

## innovations

# Michigan HTS Screening Center: Designing a Compound Library for Maximum Diversity

Mark Johnson, Pannanugget Consulting  
Gordon Bundy, Independent Chemistry Consultant  
Darryl Chapman and Robert Kilkuskie,  
Michigan High Throughput Screening Center

The Michigan High Throughput Screening Center (MHTSC) is a not-for-profit, contract research laboratory, developing assays and running screens for academic and industrial clients. One important aspect of the Center is our compound library, containing roughly 100,000 compounds from various commercial providers.

We needed a diverse, druglike library to create an efficient and economic set of compounds. With a maximally diverse library, we can identify new ligands for a wide variety of targets. To this end, we worked with a team of medicinal chemists, biochemists and statisticians to select the compounds. The selection process involved a filtering phase followed by a diversity-design phase. The filtering phase consisted of a number of critical considerations that eliminated non-druglike compounds. The diversity-design phase used novel algorithms to identify the most diverse library available from our vendors.

Our goal was to populate the MHTSC screening collection with approximately high-quality, diverse structures from commercial sources. The initial list of vendors was kept small (4-6) to minimize the logistical problems of ordering from multiple sources. Vendors were selected based on their having substantial numbers of diverse, high-quality compounds available and having an end-user-available electronic catalog. The vendors selected were: ChemBridge, ChemDiv, Maybridge, and Tripos.

### Ensuring Drug-Like Potential

After selecting the vendors, we requested that they remove from their inventory lists the kinds of compounds historically rejected for purchase by other HTS operations. Examples of these exclusion criteria include: (a) chemically reactive compounds, e.g., sulfonyl chlorides; (b) compounds considered to be toxic, e.g., poly-halo- and polynitro-aromatics; (c) chelators; (d) non-drug-like compounds, i.e., serious violators of Lipinski's guidelines (Lipinski et al., 1997); and (e) metabolically labile compounds, e.g., azo compounds. The vendors' initial lists included 1.4 million compounds. The exclusion criteria eliminated nearly 1 million compounds. A subset of the remaining 480K structures was examined visually by

medicinal chemists who determined that the percentage of undesirable structures was negligible.

### "Lock and Key" Approach

Following electronic removal of unacceptable compounds, the vendors' lists were combined and a suitable number of compounds were chosen for purchase based on a computerized assessment of their diversity and coverage of critical functionalities. Several studies provide detailed examples of popular diversity assessment protocols (Harper et al., 2004; Lajiness and Shanmugasundaram, 2004). These protocols are designed to reduce the number of highly similar structures. However, they lack any direct means of assuring complete coverage of the major functionalities or chemotypes (ring systems and functional groups) represented in the source libraries.

To pick compounds for the MHTSC screening collection, we utilized a new diversity and coverage analysis protocol implemented as part of the software program known as MEQI (see [www.pannanugget.com](http://www.pannanugget.com)), which assures such coverage. Ring systems and functional groups are defined with respect to properties of the atoms and bonds of each structure. For ring systems, the bond property is simply that of being a ring bond. For functional groups, the bonds are not carbon-carbon single bonds or carbon-carbon aromatic bonds.

The coverage methods used in our protocol arise out of ideas in molecular equivalence analysis (Xu and Johnson, 2002). Our approach reflects a view of the "lock and key" analogy of receptor binding in which that aspect of a new lead that constitutes the critical part of the "key" in some drug-receptor interaction is one of these structurally positioned chemotypes. Under this view, a structure is a patterned "key chain" with as many keys as it has ring systems and formally-defined functional groups, and a compound library is a collection of patterned key chains. A popular key associated with, for example, a particular positioned hydroxyl group, may reside on many key chains. A rare key associated with a large and complex functional group or ring system may reside on only one key chain. The primary constraint of our coverage approach is to assure that every key found in any of our vendor libraries is also found on at least one compound of our MHTSC compound library.

### Defining Individual Chemotypes

We used ring-systems and formally defined functional groups as our basic chemotypic categories. But to define individual chemotypes, we have other considerations. We took a number of different aspects of the structural environments of these functionalities into account. We augmented the functionality by adding the adjacent alpha atoms. We "locally positioned" the functionality by associating with each of its atoms the involvement of side-chain, ring system or bridge bonds and by indicating for each of its atoms whether or not it connected the augmented functionality with the rest of the structure.

We "globally positioned" the functionality by associating a "layer" with each of its bonds where, using a 1,4-diphenyl-2-methylbenzene as an example, the bonds of the methyl side chain are all layer 1, those of the two outer rings are all layer 2, the two bridge bonds are both layer 3, and the bonds of the inner benzene ring are all layer 4. Any ring sys-

tem or functional group characterized in this way is considered a desired chemotype.

After removal of unacceptable structures, the number of unique ring systems and functional groups was counted for each source library (see the table, below). We used several different protocols to select 100K compounds from the source libraries. Initially, we considered the diversity of the source libraries, and selected compounds from the more diverse libraries first. To try to limit the bias of selecting certain libraries first, we randomly selected about 2000 compounds from each source library and generated a "starter library" by comparing these structures, accepting into the starter library, compounds with unique chemotypes. This resulted in a starter library of about 8000 compounds.

We added to the starter library by sequentially picking from each vendor library additional structures that contained at least one chemotype that did not occur on three or fewer structures in the then current list. This selection process added about 73,000 compounds to the library. We added more compounds by selecting any structure unless it possessed a chemotype with at least 5 atoms and no side-chain bonds that had already occurred on at least 10 currently accepted structures. This process added about 19,000 compounds, resulting in a total of approximately 101,000 compounds. These compounds were ordered from the vendors.

We also identified a 10,000-compound subset of our library (the Select set) for projects in which a smaller screening effort is desired. To ensure maximum diversity, this library was selected from the full library by accepting structures containing any functional-group chemotype with at least 5 atoms and no side-chain bonds, but only the first occurrence of that chemotype.

The number of different chemotypes in the vendor libraries and in the MHTSC libraries are shown in the table (below). The full MHTSC library has more chemotypes than any of the individual libraries. The Select Set contains more diversity than four of the five libraries. Only the Chembridge library of more than 300,000 compounds contains more chemotypes than the Select Set.

### Known Inhibitor Library

Another important resource of the MHTSC is the known inhibitor library. This is a small library of about 400 compounds that was selected from commercially available compounds (Sigma and EMD Biosciences) with known pharmacology. These are all off-patent drugs, with known targets and mechanisms of action. They include anti-arrhythmics anti-infectives, anti-tumor agents, cardiac agents, central nervous system agents, lipid-lowering agents, and a variety of steroids and non-steroidal anti-inflammatory agents. The known inhibitor library is included in every screen conducted at the MHTSC.

The methods described produced a novel compound library. The criteria used to remove "objectionable" compounds plus the application of MEQI to identify chemically diverse compounds generated a unique set of commercially available compounds that adds significantly to the value of the screens conducted at the MHTSC. Starting from approximately 1.4 million compounds contained in the vendors' libraries, 100,000 diverse compounds were selected. This compound library is a unique resource available to investigators through the MHTSC.

For more information, see <http://mhtsc.kvcc.edu>. \*

Library	Structures	Ring-system Chemotypes	Functional-group Chemotypes
Maybridge	44,681	9,374	10,946
CDL_40 <sup>1</sup>	40,000	11,863	10,694
CDL_20 <sup>1</sup>	20,000	4,864	4,809
Chembridge	324,856	18,484	19,015
Tripos	53,682	2,435	3,365
MHTSC	100,901	30,985	32,476
MHTSC Select Set ~10K	10,010	11,237	11,730

<sup>1</sup> ChemDiv Laboratories

### References

- Harper G., Pickett S.D., Green D.V. (2004) Design of a Compound Screening Collection for use in High Throughput Screening. *Comb.Chem.High Throughput Screening*, 7, 63-70.
- Huang, J., Zhu, H., Haggarty, S.J., Spring, D.R., Hwang, H., Jin, F., Snyder, M. and Lajiness, M.S. and Shanmugasundaram, V. (2004) Strategies for the Identification and Generation of Informative Compound Sets. In *Chemoinformatics, Concepts, Methods, and Tools for Drug Discovery*, J.Bajorath, ed.; Humana Press, Totowa, N.J.
- Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, R.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* 23, 3-25.
- Xu, Y.-J. and Johnson, M. (2002) Using Molecular Equivalence Numbers to Visually Explore Structural Features that Distinguish Chemical Libraries. *J. Chem. Inf. Comput. Sci.* 42, 912-926.