0026-895X/07/7201-1–7
MOLECULAR PHARMACOLOGY
U.S. Government work not protected by U.S. copyright
Mol Pharmacol 72:1–7, 2007

Vol. 72, No. 1
35113/3216591
Printed in U.S.A.

MINIREVIEW

Building a Pharmacological Lexicon: Small Molecule Discovery in Academia

John S. Lazo, Linda S. Brady, and Ray Dingledine

Department of Pharmacology, Pittsburgh Molecular Library Screening Center, University of Pittsburgh Drug Discovery Institute, University of Pittsburgh, Pittsburgh, Pennsylvania (J.S.L.); Molecular Screening Centers Network, Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, Bethesda, Maryland (L.S.B.); and Department of Pharmacology, Emory Chemical-Biology Discovery Center, Emory University, Atlanta, Georgia (R.D.)

Received February 11, 2007; accepted March 29, 2007

ABSTRACT

Small molecules are powerful pharmacological tools to dissect biological events. There is now considerable interest in expanding efforts to identify and use small molecules targeting proteins encoded in the genomes of humans and pathogenic organisms. Integration of the principles of molecular pharmacology with contemporary high-throughput and high-content screening technologies is essential for the success of these discovery activities. We present some of the challenges and opportunities provided by the Molecular Library Screening Centers Network (MLSCN), which is a National Institutes of Health Roadmap Initiative.

"Four hundred years ago there was not such convenience available on any English bookshelf."—Simon Winchester, *Professor and the Madman.*

Can you image writing an article or a book without a dictionary or a thesaurus? How would you find the definition of exon or intron, prion or ion? That is what William Shakespeare faced when he wrote "*A Midfommer nights dreame*." The first English dictionaries appeared at about the time of his death and were arranged by subject not alphabetically. The English language was spoken and written, but it was not defined. Perhaps pharmacologists a few decades from now will have similar thoughts when they consider the status of the Pharmacology and Chemical Biology of today. We take for granted that small molecules have biological effects, but we lack a readily accessible annotated database of the pharmacological effects of chemicals. One attempt to remedy this is an effort by the National Institutes of Health's Roadmap Initiative. The purpose of this article is to explore the potential and describe recent activities of the Roadmap's Molecular Library Screening Center Network (MLSCN) (http://www.mli.nih.gov/mlscn/).

The power of readily available lexicons is evident even to a nonscientist. Consider the venerable Oxford English Dictionary (OED). Some would argue that the current hegemony of the English language rests largely on the development of the OED, which was a radical experiment sponsored by the Philological Society of London. On April 26, 1878, the Philological Society invited Professor James Murray to edit what would become the OED. The uniqueness of the project was effectively captured in the book *The Professor and the Madman*, written by Simon Winchester (Winchester, 1999), and it was partially the inspiration for this article. Professor Murray needed to convince English-speaking people to scan the literature for the first use of words. The thought of asking tens of thousands of people to voluntarily comb through their book collections and local libraries and, without any compensation, donate their findings to an Oxford professor might seem futile or even mad. Nonetheless, starting with A and proceeding over several decades, they defined 414,825 words and

This work supported in part by National Institutes of Health Grants 5U54-HG003918 (R.D.) and 5U54-MH074411 (J.S.L.).
Article, publication date, and citation information can be found at http://molpharm.aspetjournals.org.
doi:10.1124/mol.107.035113.

ABBREVIATIONS: DMSO, dimethyl sulfoxide; HEK, human embryonic kidney; HTS, high-throughput screening; MLSCN, Molecular Library Screening Centers Network; OED, Oxford English Dictionary; TR-FRET, time-resolved fluorescence resonance energy transfer; MLSMR, Molecular Libraries Small Molecule Repository.

illustrated them with 1827,306 textual examples, finally fixing the magnitude of the English language. The madman in this story, by the way, was the prime OED contributor, Dr. William Chester Minor, an American Civil War captain and Yale-trained physician who was also an inmate at an asylum for the criminally insane. The OED model of public contribution to build language lexicons lives on in the current Wikipedia experiment (http://en.wikipedia.org).

We also have two powerful examples of the consequences of freely available lexicons in biology: PubMed and the Human Genome Project. It is difficult to remember the days before Web-based searches of hundreds of genomes with publicly available free software programs. Nevertheless, the initiation of the Human Genome Project was surrounded with controversy (Palca, 1989; Roberts, 1989, 1990) that is reminiscent of the contemporary public debate over funding precipitated by the new National Institutes of Health Roadmap Initiative (Bravo, 2006; Marks, 2006; Weissmann, 2006). Public debate about the allocation of National Institutes of Health funds would seem to be an inevitable and productive process associated with public funding mechanisms and has been addressed elsewhere (Lazo, 2006). In this review, we explore how small molecule screening and the MLSCN could affect the future of pharmacology.

Should I Use Chemistry or Genetics to Probe Protein Function?

The field of pharmacology, now more than 100 years old, has always been at the interface between chemistry and biology. The earliest pharmacologists relied on crude organ preparations or animal models to probe how recently isolated or extracted compounds, such as cocaine, morphine, or atropine, functioned. The advent of more reductionist approaches revealed molecular, even atomic, interactions of small molecules with protein targets. The Human Genome Project resulted in the discovery of thousands of new genes encoding tens of thousands of proteins. Now many are attempting to integrate the knowledge obtained from molecular biological and genetic studies in the context of the whole organism. How can we develop, systematically and relatively quickly, an understanding of the biological functions of all these new proteins? Sophisticated gene knockout technologies can produce highly specific deletion of individual proteins, although homeostatic compensation must always be borne in mind. Moreover, knockout technologies are currently restricted to just a few species and are often irreversible, leading to a hypomorph lacking half of or the entire gene product. Like silencing RNA strategies, gene knockout usually eliminates the entire protein, which reduces the value of these approaches for studying splice variants or the functions of multiple independent domains within an individual protein.

As an alternative to the genetics approach, small molecules can provide powerful tools to dissect biological events. They allow one to reversibly affect protein function in a graded rather than all-or-none fashion, modify the subcellular location of a macromolecule, or disrupt a specific protein-protein interaction. The motto of small molecule users is "any species, any place (in the body), any time (during development)." Moreover, small molecules can be designed that inhibit *or* potentiate protein function and can be used to probe the function of individual subunits in a multimeric protein complex, or even different domains of the same protein subunit. Finally, a selective, potent small molecule targeted to a disease-related protein can serve as a lead for therapeutic development. However, achieving sufficient target selectivity is often very difficult, making the small molecule approach complementary to genetics rather than a substitute.

The recognition that small molecules can be powerful reagents has fueled enthusiasm for using high-throughput screening (HTS) and high content screening methods to identify biologically friendly small molecules that would be readily available to scientists. Evidence for the power of such endeavors are found throughout the contemporary pharmacological literature (Oltersdorf et al., 2005; Huang et al., 2006; Sanna et al., 2006).

Differentiating Academic and Industry Goals. The phrase "compound screening" sometimes evokes a disapproving remark in the halls of academia, even though other screening activities, such as screening a DNA library for expression differences or novel genes, are well accepted and occasionally even admired. This might reflect a belief that compound screening is something that should be relegated to an industrial environment, being an applied task that is not hypothesis-driven. That small molecule screening is reagent building cannot be denied, but there are many chemical biology screens that provide valuable probes and yield unique biological insights. Not to seek and use these reagents seems foolish. Indeed, innovative and high-impact advances in therapeutics will probably come from aggressive efforts to provide a bridge that allows translation of advances in basic science to novel therapeutics and marketable products. This bridge is the identification of novel small molecules and their mechanisms of action that specifically perturb the function of disease-related proteins studied in academic laboratories. Three developments in the past 5 years have made the small molecule-based chemical genetics approach a realistic goal in academic institutions: the development of commercially available small molecule libraries, a reduction in cost of screening instrumentation, and the beginning of a flow of talent from industry back to academia. Currently there are more than 30 academic small molecule screening centers, including the MLSCN centers (Gordon, 2007).

The mission and architecture of academic and industrial compound screening can be readily differentiated. A significant fraction of the research and development budget for compound screening in large pharmaceutical companies is focused on identifying potential "blockbuster" drugs. Thus, 90% of the current commercial research and development resources are spent on only 10% of the current worldwide human disease burden (Munos 2006). Academic investigators are often less encumbered by profit motive limitations when they select targets for screening. Moreover, investigators in the nonprofit sector often are just seeking potent biological probes, and the resulting compounds need not possess the requisite pharmacokinetic and metabolic profiles for a good drug. Thus, academics are free to interrogate a chemical library that may not be viewed as "drug-like" but still has pharmacologically unique components. Funding for academic scientists often has longer time lines, measured in half decades, than that for industrial scientists, who must place a premium on flexibility and rapid assay development and analysis as corporate goals change. Finally, the product of academic screening is inherently open access, such as in

PubChem (see below) or ChemBank, whereas the result of an industrial screening exercise generally remains proprietary.

The differences in core missions between the public and private sector means there is opportunity for mutual benefits or collaborations. Indeed, there are some interesting public-private partnerships that have already emerged (Munos, 2006). One of these is the Medicines for Malaria Venture, which was established in 1999 to discover and develop new and affordable antimalarial drugs. This group has brought 40 public and private institutions together in a network comprising 300 scientists. Another example is the GAVI Alliance (http://www.gavialliance.org), which brings together private foundations, national governments, UNICEF, WHO, The World Bank, the vaccine industry, and public health institutions in a $3 billion effort to increase children's access to vaccines in poor countries. The Initiative on Public-Private Partnership for Health (http://www.ippph.org) lists 92 different public-private partnerships focused on neglected diseases. These recent developments, most within the last 5 years, could have a major effect on the research focus of some academic centers. Furthermore, academic centers that support small molecule screening should be in an excellent position to train the next generation of drug discovery scientists.

National Institutes of Health Roadmap and Contemporary Pharmacology. The National Institutes of Health has launched an ambitious program to optimize its entire research portfolio, called the Roadmap Initiative (http://nihroadmap.nih.gov). One of the three main areas of the initiative is the New Pathways to Discovery component, which is expected to empower the research community with small molecule compounds for tools to perturb genes and pathways, as imaging probes in basic or clinical applications, or as starting points for the development of new therapeutics for human disease. To achieve this goal, the National Institutes of Health has established 10 new MLSCN Centers. The eventual goal of each MLSCN Center is to conduct 15 assays with 300,000 compounds each year and to deposit these 4.5 million assay results on PubChem for public use.

How Do I Get Involved in the MLSCN?

Overview of Target-to-Probe Process. Academic investigators typically bring a Zen-like intensity to their attempts to understand the biological functions of their favorite proteins. When it becomes important to develop a small-molecule tool for more detailed exploration, the MLSCN stands ready to help. R03 and R21 National Institutes of Health grant mechanisms are available to prepare your assay for HTS, often in collaboration with one of the existing MLSCN Centers (see below). In your own laboratory (or in collaboration with one of the MLSCN Centers), it is first necessary to convert your bench top assay from microcentrifuge tube or microscope slide to 96- or 384-well plate format, by carrying out the experiments explained in the next section. From this point, the process is summarized in Fig. 1. The next step is to submit an R03 or X01 grant application to National Institutes of Health (http://www.mli.nih.gov/funding/assay_solic_fund_ops.php). Once the project is accepted and assigned to a Center, further assay optimization and, if necessary, miniaturization, is carried out at the Center, then a high-throughput primary screen of up to 100,000 compounds is performed generally at a single compound concentration (typically 10–30 μM), although one center now routinely engages in concentration profiling as a primary screening approach (Inglese et al., 2006). Hits are confirmed in one or more secondary screens, concentration-response relations are obtained to inform structure-activity analysis, and a decision is made whether to invest medicinal chemistry resources of the network to the project. Assuming a positive answer, a wide range of cheminformatics tools is brought to bear, both to identify commercially available analogs of promising hits and to assist medicinal chemists in the design and synthesis of novel analogs for bioassay. One or more iterations of synthesis and bioassay are performed until the desired goal of the project is achieved. All screening data are deposited in a public database, PubChem (see below). Thus, the dual deliverables of the MLSCN are novel, useful chemical probes for the biomedical investigator and a public database relating chemical structures to biological effects.

From Bench to Robot. Adaptation of a successful bench assay to high-throughput mode requires optimizing the robustness and stability of the assay readouts, taking into consideration cost per well as well as labor costs. From a practical point of view, it is good to keep in mind that if reagent costs are even as low as $0.20 per well, a 100,000 compound primary screen (without duplicates) will cost $20,000 plus labor. The MLSCN is gearing up for 300,000 compound screens, and therefore it is highly worthwhile to spend time up-front optimizing the assay. A detailed description of assay optimization for high throughput screening has been compiled by investigators at Eli Lilly and the Chemical Genomics Center at National Institutes of Health, available at http://www.ncgc.nih.gov/guidance/manual_toc.html. Very briefly, the process typically begins by optimizing the concentration of protein and substrate in each well for a biochemical assay, or the number of cells per well in the case of cell-based assays. The parameters to optimize are cost of reagents, the signal-to-noise or signal-to-background ratio, and the Z′ factor. The Z′ factor measures the quality of the assay itself without intervention of test compounds. This measure of assay robustness is calculated by the equation: $Z' = 1 - 3 \times (SD_s + SD_b)/(\mu_s - \mu_b)$, where the subscript s refers to the maximum assay signal (e.g., in the presence of a screening concentration of agonist), subscript b is the minimum signal

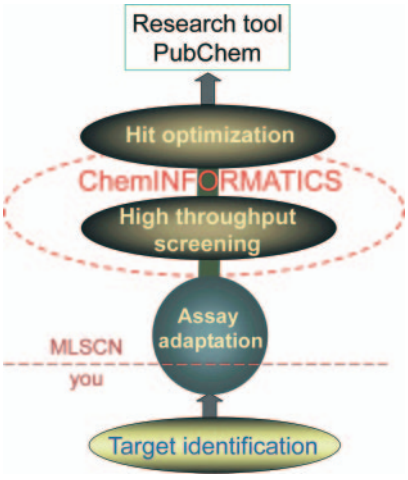

Fig. 1. Overview of the probe development process.

(e.g., in the absence of agonist), SD is the standard deviation, and μ is the mean signal in each condition. The signal-to-background ratio (S:B) is defined as μ_s/μ_b, and the signal-to-noise ratio (S:N) is defined by the equation (S:N) = $(\mu_s - \mu_b)/[\sqrt{(SD_s^2 + SD_b^2)}]$.

A simple example for a cell-based cytotoxicity assay using A549 lung tumor cells in culture is shown in Fig. 2A. The assay is based on the ability of healthy, viable cells to reduce resazurin to resorufin, which is excited at 560 nm and emits fluorescence at 590 nm. Injured or dead cells lack active reducing enzymes. Various numbers of cells ($5–15 \times 10^4$) were seeded into 96-well plates and treated with vehicle or 1 μM doxorubicin for 24 h. Resazurin was then added to each well, and the reaction was allowed to proceed in the culture incubator for 15 min to 24 h before reading the plate in an Analyst HT multimode plate reader. Best results were obtained after 3.5- to 4-h incubation. A plot of Z′ and S:B against cell density after 3.5 h of incubation (Fig. 2A) indicates that 1 to 2×10^4 cells per well yield an optimum Z′ (= 0.8) and signal-to-background (S:B) ratio of 12, although these robustness measures are not much reduced if only 4000 cells per well are used. As a rule of thumb, conditions resulting in Z′ > 0.5 and S:B > 5 produce an assay robust enough for most high-throughput screens.

The next step in assay optimization is to determine whether plate-to-plate and week-to-week variability is adequate. Figure 2, B and C, shows results from four 96-well plates repeated on 3 separate weeks, each week representing a different cell splitting. The assay shows acceptable stability over the 3-week period, without evidence of "edge effects" (data not shown). It is important to evaluate the effect of DMSO on the assay because DMSO is the vehicle for the compound libraries and so is often present at concentrations of 0.5 to 2% in the assay. In this case, Z′ was reduced to ~0.65 in the presence of 1% DMSO (data not shown). Finally, it is important to show that a positive control, here doxorubicin, affects the assay readout in a stable manner over a multiday test period. Some assays, for example of novel targets, may

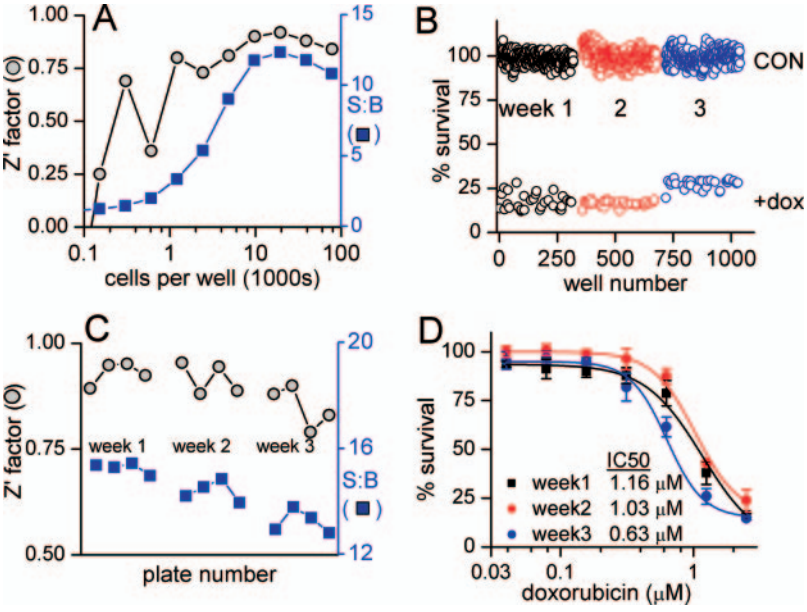

Fig. 2. Optimization of an Alamar blue cytotoxicity assay for A549 cells treated with doxorubicin in 96-well plates. A, Z′ factor and S:B as function of number of cells per well. B and C, stability of assay over 3 weeks (four plates per week). D, doxorubicin concentration-inhibition curve is stable over 3 weeks. Y. Du and H. Fu, unpublished.

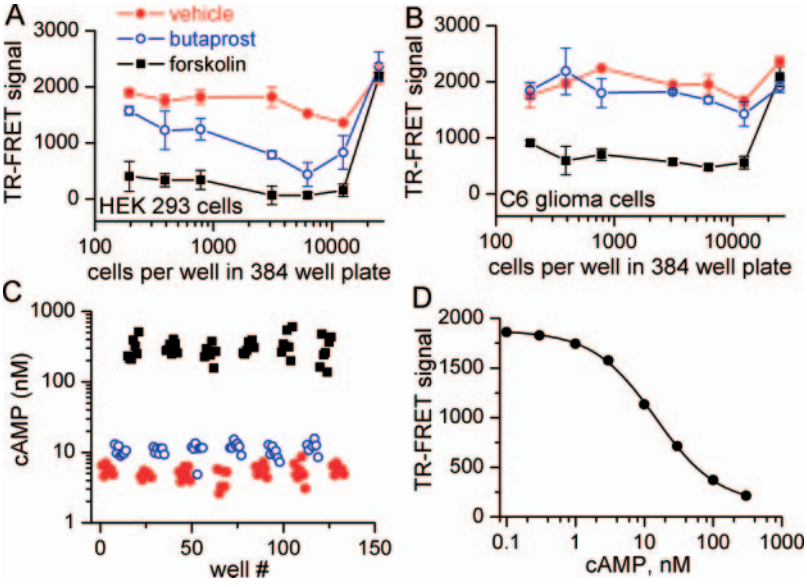

Fig. 3. Activation of EP2 receptors in HEK293 cell line measured by immuno-FRET based assay for cAMP. A and B, raw TR-FRET signals show that HEK293 but not C6 glioma cells respond to butaprost, whereas forskolin activates cAMP formation strongly in both lines. C, assay stability across a 384-well plate (symbols as in A), with FRET signal converted to cAMP concentration by standard curve in D. R. Dingledine and J. Wetherington, unpublished.

not have positive controls. Screening a small library, such as the Library of Pharmacologically Active Compounds (LOPAC), often results in identification of a compound that can be used as a positive control for the purpose of assessing assay quality, even if its mechanism of action is not the desired one. Figure 2D shows that the IC_{50} of doxorubicin varied less than 2-fold over this 3-week period. This cell viability assay was viewed as ready for high-throughput screening of small molecules applied in a final concentration of 25 μM (1% DMSO).

Another example is an assay being developed to screen for potentiators or novel agonists of the EP2 prostanoid receptor, which activates adenylyl cyclase. An HEK293 cell line that expresses human EP2 receptors was used to optimize a time-resolved fluorescence resonance energy transfer-based immunoassay for cyclic AMP formation in response to the selective EP2 agonist butaprost. The method depends upon competition by cell-derived cAMP for binding of labeled cAMP to a cAMP antibody. The FRET donor, an anti-cAMP antibody conjugated to europium cryptate, is excited at 337 nm and emits at 620 nm. The half-life of europium emission is several hundred microseconds or longer, so in practice, a time delay of 50 to 100 μs is imposed between excitation and emission readings to allow the intrinsic fluorescence of library compounds (typical half-life of hundreds of nanoseconds) to subside. The FRET acceptor is a cAMP molecule conjugated to a fluorophore that is excited at 620 nm and emits at 665 nm. The TR-FRET signal is then the ratio ($\times 10^4$) of emission readings at 665 and 620 nm. A ratiometric measure reduces well to-well variability as a result of the presence of colored compounds, phenol red in culture medium, etc. As expected, the FRET signal decreases as cAMP concentration rises (Fig. 3D). In the experiment, HEK293 or C6 glioma cells seeded into a 384-well plate were treated for 30 min with vehicle (0.3% DMSO), forskolin (a strong activator of adenylyl cyclase), or butaprost, all in the presence of 200 μM 3-isobutyl-1-methylxanthine to block phosphodiesterases. Figure 3 shows that forskolin strongly reduces the FRET signal (i.e., elevates cAMP) in both HEK and C6 cell lines, but butaprost is an effective activator of cAMP production only in the HEK293 cell line. A cell density of 3000 cells per well provided the optimal signal-to-background ratio (Fig. 3A). Experiments with butaprost or forskolin as agonist indicate adequate signal stability across a 384-well plate (Fig. 3C), with $Z' = 0.62$ and S:B = 9. The next steps for assay development would involve evaluation of plate-to-plate and week-to-week stability.

TABLE 1
Core capabilities of the MLSCN network

Assay Expertise	Technologies
Enzymes/proteases	Ultra HTS
G-protein coupled receptors	Quantitative HTS
Kinases	High-throughput microscopy
Ion channels/transporters	Flow cytometry
Cytotoxicity	NMR-based methods
Protein-protein interactions	Virtual screening
Protein misfolding/degradation	
High-content screens	
Yeast-based assays	
Zebrafish	

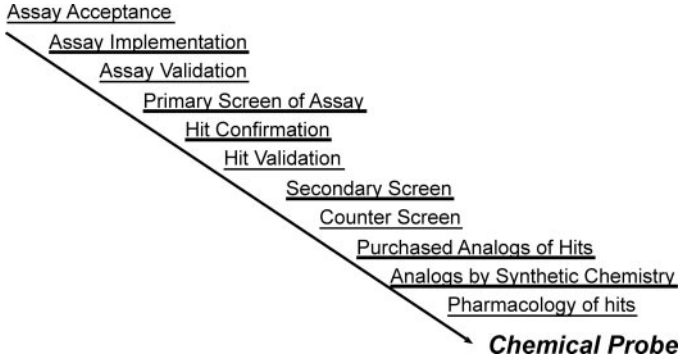

Fig. 4. HTS assay campaign: path from assay to probe.

What Does the MLSCN Offer Pharmacologists?

Diverse Screening Platforms. The MLSCN offers the research community an opportunity to leverage biology and chemistry resources to identify small-molecule probes for innovative or challenging targets with the potential for insight into biological pathways that affect public health. The MLSCN initiative is a consortium of 10 centers that, in its initial pilot phase, is developing the capacity to screen a large set of shared compounds (>100,000) maintained in a central repository, in a highly diverse set of assays solicited from the scientific community. Simultaneously, the centers are developing informatics and chemistry capacities to optimize "hits" identified in the initial screening to produce chemical probes that can be used for in vitro studies to interrogate the targets or phenotypes studied in the assays. All of the HTS screening data from the MLSCN assays, including functional and pharmacologic selectivity data, are deposited into PubChem (http://pubchem.ncbi.nlm.nih.gov/), an open access database that provides biological and chemical information to researchers for use in studying biology and disease.

Access to MLSCN Resources. Launched in June of 2005, the 10 MLSCN centers comprise a network of shared management, experience, and expertise that provides greater capability and productivity for biological discovery than would be found in any single academic center. Together, the centers network has an array of cutting-edge HTS technology

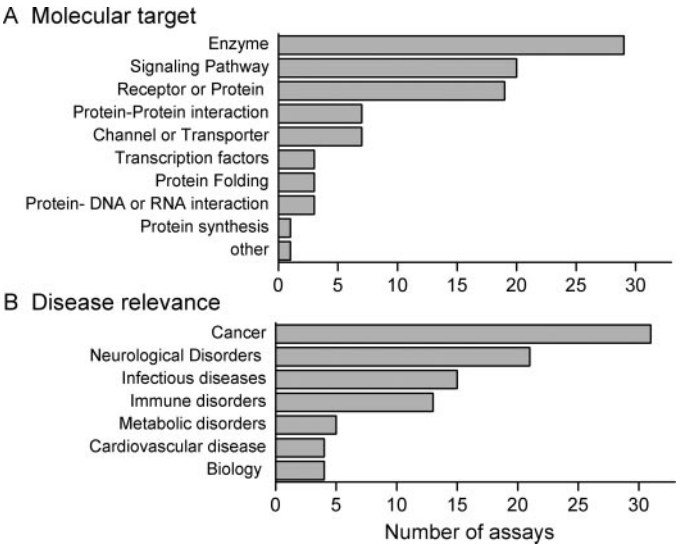

Fig. 5. Diversity of assay targets and disease relevance of HTS assays assigned to the MLSCN.

and detection platforms that include ultra-HTS capabilities afforded by Kalypsys technology, quantitative HTS approaches (Inglese et al., 2006), high-throughput microscopy for phenotypic screening, NMR-based approaches, and flow cytometry that allows for multiplexing of related assay targets (Table 1). The comprehensive technology and detection platforms give the network the ability to implement a diversity of target-based (e.g., G-protein-coupled receptors, ion channels, transporters, kinases, enzymes), cell-based, phenotypic high content screens, and model organism-based assays that address areas of pharmacologic research in need of a chemical probe to dissect signaling pathways and drive biological discovery. More detailed information about the capabilities and expertise of each of the centers can be found at http://www.mli.nih.gov/mlscn/descriptions.php.

The network of centers has gained experience implementing a diverse array of HTS assays and has developed the capacity to conduct nearly 150 assay campaigns per year against the library of 150,000 compounds, generating more than 10 million data points in PubChem to date. Collectively, the centers can conduct most assays, including biochemical and cell-based, phenotypic, multiplexed, infectious agent-based, and a limited number of whole organism-based assays. The National Institutes of Health is soliciting HTS assay projects directed toward novel proteins, cellular phenotypes, biological functions, and disease mechanisms from the research community (http://grants.nih.gov/grants/guide/notice-files/NOT-RM-07-003.html). This program provides scientists access to MLSCN resources at no cost to the investigator. HTS assay projects are selected for implementation based on peer review. Projects are prioritized based on the significance of the biological target, need for a chemical probe, and HTS readiness, and the assays are then assigned to each MLSCN center based on its expertise and technology. Collaboration among the assay provider, MLSCN biologists, and chemists is critical for development of a critical path work plan for progressing from assay to probe, which describes assay automation, secondary and counter-screening assays, and identification of hits that are amenable to structure-activity relationship, computational modeling, and synthetic chemistry to optimize small molecule probes (Fig. 4). Through February 2007, 92 assays have been selected for HTS campaigns by the MLSCN. The assays span a variety of target classes (Fig. 5A) and disease relevance (Fig. 5B).

TABLE 2
Chemical composition of the Molecular Libraries repository

Compound Class	Number
Diverse structures	109,829
Targeted libraries	
GPCR	2339
Kinases	2677
Ion channels	1676
Proteases	662
Nucleases	110
Specialty sets	853
Non-commercial compounds from PSLs, CMLDs, MLSCN hit followup	24,489
Natural products	<1000
Total	143,635

Access to a High-Quality Compound Repository. The MLSMR, established in September 2004, currently houses a collection of nearly 115,000 chemically diverse molecules with both proven and unknown biological activities. Biofocus DPI/Gallapagos collects, maintains, and distributes the compound library to the MLSCN centers (http://mlsmr.glpg.com/MLSMR_HomePage/) for HTS. The repository currently contains diverse compounds, targeted libraries (e.g., G-protein-coupled receptors, ion channels, nuclear receptors, kinases, proteases), natural products, a specialty set of compounds with known biological activity, including approved drugs, failed clinical candidates, veterinary medications, toxins, metabolites, etc., and novel chemical structures from the Molecular Libraries Pilot Scale Libraries (PSL) Program (http://

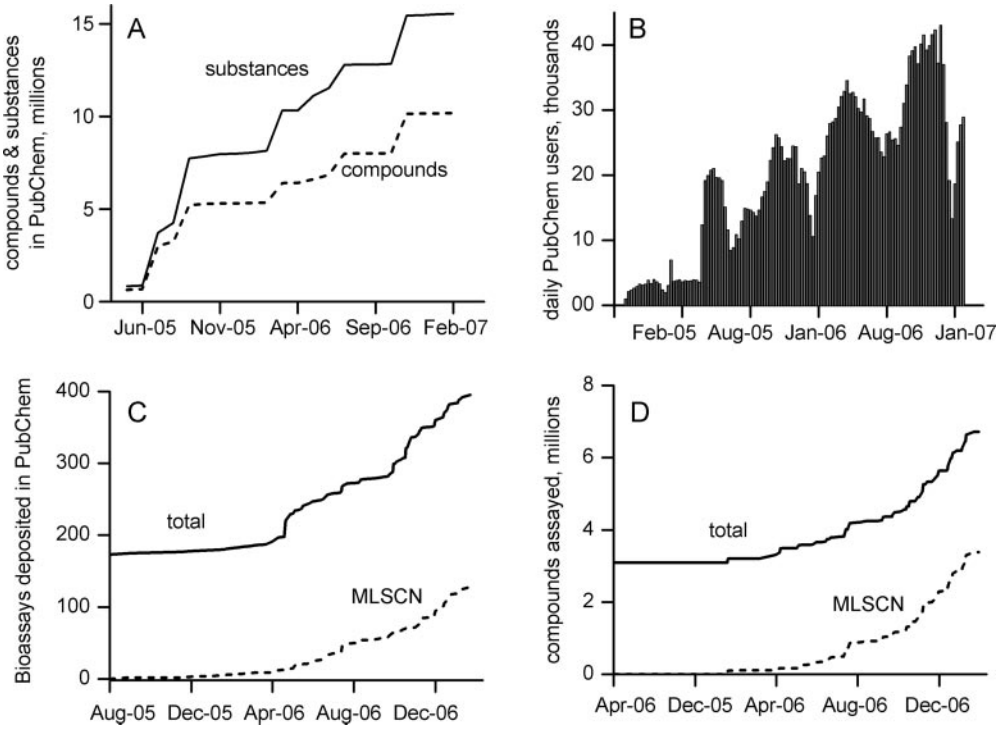

Fig. 6. Growth of PubChem composition and usage.

mli.nih.gov/funding/chem_div_fund_res_pilot.php) (Table 2). The Chemical Methodologies and Library Development (CMLD) centers, which are supported by the National Institute of General Medical Sciences (http://www.nigms.nih.gov/Initiatives/CMLD/Centers/), have also begun to contribute novel structures to the compound repository.

Over the next 3 to 5 years, the repository will grow to 500,000 compounds. There will be commercial acquisitions based on rational selection strategies and efforts to obtain novel compounds from noncommercial sources. The PSL initiative will continue to expand the library by generating small molecules to explore areas of "chemical diversity space" through target-oriented synthesis of complex molecules, and natural products isolation and derivatization (http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-06-003.html). In addition, the MLSMR is seeking unique chemical contributions, preferably in solid form, from all sources meeting the criteria of at least 90% purity, sufficient water solubility for use in HTS, and reasonable stability at room temperature. Compounds obtained by high-throughput synthesis, medicinal or synthetic organic chemistry, and purified discrete natural products from microorganisms, plants, or marine organisms are of interest (http://grants.nih.gov/grants/guide/notice-files/NOT-RM-07-005.html). Over time, the use of the shared library by the MLSCN should provide extensive biological annotation, generating a unique and rich dataset available in the public domain.

Access to Biological and Chemical Datasets. PubChem is a public sector cheminformatics database of small organic molecule modulators developed by the National Center for Biotechnology Information and launched in September of 2004 (http://pubchem.ncbi.nlm.nih.gov/). It is an online resource providing comprehensive information on the biological activities of small molecules and search, retrieval, and data analysis tools such as structure search, structure-activity analysis, and structure clustering to optimize the utility of chemical structure and bioactivity information, as well as integration with other National Institutes of Health biomedical information sources such as PubMed and Genome, Protein, and Structure databases. To date, PubChem contains over 15.5 million substance records from more than 50 depositors (Fig. 6A), more than 6 million unique compound structures with links to bioassay descriptions, relevant literature, references, and assay data points (Fig. 6D), and nearly 400 bioassay data sets; more than 128 of these have been contributed by the MLSCN thus far (Fig. 6C). Examples of MLSCN bioassay summary data including target/assay description, assay protocol, and definition of compound bioactivity can be found at (http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=439; http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=360). Examples of probe reports generated by the MLSCN centers can be found at (http://molscreen.florida.scripps.edu/probes.html; http://ncgc.nih.gov/db/?aid=103). As the deposition of chemical structures and assays continue to rise over time, so does the number of users (Fig. 6B), making this a valuable resource to the public and private sector.

Conclusions and Summary

The MLSCN, along with PubChem and the MLSMR, offers a new dimension in research opportunities for pharmacologists, chemists and biologists in the academic and nonprofit sector. The sharing of small molecules, biological assays, and screening data with the larger scientific community represents a new public sector paradigm that promises to facilitate the understanding of basic biological mechanisms and shorten the timeline for drug development, with resulting benefits to public health, especially for rare and neglected disorders.

Acknowledgments

We appreciate the assistance of Drs. Carson Loomis, Ingrid Li, Jamie Driscoll, and Steve Bryant for providing the data for Figs. 4 to 6 and Table 2.

References

Gordon EJ (2007) Small molecule screening: it takes a village. *ACS Chem Biol* **2:**9–16.

Huang H, Regan KM, Luo Z, Chen J, and Tindall DJ (2006) CDK2-dependent phosphorylation of FOXO1 as an apoptotic response to DNA damage. *Science* **314:**220–323.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, and Austin CP (2006) Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc Natl Acad Sci U S A* **103:**11473–11478.

Lazo JS (2006) Roadmap or roadkill: A pharmacologist's analysis of the NIH Molecular Libraries Initiative. *Mol Interv* **6:**240–243.

Marks AR (2006) Rescuing the NIH before it is too late. *J Clin Invest* **116:**844.

Munos B (2006) Can open-source R&D reinvigorate drug research? *Nat Rev Drug Discovery* **5:**723–929.

Oltersdorf TS, Elmore W, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, Bruncko M, Deckwerth TL, Dinges J, Hajduk PJ, et al. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **435**:677–681.

Palca J (1989) Hard times at NIH. *Science* **246:**988–990.

Roberts L (1989) Plan for genome centers sparks a controversy. *Science* **246:**204–205.

Roberts L (1990) Tough times ahead for the genome project. *Science* **248:**1600–1601.

Ruiz Bravo N (2006) Response to: "Rescuing the NIH before it is too late" from the Deputy Director for Extramural Research. *J Clin Invest* **116:**1463–1464.

Sanna MG, Wang SK, Gonzalez-Cabrera PJ, Don A, Marsolais D, Matheu MP, Wei SH, Parker I, Jo E, Cheng WC, et al. (2006) Enhancement of capillary leakage and restoration of lymphocyte egress by a chiral S1P(1) antagonist in vivo. *Nat Chem Biol* **2:**434–441.

Weissmann G (2006) Planning science (a generation after Lewis Thomas). *J Clin Invest* **116:**1463.

Winchester S (1999) *The Professor and the Madman*, HarperCollins, New York.

Address correspondence to: John S. Lazo, Department of Pharmacology, Pittsburgh Molecular Library Screening Center, University of Pittsburgh, Biomedical Science Tower 3, Suite 10040, Pittsburgh, PA 15261-0001. E-mail: lazo@pitt.edu; Raymond Dingledine, Department of Pharmacology, Emory Chemical-Biology Center, 1510 Clifton Rd., Emory University, Atlanta, GA 30322. E-mail: rdingledine@pharm.emory.edu