This paper appeared in Genetic Engineering & Biotechnology News, October 15, 2018.

## Abstract

When analysing the results from a high throughput screening (HTS) campaign the goal is to identify diverse hit series with high activity, structure-activity relationships (SAR) that indicate the opportunity for further optimisation and good ‘lead like’ properties. The common practise is to apply filters to these large datasets, for example an activity threshold or simple properties such as molecular weight, logP, numbers of hydrogen bond donors and acceptors or the presence of substructures that may indicate non-specific binding. However, this process draws artificially harsh distinctions between compounds, given the inherent variability in HTS data and the low correlation between simple properties and the ultimate *in vivo* disposition of a compound. This leads to selection of ‘false positives’, i.e. active compounds that are not good starting points for further optimisation and ‘false negatives’, i.e. potentially good compounds that have been inappropriately rejected. We will illustrate how a true multi-parameter approach enables appropriate weight to be given to these data to confidently identify high quality, potent hits while avoiding missed opportunities.

Mapping this information across the chemical diversity of the compounds explored in an HTS campaign, by clustering or visualisation of a ‘chemical space’, helps to find ‘hot spots’ representing high quality series of compounds for further investigation while also considering diverse chemistries to provide potential backup series. Finally, exploring the SAR within these series then helps to identify further opportunities for optimisation. We will show how this can all be achieved in a high visual and intuitive way, to move quickly and confidently from initial HTS hits to high quality lead series.

You can download this publication as a PDF .