

The academic pursuit of screening

High-throughput screening has become increasingly important in academic research over the last decade. The diversity of chemical and biological space being probed by academic chemical screening, coupled with the public reporting of results, has created an important new resource of data for chemical biologists.

High-throughput screening (HTS) has been the dominant paradigm for industrial drug discovery since at least the early 1990's. More recently, HTS has become increasingly important in academic research. Founded in 1997, the Institute for Chemistry and Cell Biology at Harvard University was one of the first academic screening centers (*J. Biomol. Screen.* **8**, 615–619, 2003). Since then, an expanding number of academic screening centers have been established worldwide, including the ten screening centers of the Molecular Libraries Screening Centers Network (MLSCN), financed by the US National Institutes of Health (NIH) (*Science* **306**, 1138–1139, 2004). Although the goals of screening centers vary from basic biological research to drug discovery, HTS in an academic setting creates new opportunities and presents different challenges from industrial drug discovery. Ten years into the experiment with academic screening, it is time to evaluate the results and look to expanding the impact of academic HTS.

HTS in industry is focused largely on assaying druggable targets for lead compounds with drug-like properties. In academic research, on the other hand, investigators may be interested in identifying small molecule modulators of biological targets that are not considered druggable or that have no connection to disease. With the broader range of biology under investigation, and without the requirement for optimal pharmacology, it is not necessary, and often not desirable, to limit screening libraries to drug-like molecules. This broad chemical biology purview pushes the boundaries for HTS assays (Review, p. 466) and changes the demands for the composition of chemical libraries (Commentary, p. 442). Thus, HTS in academic research is expanding our ability to probe chemical and biological space.

The increasing availability of screening facilities has significantly lowered the barrier to entry for academic researchers. However, the apparent ease of screening is somewhat deceptive. Rarely will a hit from an initial chemical screen provide a selective and potent modulator of a biological process, and often a significant medicinal chemistry effort is required to generate a small molecule that can be used to obtain biologically meaningful results. Although some screening centers provide users with medicinal chemistry support, more chemical resources will be essential for opening up screening to researchers from broad chemical biology backgrounds.

Within industry, the details of HTS screens are often incompletely reported, if they are reported at all. The use of HTS in an academic setting opens up the opportunity for more detailed reporting of screening results. However, there is not yet a community consensus for the information to include when publishing a chemical screen, and the assay details that are reported vary considerably from one paper to another. In this issue, in an effort toward increasing the transparency of screening results and to aid in comparing results between screens, Ingles, Shamu and Guy propose guidelines for reporting small molecule HTS data (Commentary, p. 438).

To make available the full data resulting from MLSCN screens, the NIH has created PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), a cheminformatics database of chemical structures and their biological activities. PubChem contains information about compounds, which are unique chemical structures; substances, which are chemical samples that contain more than a single compound or for which the compound structure is not known; and bioassays, which include results from all MLSCN-run screens, as well as assay data deposited from other sources. Chemical information also comes from many sources; for instance, *Nature Chemical Biology* contributes to PubChem through the deposition of the chemical compound information in its published articles (*Nat. Chem. Biol.* **3**, 297, 2007). PubChem content can be searched by chemical names or chemical structures, and by chemical similarity or other chemical properties. Ideally, PubChem will provide researchers with a resource, comparable to those that exist in industry, in which compounds are annotated with information about their performance across many biological assays. Besides the advantage of more easily eliminating nonspecific hits, these annotations can open up new insights into the interplay between chemical structure and biological activity.

Despite the clear importance of the information that is currently contained in PubChem, there is significant room for increasing the value of the database to the scientific community. Much of the information is user deposited, and thus can and does contain errors. Ensuring that the chemical information in the database is correct and nonredundant is essential for guaranteeing the usefulness of the database. At the moment, it is not easy to quickly obtain all the information in the database for a particular compound. Developing a more user-friendly interface would significantly increase the number of database users. Additionally, PubChem is currently integrated with other databases of the National Center for Biotechnology Information. Further integration with other chemical and biological resources would increase the value of the database's chemical information (Commentary, p. 447). To achieve the level of curation and integration seen in other chemical databases, a financial commitment for additional funding and staffing resources at PubChem will be critical.

From genome sequencing to DNA microarrays, biological researchers have embraced experiments that produce large data sets and have championed databases to make the results freely available to the scientific community. With the increased use of HTS in academics, the chemical community has taken an important step in the same direction. However, more effort will be needed to increase the availability of HTS methods and data. With a continued commitment to public funding of screening centers, cheminformatic databases, and related resources, along with forums for community feedback on HTS-related initiatives, large-scale chemistry will only grow in importance. ■