

Using combined active and inactive data from HTS to reduce false positives and yield better enrichments

Al Maynard, Jon Sutter, Marvin Waldman, and Samuel Toba, Accelrys, Inc.

The wide use of high throughput screening (HTS) as biological assays has resulted in a large volume of valuable data for researchers. However, results from HTS only yield a very small (typically <<1%) number of active compounds and the rest of the inactive data are not utilized. These inactive compounds contain valuable information that is lost when not considered.

In Catalyst 4.10, we present a new scientific algorithm (HipHopRefine) that allows researchers to capture both the active and inactive compound data in their pharmacophore model generation by the use of excluded volumes (EV). Models generated using combined active and inactive compounds in the training set help reduce false positives and increase the enrichment rate when screening a virtual database.

We have selected a real-life example dataset of 50K compounds from HTS against dihydrofolate reductase (DHFR).^{*} The data contains two parallel screens, of which 32 compounds are identified as active with >75% inhibition in both screens. Subsequent IC₅₀ determination of the 32 actives found 11 competitive inhibitors with a K_i range of 3.8-11,000 nM.

To generate the pharmacophore model, we randomly split the data into a training set and a test set of 25K compounds each. We used the six most active compounds from the training set to build the initial model (EV=0), followed by the subsequent refinement using inactive compounds that map to the pharmacophore. The refined pharmacophore model contains 25 excluded volumes (EV=25). The models are then applied to virtually screen the 25K test set database.

	Number of Actives	Total Retrieved	% Yield	Enrichment	% Actives
Database	5	25,000	0.02%	1	100%
HipHop (EV=0)	5	1533	0.33%	16	100%
HipHopRefine (EV=25)	5	283	1.8%	88	100%

The new scientific algorithm HipHopRefine, which is contained in Catalyst 4.10, uses inactive data to generate excluded volumes. As seen in the table, HipHopRefine gives enrichment of 88x over the database and 5x over the existing pharmacophore model.

^{*}DHFR HTS information and data sets available from McMaster University: <http://hts.mcmaster.ca/HTSDataMiningCompetition.htm>

Accelrys Corporate Headquarters
10188 Telesis Court, Suite 100
San Diego, CA 92121, USA
Tel: +1 858 799 5000

Accelrys European Headquarters
334 Cambridge Science Park
Cambridge, CB4 0WN, UK
Tel: +44 1223 228500

Accelrys Asia Headquarters
Nishi-shimbashi TS Bldg 11F
Nishi-shimbashi 3-3-1, Minato-ku,
Tokyo, 105-0003, Japan
Tel: 81 3 3578 3860

Industry Sector

Pharmaceuticals

Organization

Accelrys

Key Products

Catalyst[®]

(HipHop and HipHopRefine)

