Description of UCLA MSSR Compound Library- information

Information provided by Dr. Robert Damoiseaux, July, 2015

Our compounds libraries are split into 4 segments: Pharmacological validation and repurposing libraries (biomol, LOPAC, Prestwick and Microsource spectrum, NIH clinical collection), targeted libraries (TAR), lead-like libraries (DL), diverse libraries (UCLA) and diverse/smart libraries (EAM and LS). All of our compounds are at least 90% pure, typically better. On average, we can resupply 90%-95% of the hit compounds as powder for follow up. With the exception of the diverse library (UCLA set) which was a pre-plated set, all of our sets are custom sets and are not likely to be found in another screening facility. We have applied extensive filtering against liabilities such as reactive groups, aggregators etc. (See the chapter on "Molecular Screening" in the Development of Therapeutic Agents Handbook by R.Damoiseaux, published by Wiley for some more details.)

The drug-likeness and usefulness of the pharmacological validation and repurposing libraries is well established. In this context, it is interesting to note that the NIH clinical collection serves to fill gaps in this area that were previously not addressed, hence we included this library in our deck.

The targeted library set if a set of 8k compounds which were subjected to high-throughput docking to kinases,proteases, ion channels and GPCR's and included in this set based on their predicted ability to bind to these high-value targets. The drug-likeness of this set is excellent, compounds from this set obey the rule of 5 - many of them even the rule of 4.

The lead-like library is a set of 20k compounds which obeys the more stringent rule of 4 rather than the more permissive rule of 5. This library was selected from a set of about 250,000 compounds to yield compounds which have more favorable properties for subsequent medchem optimization as the average molecule from this set is smaller, has fewer h-acceptor or donor sites and a better logP. It addresses the typical problem that during chemical optimization the compounds typically get heavier and less drug-like leading to more potential for ADME-problems in subsequent stages.

The diverse library is the DiverSet E from Chembridge - it's a well established 30k compound set which contains a lot of interesting structures and has generated many interesting hits in our hands. This set contains a vast structural diversity which has been selected from over 500,000 compounds. Another 50k custom diverse set from Life Chemicals was selected from over 600,000 compounds.

Our diverse smart libraries were selected from large compounds sets of 600k and 250k compounds. Using large computer clusters (250 nodes and more) at the CNSI, we filtered the sets for compounds which were drug like and did not have any other liabilities, fewer than 8 rotatable bond etc. and then broke the compounds into clusters of similar structures. This was a challenging process which took about 3 weeks computing time on the cluster. We then sampled these clusters for diversity and included the resulting compounds into our libraries. The resulting compound libraries have excellent properties (at least rule of 5 or better) and they are proprietary to the MSSR. The EAM set is 20k big and the LS set is 40k big.

http://www.mssr.ucla.edu/