, Department of Biological Chemistry and
Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Received 11 December 2003

With advances in determining the entire DNA sequence of the human genome, it is now critical to systematically identify the function of a number of genes in the human genome. These biological challenges, especially those in human diseases, should be addressed in human cells in which conventional (e.g. genetic) approaches have been extremely difficult to implement. To overcome this, several approaches have been initiated. This review will focus on the development of a novel ‘chemical genetic/genomic approach’ that uses small molecules to ‘probe and identify’ the function of genes in specific biological processes or pathways in human cells. Due to the close relationship of small molecules with drugs, these systematic and integrative studies will lead to the ‘medicinal systems biology approach’ which is critical to ‘formulate and modulate’ complex biological (disease) networks by small molecules (drugs) in human bio-systems.

Keywords: Chemical genomics, Medicine, Networks, Systems biology

Introduction

“What is the best way to study biological processes and genomic networks in human cells?” Biochemistry and genetics are both powerful approaches, but ideally we would like to study protein function in human cells under physiologically relevant contexts (i.e. their own cellular and molecular environments). Many complex biological questions about development and disease are best addressed in human cells. Scientists have become acutely aware of the problems of conventional techniques in working on human cells. One of

attempts to solve these problems includes the development of a novel chemical genetic/genomic approach that substitutes small molecules for genetic/genomic mutations, and uses these chemical tools to probe the function of specific pathways and networks in human cells. One such case was the small molecule (FK506)-based studies of human immune response. The discovery that a small molecule, named FK506, which blocks the production of several cytokines that are produced in an immune response, was one of the first successful applications of this chemical genetic/genomic approach (Schreiber, 1998). Subsequent studies found that FK506 targets protein calcineurin, indicating that calcineurin is important for inducing the immune response. This study revealed to scientists a previously unknown biological function of calcineurin, and it identified a small molecule, FK506, with possible medical use as an immunosuppressant drug. Harvard Institute of Chemistry and Cell Biology-Initiative for Chemical Genomics (ICCB-ICG) was established to explore this new science of chemical genetics/genomics. These novel ideas and unique approaches to solve complex human biological problems have been substantially supported and validated by pioneering chemical genetic/genomic studies during the last 4-5 years. Now a chemical genomics program is extensively supported by the National Cancer Institute as a central tool for the discovery of drugs and therapeutic disease genes. Reflecting these achievements, I will briefly describe my own perspectives on the key ideas and technologies of chemical genomics with an emphasis on those that evolved at Harvard and its connections with medicinal systems biology for innovative medicine. Furthermore, I will highlight the potential values and impacts of chemical genomics and medicinal systems biology in human biology and drug discovery.

*To whom correspondence should be addressed.

Tel: 82-42-869-2634; Fax: 82-42-869-8160

E-mail: tkkim@kaist.ac.kr

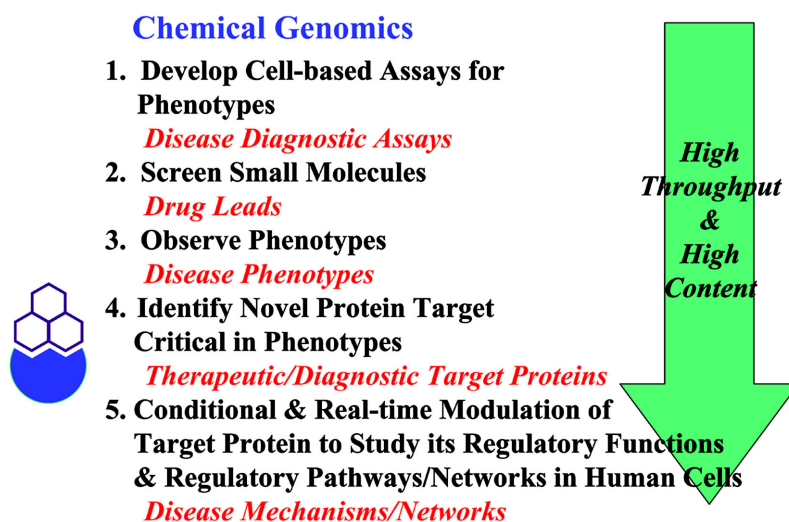


Fig. 1. Processes in chemical genomics.

Key Concepts of Chemical Genomics and Medicinal Systems Biology

The small molecule-based genomic approach (Fig. 1) involves the following: (1) screening and identifying small molecules that affect specific cellular pathways and networks with unbiased phenotype-based screens; (2) characterizing their action mechanisms and identifying the proteins whose activities are affected by these chemical ligands; and (3) real-time studying the biological consequences of conditional inhibition or activation of the target protein using the small molecules as a tool (Schreiber, 1998; King, 1999; Kim, 2000). These procedures yield information that is similar to what could be learned from cloning a mutant gene from a genetic/genomic screen and using a conditional allele to study the protein function that it encodes in the classical genetic/genomic approach with model organisms. This small molecule-based approach has been extensively complemented by functional genomic approaches, which will provide the relevant and creative directions for efficient discovery of bioactive small molecules (drug leads) and genes (therapeutic/diagnostic genes) in specific biological (disease) processes and networks (Strausberg and Schreiber, 2003).

These novel ideas and unique approaches to solve complex human biological problems have been substantially supported and validated by pioneering chemical genomic studies at Harvard ICCB-ICG during the last 4-5 years. We have done the following: (1) performed a number of proof-of-concept experiments to validate a novel chemical genomic approach; (2) initiated to develop core chemical genomic technologies; and (3) initiated the application of these unique approaches to the identification of small molecules and their target proteins that affect a chosen biological (e.g. tumorigenic) pathway in human cells.

With the validation of scientific platforms for chemical genomics, it is now necessary to focus on the development of

highly integrative technological platforms for systematic chemical genomics in human disease networks. To this end, we would like to accomplish the following: (1) systematically establish these small molecule-based approaches in a high-throughput and high-content format to increase the efficiency of discovery/screening processes; (2) develop novel key technological and scientific platforms (e.g. higher order assays, screenings, and molecular target identifications) for the full realization of chemical genomics in human bio-systems; and (3) systemize a variety of screening data and chemical modulators from human cells in multi-dimensional database and systems informatics. These efforts in chemical genomics will lead to the “medicinal systems biology” which is critical to identify and catalog therapeutic genes (i.e. targets of therapeutic small molecules) for specific disease pathways and networks, and to formulate disease networks and database in human bio-systems using systems biology. Ultimately, the goal of the chemical genomics and medicinal systems biology is to control complex biological (especially disease) networks through the development of novel drug molecules and therapeutic proteins (Fig. 2).

Key Technological Platforms for Chemical Genomics and Medicinal Systems Biology

Higher-order and multiplex assays Development of unbiased cell-based (or organism-based) assays is important for the screening of chemical libraries toward human genomic networks (Tian *et al.*, 1998; Komarov *et al.*, 1999; Mayer *et al.*, 1999). Cell-based assays for screening small molecules have been hampered by high false-positive rates, non-specific effects, and inefficient assays. For example, specific inhibition of reporter gene activity can be confused with cytotoxic effects of the compounds. Significant effects of specific compounds can also be masked by the screening noise in each

Chemical Genomics & Medicinal Systems Biology

Small Molecule-based
**Systematic Identification of
 Human Genome Functions**
and
**Formulation & Modulation of
 Genomic Disease Networks in Human Systems Biology**
for
**“High Throughput Discovery of
 Therapeutic/Diagnostic Genes, Drug Leads & Disease Mechanisms”**
with
Multi-dimensional Systems Database
*Chemical Structures-Genome, Proteome Activities-
 Target Proteins, Target Pathways/Networks-Phenotypic, Functional Alterations*

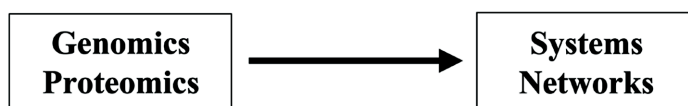


Fig. 2. Overview of chemical genomics and medicinal systems biology.

well of high-density assay plates. To overcome these limitations, multiplex readout assay systems have been developed for specific biological pathways and networks. Multiple and simultaneous readouts in single assay wells of high-throughput screening formats enable more convenient and normalized processing of many compound effects. They also provide reliable data for specific compounds in the chemical libraries, facilitating identification of specific modulators in biological pathways and networks.

High-throughput and high-content screenings Chemical genomics focuses on the development of cell-based (rather than biochemical) assays and screenings for biological pathways and networks. To increase high-throughput in the screening, one of developments includes miniaturization (Walling *et al.*, 2001). Cell-based screening can now be performed in 384 well plates at ~20 μ l/assay and in 1536 well plates at ~2-4 μ l/assay. A developed screening facility provides equipment and expertise for chemical genomics. These include the following: (1) compound storage; (2) assay systems; (3) compound transfer; (4) liquid handling; (5) plate washing; (6) plate imaging system; and (7) automated microscopic screening systems. Molecular cytology, for example immunofluorescence, is arguably the most important single method in imaging-based research (Rosania *et al.*, 2000). To screen for chemicals using this information-rich (high-content) approach, we built an automated microscopic imaging screening system at Harvard ICCB-ICG (Yarrow *et al.*, 2003). This is an inverted microscope with an automated X-Y stage and a piezo Z-focus that sequentially visits each

well of a 384 well plate, focuses, and collects images in multiple fluorescence channels and transmits light to a cooled CCD camera. It takes 15-20 minutes per plate to collect multiple images of each well, depending on the brightness of the image. The software and sophisticated algorithms have been developed for automatically analyzing the images. The plate crane and bio-chamber can be implemented to allow automated, real-time kinetic operations.

Small molecule libraries The available libraries for chemical genomics include those made and/or available at institutes and pharmaceutical companies and those made at Harvard ICCB-ICG by split-pool synthesis on beads (Schreiber, 2000; Clemons *et al.*, 2001). Limited sources of natural products can also be utilized with careful consideration of supplied amounts. Acquired compound libraries are currently our most frequently used resource for chemical genomics at Harvard. Libraries from companies are rich in heterocyclic aromatic compounds, a class of synthetic molecules with a good history of providing drug leads for enzyme inhibitors and agonists/antagonists of receptors. The library from the National Cancer Institute includes molecules with relatively smaller average molecular weight and complexity; one advantage may be that compounds have been developed toward anti-cancer or other therapeutic applications. One of considerations is that these various sources would be complementary for the studies of chemical genomics and medicinal systems biology. Other small molecule resources are the unique libraries made at Harvard ICCB-ICG. These compounds may be larger and more

chemically interesting than those in our acquired libraries: (1) complexity--chemical genomics aims to imitate the structural complexity of natural products that target diverse proteins with high affinity and selectivity; and (2) diversity--chemical genomics aims to achieve as much diversity as possible from a given set of reagents and reactions for highly divergent structures. Thus, these libraries will include complex and diverse structural entities in their overall size, various numbers and sizes of fused ring systems, variety of building blocks, and stereochemical diversities (Schreiber, 2000).

Molecular medicinal-systems informatics One of goals in chemical genomics is to develop the database that integrates chemical structure and screening results with a variety of information on cellular phenotypes, cellular functions, biological pathways/networks, and protein targets. This system will be developed by implementing multiple application-specific databases that are specialized to handle the separate problems of chemical structures and multi-dimensional biological activities. We initially captured and stored all of the synthetic chemical library data as well as the biological screening data by implementing a combination of Harvard ICCB-ICG database systems. Harvard ICCB-ICG, based on the Oracle database system, stores chemical structures, provides chemical inventory support, captures screening information, generates reports to analyze screening results, generates structure-activity relationships (SAR) via its chemical searching capabilities, and identifies compounds that score in common among multiple assays. The latter feature has been especially helpful in performing pathway/network-based assays, as comparisons of related assays can help in identifying potential targets of active compounds. Furthermore, this database provides valuable information on specificity, and can therefore allow compounds to be prioritized for follow-up SAR studies. This multi-dimensional database will also capture imaging data from automated imaging screening systems. Image database is one of the critical and unique features in chemical genomics and medicinal systems biology, as not only a storage tool but also an analytical tool (Swedlow *et al.*, 2003). It is now necessary to adapt and develop standard digital image analysis algorithms for high-throughput and high-content screening. Establishment of this multi-dimensional database, including image and systems informatics, will be critical for small molecule-based formulation and modulation of biological networks in human bio-systems.

Molecular target identifications Genome-based unbiased chemical screening will discover small molecules that target novel proteins in specific pathways and networks. Target identification is one of the key elements of chemical genomic screen with small molecules. To identify such target proteins, several approaches have been developed: (1) high-density, proteome-based affinity purification (Harding *et al.*, 1989; Taunton *et al.*, 1996); (2) genetic/genomic screening

(Launhardt *et al.*, 1998; Glaever *et al.*, 1999); and (3) *in vivo* and (4) *in vitro* expression screening (Caterina *et al.*, 1997; King, 1999; Tanaka *et al.*, 1999). It is necessary to systematically analyze these various approaches to test their feasibility for identification of the proteins whose activity is affected by chemicals. These studies will provide important insights into how, once specific bioactive chemicals are identified with various effective concentrations, their targets can be efficiently identified in chemical genomics. In addition to these approaches to the identification of direct targets as previously described, the target of a small molecule can be identified through systems informatics based on the multi-dimensional database, including image-based profiling with markers for multiple cellular structures and functions as well as genomic (e.g. transcripts) and proteomic (e.g. expressed proteins) profilings (Marton *et al.*, 1998; Butte, 2002; Kuruvilla *et al.*, 2002). In particular, it is helpful to detect molecular target interactions with microarrayed protein and chemicals under genome and proteome scale in a high-throughput manner (Hergenrother *et al.*, 2000; MacBeath and Schreiber, 2000). These parallel and synergistic multi-disciplinary approaches to the various target identification systems would require highly sensitive and efficient detection of protein-chemical interactions *in vitro* and *in vivo*, for example using advanced nano-technologies.

Potential Values and Impacts of Chemical Genomics and Medicinal Systems Biology

Impacts on human biology (1) Mice have been frequently used as a model bio-system of human. However, information from mice is limited because of species-specificity. Thus, many complex questions about the development & disease are best addressed in human cells. Cell-permeable bioactive small molecules are immediately applicable to most biological systems, including those where conventional (e.g. genetic) approaches are difficult, such as primary human cells. A small molecule can be used to rapidly assess the potential role of its target protein in a biological process, without the manipulation of a specific gene. (2) Another important and unique feature in chemical genomics is that a cell-permeable small molecule can affect a target protein within seconds of addition, and its effect can be rapidly reversed by washing out. The only genetic approach with comparable kinetics is temperature sensitive (ts) mutations. Good ts mutations are extremely difficult to find in human cell systems, and temperature-change produces physiological effects of its own. The kinetic advantage of small molecule-based approaches to perturbing protein functions is especially important for a real-time mechanistic analysis of highly dynamic human bio-systems. (3) Related to these dynamic modulations, certain proteins in biological processes turned out to be essential for the survival of human cells. Small molecules can therefore be used to study the function of these essential proteins through

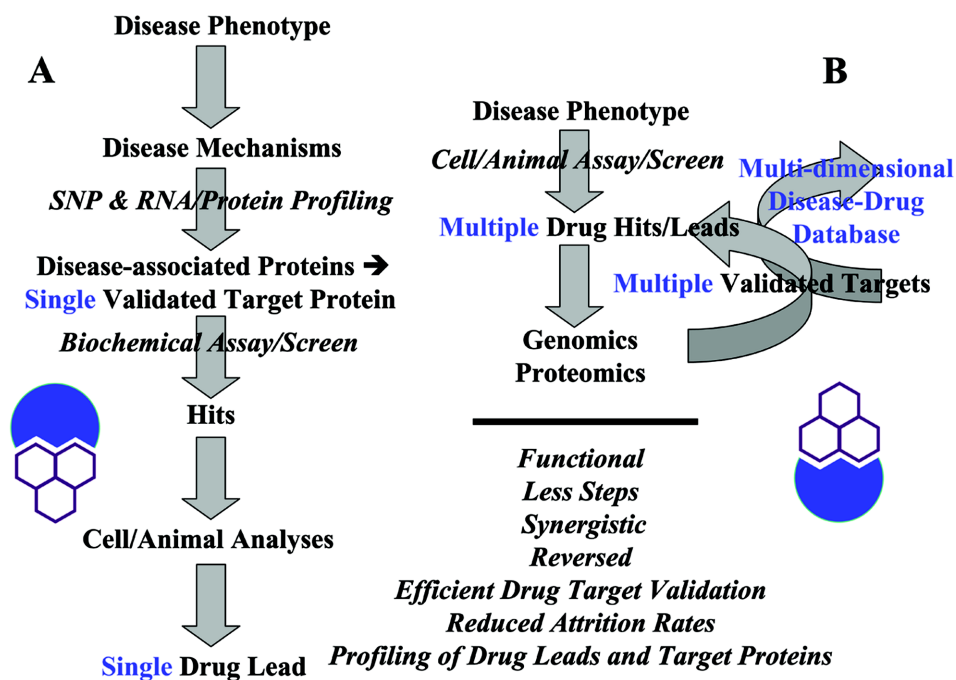


Fig. 3. Comparison of pharmaceutical and chemical genomic approaches to the drug discovery and development. (A) Pharmaceutical approaches. (B) Chemical genomic approaches.

conditional modulation within certain time windows. This is difficult with traditional genetic approaches in human cells. (4) Many complex pathways in human development & disease are specific to different molecular and cellular contexts. For example, tumorigenesis is regulated in a cell lineage-specific manner. Small molecules can be used as unique tools to systematically analyze these processes in different cellular and molecular contexts. Otherwise, the functions of the genetic manipulation of all of the involved genes in different cell types are impractical to analyze.

Impacts on drug discovery and development (1) Small molecule modulators that are screened for disease pathways and networks can lead directly to therapeutics as drug leads. For example, the modulators of tumorigenic pathways that we discovered have generated much interest in pharmaceutical companies as anti-cancer drug leads. (2) Chemical genomic studies will provide the novel therapeutic/diagnostic targets in the pharmaceutical approach. For example, pharmaceutical approaches are typically biased toward finding compounds that interact with validated therapeutic targets (specific proteins with known functions). In contrast, the chemical genomic approach involves screening of small molecules that affect particular pathways/networks or phenotypes, rather than a single protein. Thus, the identification of targets of chemical modulators from these unbiased screens will lead to the discovery of novel regulatory proteins as novel therapeutic/diagnostic targets in the pharmaceutical approach. (3) Related to these unique features, chemical genomics can lead to the efficient profiling of drug leads and therapeutic/diagnostic

target proteins in human diseases (Fig. 3). Furthermore, numerous human diseases, including cancer, result from multi-genetic changes. It is therefore difficult to treat these diseases through the modulation of a single protein in the pharmaceutical approaches. Chemical genomic approaches to the modulations of disease phenotypes may be more effective for these human therapeutics. (4) Several technological platforms in chemical genomics and medicinal systems biology will be incorporated into drug development processes in numerous pharmaceutical companies. For example, as shown in Fig. 3 for the drug discovery and development processes, chemical genomic approaches (B) differ from conventional pharmaceutical approaches (A) in many aspects. These include: (i) functional screening for efficient drug discovery/development based on unbiased cell-based assays toward disease phenotypes and networks; (ii) fewer steps in drug discovery/development when compared to step-wise pharmaceutical approaches; (iii) more synergistic and highly interactive drug discovery/development processes; (iv) bottom-up/reversed drug discovery/development approach; (v) more efficient drug target validations through functional approach; (vi) multiple drugs/therapeutic targets discoveries and profilings; and (vii) establishment of multi-dimensional drug-disease database. These “disruptive drug value chains” may contribute to reducing the high attrition rates of screened drug hits, which is the one of the intrinsic major problems in pharmaceutical approaches. Indeed, several core technologies that were developed at Harvard ICCB-ICG laid the foundation of several companies and the alliance of companies in the United States.

Other potential values in scientific and technological developments The development of chemical genomics and medicinal systems biology will be helpful to lead scientific discoveries to industrial achievements (e.g. drug developments). To this end, chemical genomics and medicinal systems biology involves these key features. These include: (1) translational research that links our basic bioscience (e.g. knowledge in basic cancer cell biology) with applied and clinical bioscience (e.g. discovery of therapeutic genes and drug leads in human cancer) to treat human cancer; (2) multi-disciplinary approach involving pharmacology, biology, chemistry, computer science, engineering, and medicine with holistic and systemic viewpoints on human diseases; (3) creative and innovative research based on higher-order fusion of mainstream approaches (e.g. technologies and concepts in school) with other related synergistic approaches (e.g. those in company); and (4) technology-focused research whose developed technologies and tools can be used as part of a systematic means to explore many aspects of bioscience and bioindustry in a high-throughput manner. These developments will facilitate the establishment of a novel bridging infrastructure for the successful development of both bioscience and pharmaceutical/biotechnological industries.

Acknowledgments I am grateful to all of the chemical genomics members of Harvard ICCB-ICG and KAIST. These chemical genomics studies were initiated and supported by Harvard University, National Cancer Institute, National Institute of Health, Merck & Co., E. Merck & Co., and Keck Foundation in United States; as well as the Korea Brain 21 Project, the Molecular Medicine Research Group Program, the Center for Biological Modulators of the 21st Century Frontier R&D Program, and R&D Program for Fusion Strategy of Advanced Technologies in Korea.

References

- Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D. and Julius, D. (1999) The capsaicin receptor: a heat-activated ion channel in pain pathway. *Nature* **389**, 816-824.
- Clemons, P. A., Koehler, A. N., Wagner, B. K., Sprigings, T. G., Spring, D. R., King, R. W., Schreiber, S. L. and Foley, M. A. (2001) A one-bead, one-stock solution approach to chemical genetics. *Chem. Biol.* **8**, 1183-1195.
- Giaever, G., Shoemaker, D. D., Jones, T. W., Liang, H., Winzeler, E. A., Astromoff, A. and Davis, R. W. (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* **21**, 278-283.
- Harding, M. W., Galat, A., Uehling, D. E. and Schreiber, S. L. (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* **341**, 758-760.
- Hergenrother, P. J., Depew, K. M. and Schreiber, S. L. (2000) Small Molecule Microarrays: Covalent Attachment and Screening of Alcohol-Containing Small Molecules on Glass Slides. *J. Am. Chem. Soc.* **122**, 7849-7850.
- Kim, T. K. (2000) Chemical genetics and chemical genomics: high throughput discovery of drugs and disease genes. *Mol. Biol. News* **1**, 30-32.
- King, R. W. (1999) Chemistry or biology: which comes first after the genome is sequenced? *Chem. Biol.* **6**, R327-R333.
- Komarov, P. G., Komarova, E. A., Kondratov, R. V., Christov-Tselkov, K., Coon, J. S., Chernov, M. V. and Gudkov, A. V. (1999) A chemical inhibitor of p53 that protects mice from the side effects of cancer therapy. *Science* **285**, 1733-1737.
- Kuruvilla, F. G., Shamji, A. F., Sternson, S. M., Hergenrother, P. J. and Schreiber, S. L. (2002) Dissection of a glucose-sensitive pathway of the nutrient-response network using diversity-oriented synthesis and small molecule microarrays. *Nature* **416**, 653-656.
- Launhardt, H., Hinnen, A. and Munder, T. (1998) Drug-induced phenotypes provide a tool for the functional analysis of yeast genes. *Yeast* **14**, 935-942.
- MacBeath, G. and Schreiber, S. L. (2000) Printing Proteins as Microarrays for High Throughput Function Determination. *Science* **289**, 1760-1762.
- Marton, M. J., DeRisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett, D. E. Jr., Hartwell, L. H., Brown, P. O. and Friend, S. H. (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **11**, 1293-1301.
- Mayer, T. U., Kappor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L. and Mitchison, T. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971-974.
- Rosania, G. R., Chang, Y. -T., Perez, O., Sutherlin, D., Dong, H., Lockhart, D. J. and Schultz, P. G. (2000) Myoseverin, a microtubule-binding molecule with novel cellular effects. *Nat. Biotech.* **18**, 304-308.
- Schreiber, S. L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* **6**, 1127-1152.
- Schreiber, S. L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **287**, 1964-1969.
- Strausberg, R. and Schreiber, S. (2003) From Knowing to Controlling: A Path from Genomics to Drugs Using Small Molecule Probes. *Science* **300**, 294-295.
- Swedlow, J., Goldberg, I., Brauner, E. and Sorger, P. (2003) Informatics and Quantitative Analysis in Biological Imaging. *Science* **300**, 100-102.
- Tanaka, A. S., Sliva, M. M., Torquato, R. J., Noguti, M. A., Sampaio, C. A., Friz, H. and Auerswald, E. A. (1999) Functional phage display of leech-derived tryptase inhibitor (LDTI): construction of a library and selection of thrombin inhibitors. *FEBS Lett.* **458**, 11-16.
- Tian, S. S., Lamb, P., King, A. G., Miller, S. G., Kessler, L., Luengo, J. I., Averill, L., Johnson, R. K., Gleason, J. G., Pelus, L. M., Dillon, S. B. and Rosen, J. (1998) A small, nonpeptidyl mimic of granulocyte-colony-stimulating factor. *Science* **281**, 257-259.
- Yarrow, J., Feng, Y., Perlman, Z., Kirchhausen, T. and Mitchison, T. (2003) Phenotypic screening of small molecule libraries by high throughput cell imaging. *Comb. Chem. High Throughput Screen.* **6**, 79-99.
- Walling, L. A., Peters, N., Horn, E. J. and King, R. W. (2001) New technologies for chemical genetics. *J. Cell. Biochem. Suppl.* **37**, 7-12.