

CHEMICAL BIOLOGY

Branching out into chemical space



To identify new biologically active compounds in chemical space — the vast array of all potential chemical molecules — there is a need for tools that recognize complex structural and bioactivity relationships and allow interactive navigation of large data sets. Now, two papers by Waldmann and colleagues published in *Nature Chemical Biology* describe the use of ‘Scaffold Hunter’, a computer-based tool that generates compound scaffolds and annotates them with bioactivity, to identify several new ligands that are associated with biochemical activity.

The programme reads data, such as that generated by a high-throughput screen, and then extracts chemically meaningful scaffolds and iteratively removes one carbocyclic or heterocyclic ring at a time from the larger ‘parent’ scaffolds to generate smaller ‘child’ scaffolds according to a set of chemistry-derived rules. Hierarchical arrangement of parents and children create ‘branches’

that are combined to form a ‘tree’. Importantly, scaffolds that do not exist in the dataset — known as virtual scaffolds — but occupy intermediate positions between scaffolds in the dataset are constructed *in silico* and included.

As a proof of concept that Scaffold Hunter can be applied to biological data sets to explore the gaps in chemical space, the authors analysed the PubChem pyruvate kinase screen of more than 50,000 molecules. Four virtual scaffolds were chosen on the basis of the potency of compounds in neighbouring scaffolds on the tree, from which nine previously undescribed pyruvate kinase inhibitors and activators were identified.

The authors next conducted a more detailed branch analysis of a dataset that was generated from the World of Molecular Bioactivity (WOMBAT) database (a collection of data from published literature on chemical structures and their associated biological activity). Unique molecular scaffolds (~46,000) were identified and then subjected to branch generation, and the branches were then annotated for biological activity. For the five main target classes — G protein-coupled receptors, nuclear receptors, kinases, proteases and ion channels — branches containing at least 5 scaffolds were identified, showing that progression from structurally complex to simpler scaffolds — that is, containing fewer ring structures — while retaining activity is feasible.

To investigate a prospective use of the scaffold–tree–branch approach,

the authors searched for branches with gaps in annotated activity, because such un-annotated scaffolds could be a good starting point for the synthesis of active compounds with expected activity. One example was identified for 5-lipoxygenase inhibitors. An un-annotated scaffold was used to design four corresponding compounds that had *in vitro* activity, and demonstrated that this branching approach can lead to molecules with improved ‘ligand efficiency’ (that is, the negative log of the half-maximal inhibitory concentration divided by the number of heavy atoms in the molecule — a metric that is useful in lead selection). Another example of a branch with un-annotated scaffolds was the oestrogen receptor- α (ER α) branch. Eight molecules were designed and synthesized that had varying affinity and selectivity for ER α . Data suggested that selectivity patterns might be identified by branching towards smaller scaffolds.

So, although not designed as a tool for the delineation of classical structure–activity relationships, Scaffold Hunter — which is freely available as an open source — can be used to identify regions in chemical space that could contain biologically active compounds and so guide the synthesis of new ligands.

Charlotte Harrison

ORIGINAL RESEARCH PAPERS Wetzels, S. *et al.* Interactive exploration of chemical space with Scaffold Hunter. *Nature Chem. Biol.* 28 June 2009 (doi:10.1038/nchembio.187) | Renner, S. *et al.* Bioactivity-guided mapping and navigation of chemical space. *Nature Chem. Biol.* 28 June 2009 (doi:10.1038/nchembio.188)