The scaffold diversity of exemplified medicinal chemistry space

Sarah R. Langdon¹,², Nathan Brown,¹ Julian Blagg²

¹ In Silico Medicinal Chemistry
² Medicinal Chemistry
Cancer Research UK Centre for Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, UK, SM2 5NG

Many large public and corporate databases of compounds exist to aid the drug discovery process. Such databases may be a limitation to hit finding if they only cover a small area of chemistry space leading to a lack of innovative drug design and crowding of intellectual property space. This work investigates the scaffold diversity of such databases. Scaffolds are defined by the Scaffold Tree¹ and Murcko frameworks.² We find that the majority of compounds in the databases we studied were based on only a handful of scaffolds. Thus a few scaffolds represent a large proportion of the database and the remaining scaffolds only represent a few or one compound. This demonstrates that the databases we are using in drug discovery have a limited scaffold diversity and identifies the need to explore methods for designing diverse and novel scaffolds that may lead to more successful hit discovery metrics across wider target classes and avoid known intellectual property space.

1. J. Chem. Inf. Model., 2007, 47, 47-58
2. J. Med. Chem. 1996, 39, 2887-2893