

## NIH Molecular Libraries Initiative

Christopher P. Austin,<sup>1\*</sup> Linda S. Brady,<sup>2</sup> Thomas R. Insel,<sup>2</sup> and Francis S. Collins<sup>1</sup>

The purpose of the Molecular Libraries Initiative (MLI) component of the NIH Roadmap for Medical Research (1, 2) is to expand the availability, flexibility, and use of small-molecule chemical probes for basic research. Because this initiative is particularly novel and far-reaching, it has been the subject of considerable discussion (3–5), and sometimes misinterpretation (6), in the research community.

Two imperatives motivated the development of the MLI. The first, related to NIH's mission in basic biomedical research, was the need for fundamentally new approaches to determine function and therapeutic potential for all genes in the newly sequenced human genome. The second, related to NIH's mission to improve public health, was the need to accelerate the translation of basic research discoveries into new therapeutics.

Expense (particularly large capital costs), expertise, and cultural divides between public and private sectors have historically kept discovery and optimization of small molecules largely restricted to pharmaceutical and biotechnology companies. Dissemination of small-molecule research tools and technologies into the public sector via the MLI is timely for several reasons.

First, sequencing of the human genome has provided an abundance of new targets for study, and small-molecule research tools will accelerate the translation of genome sequence into biological and therapeutic insights (7). Small molecules are complementary to nucleic acid–based translational tools such as knockout mice and siRNAs, in that they target the protein gene product rather

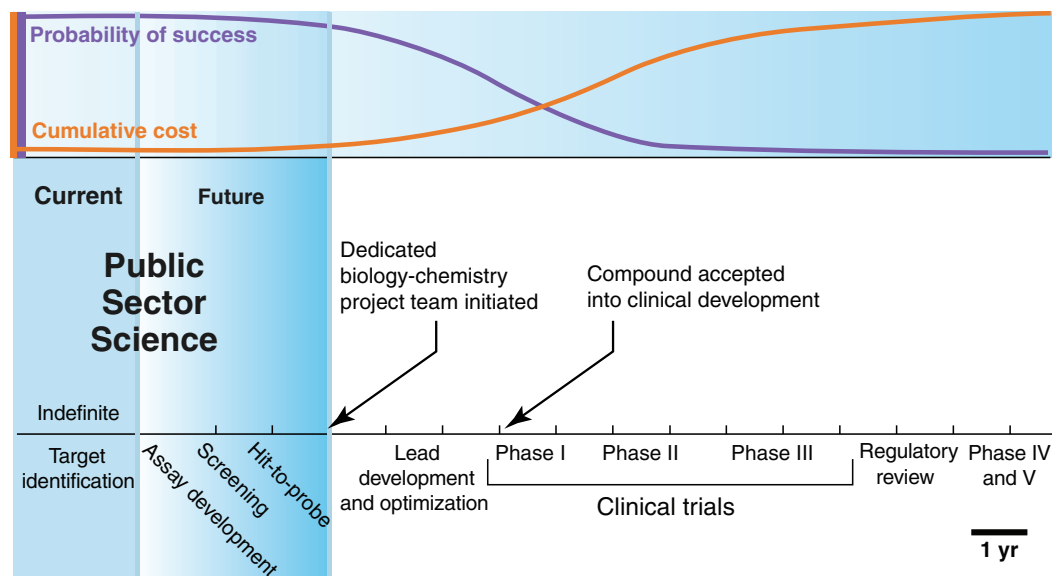
than the gene locus or mRNA, have virtually limitless structural diversity, can affect particular target functions for defined periods in isolated proteins, cells, or organisms, and can serve as either agonists or antagonists. The characteristics that make this class of molecule useful as drugs—their potential for selectivity, cell permeability, and subtle reversible modulation of important physiological functions—also make them good research tools for dissecting the functions of novel genes, pathways, and cells.

The human genome encodes 20,000 to 25,000 genes (8) and perhaps a million proteins, of which only ~500 are targeted

ers of high-quality compound libraries, small molecules can now be obtained on a large scale. At the same time, advances in robotics and informatics have made screening and analysis of such large compound libraries possible. Up to a million compounds can now be screened against a target in a single day, three orders of magnitude greater than was possible only a decade ago. Together, these developments make a public-sector small-molecule screening and chemistry initiative such as the MLI possible.

The MLI was developed over the course of 9 months through consultations with representatives of multiple NIH institutes, and external consultants from the public and private sectors. The MLI research agenda has three components focused on screening, cheminformatics, and technology development, and is being carried out via NIH grant and contract mechanisms (11).

The Molecular Libraries Screening Center Network (MLSCN) will be a consortium of five or six high-throughput



Interface of the MLI and drug development.

by currently available small molecules (9). For the most part, pharmaceutical and biotechnology companies prefer to focus on the “druggable genome” thought to be more amenable to drug development (10). The majority of the genome that is currently considered “undruggable” (i.e., unmanipulable by small molecules) is therefore a major focus of the MLI.

Large libraries of small molecules have traditionally been unavailable to academic researchers, but with the advent of combinatorial chemistry and commercial suppli-

screening (HTS) centers that will screen assays submitted by the research community on a large number of compounds (>100,000) maintained in a central compound repository, and perform optimization chemistry required to produce in vitro chemical probes of the targets or phenotypes studied in the assays (12). All results will be placed into a new public database (PubChem, see below), and probe compounds will be made available without encumbrance to all researchers, in public and private sectors, for their use in studying bi-

<sup>1</sup>National Human Genome Research Institute and <sup>2</sup>National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA.

\*Author for correspondence. E-mail: austinc@mail.nih.gov

ology and disease. The first of the MLSCN screening centers (the NIH Chemical Genomics Center) has been established within the NIH intramural program. It will begin full-scale screening in early 2005 (13). The other MLSCN centers will be funded via extramural grants; applications to this NIH Roadmap Request for Applications (12) are under review and will be awarded in the late spring of 2005. The contract for the Molecular Libraries Small Molecule Repository, which will house the screening collection, was recently awarded (14), and the composition of the compound collection is being determined by a distinguished panel of chemists from the public and private sectors.

A comprehensive database of chemical structures and their activities, PubChem (15), has been developed by the National Center for Biotechnology Information at the National Library of Medicine/NIH. PubChem links small-molecule information to GenBank, MEDLINE, and the other Entrez databases, and will serve as a public portal for MLSCN screening results and chemistry data. New algorithms and tools for computational chemistry, virtual screening, and other research aspects of cheminformatics, will be funded via new grants (11).

The ultimate goal of the MLI is to develop small-molecule modulators of thousands of cellular targets. To succeed, the MLI will be developing technology in four critical areas (11):

- (i) **Chemical diversity:** production of pilot-scale compound libraries in novel areas of chemical space (16), and methods development for natural product isolation, characterization, and chemistry. Compounds will be placed into the Small Molecule Repository and screened by the MLSCN centers.
- (ii) **Assay diversity:** development of innovative high-throughput assays for novel proteins, cellular phenotypes, biological functions, and disease mechanisms.
- (iii) **Instrumentation:** development of new technologies to allow HTS of novel assay formats and to increase throughput and accuracy of current screening technologies (e.g., methods for highly parallel noncompetitive detection of target binding, lab-on-chip technologies enabling complex multi-step assays).
- (iv) **Predictive ADME/Toxicology:** development of data sets and analysis algorithms for improved prediction of ADME (absorption, distribution, metabolism, and excretion) and toxicity properties of small molecules.

Although the MLI will utilize tools and technologies found in biopharmaceutical companies, the MLI has quite distinct

goals and deliverables from those of the private sector. For example, the need to develop compounds for human use as quickly as possible drives biopharmaceutical companies to screen only "druglike" compounds having potential for intellectual property novelty, on assays for a relatively limited group of targets (e.g., GPCRs, kinases, nuclear hormone receptors), and to not disclose the results of screens, keeping them as trade secrets or intellectual property. In contrast, the types of compounds synthesized and screened by the MLI will be broader, including metabolic intermediates, a range of natural products, and agents with potential *in vivo* toxicity. Types of assays will be broader also, including, for example, protein-protein interactions, splicing events, and diverse cellular and even organismal phenotypes. Via PubChem and the compound repository, all data and chemical probes will be available to the entire research community without encumbrance. The MLI will thereby enable study of the majority of biological and chemical space that is currently unexplored.

The small-molecule research tools produced by the MLI should accelerate validation of new drug targets (17) and thereby enable new drug development, but clinically useful compounds cannot be expected from the MLI itself. Drug development is a complex, time-consuming, and expensive process (18, 19), only the first steps of which will be performed by the MLI (see the figure on page 1138). The "probe compounds" produced by the MLI will have potency and aqueous solubility adequate for *in vitro* applications, but chemical modifications will generally be needed to confer the selectivity, pharmacokinetic, and metabolic properties required for *in vivo* use. From the probe stage, 10 to 12 chemist-years are commonly required to develop a "lead compound" with minimal pharmaceutical properties, and an additional 20 to 30 chemist-years to produce a "clinical candidate" compound appropriate for testing in humans. During this time, >3000 different compounds based on the initial probes are typically synthesized and tested. Even after this investment in chemical optimization, >90% of clinical candidates fail in human testing (19).

Examination of the cost and expected success rate at each stage of drug development demonstrates that the assay development, HTS, and hit-to-probe chemistry steps to be performed by the MLI are inexpensive and straightforward compared with the later stages of drug development (see the figure on page 1138). This should allow the MLI to produce probes for a broad range of targets with its relatively limited budget of ~\$100 million per year (com-

pared with >\$30 billion per year spent by pharmaceutical and biotechnology companies on drug development). We estimate that the probes produced by the MLI will entail only 2% of the cost, and 5% of the time, required to develop a novel drug (19). In exceptional circumstances, the NIH itself might attempt the subsequent steps of drug development, particularly if a small-molecule probe shows special potential for development into a drug for an orphan indication or a disease occurring primarily in the developing world, where there is unlikely to be commercial interest. In this case, NIH Roadmap mechanisms such as the Translational Research Core Services Program and the Regional Translational Research Centers (20) could be used. But for the most part, we expect that chemists in the public and private sectors will use the MLI probes as proof-of-concept compounds and as starting points to produce a variety of chemical analogs with improved properties. For this reason, intellectual property claims on the MLI probes will be strongly discouraged, as such claims would prevent the widest use of these tools and is contrary to the "community resource" nature of this initiative (21).

The MLI is a bold initiative to catalyze science in the genome era. By providing a new path for discovery, this program aims to accelerate science and its translation into benefits for the health of the public.

## References and Notes

1. E. Zerhouni, *Science* **302**, 63 (2003).
2. <http://nihroadmap.nih.gov/molecularlibraries/index.asp>.
3. J. Kaiser, *Science* **304**, 1728 (2004).
4. D. Janssen, *Bio-IT World* **3**, 1 (2004).
5. F. S. Collins, *Nature Rev. Drug Discov.* **3**, 640 (2004).
6. J. Couzin, *Science* **302**, 218 (2003).
7. C. P. Austin, *Curr. Opin. Chem. Biol.* **7**, 511 (2003).
8. International Human Genome Sequencing Consortium, *Nature* **431**, 931 (2004).
9. J. Drews, *Science* **287**, 1960 (2000).
10. A. L. Hopkins, C. R. Groom, *Nature Rev. Drug Discov.* **1**, 727 (2002).
11. <http://nihroadmap.nih.gov/molecularlibraries/grants.asp>.
12. <http://grants1.nih.gov/grants/guide/rfa-files/RFA-RM-04-017.html>.
13. [www.genome.gov/12512295](http://www.genome.gov/12512295).
14. <http://pubs.acs.org/cen/news/8235/8235notw5.html>.
15. <http://pubchem.ncbi.nlm.nih.gov/>.
16. N. P. Savchuk, K. V. Balakin, S. E. Tkachenko, *Curr. Opin. Chem. Biol.* **8**, 412-417 (2004).
17. M. A. Lindsay, *Nature Rev. Drug Discov.* **2**, 831-834 (2003).
18. K. H. Bleicher, H. J. Bohm, K. Muller, A. I. Alanine, *Nature Rev. Drug Discov.* **2**, 369-78 (2003).
19. J. A. DiMasi, R. W. Hansen, H. G. Grabowski, *J. Health Econ.* **22**, 151 (2003).
20. <http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp>.
21. <http://grants1.nih.gov/grants/guide/notice-files/NOT-RM-04-014.html>.
22. The Molecular Libraries Roadmap is a trans-NIH initiative conceptualized and implemented with the participation of all Institutes and Centers; see <http://nihroadmap.nih.gov/molecularlibraries/members.asp> for the members of the Molecular Libraries Project Team. The authors are leaders of the initiative.