

Consensus rank orderings of molecular fingerprints illustrate the ‘most genuine’ similarities between marketed drugs and small endogenous human metabolites, but highlight exogenous natural products as the most important ‘natural’ drug transporter substrates.

^{1,2,3}Douglas B. Kell & ^{1,2}Steve O’Hagan

¹School of Chemistry, ²Manchester Institute of Biotechnology, ³SYNBIOCHEM, The University of Manchester, 131 Princess St, Manchester M1 7DN, UK dbk@manchester.ac.uk <http://dbkgroup.org/>

We compare several molecular fingerprint encodings for marketed, small molecule drugs, and assess how their rank order varies with the fingerprint in terms of the Tanimoto similarity to the most similar endogenous human metabolite as taken from Recon2. For the great majority of drugs, the rank order varies very greatly depending on the encoding used, and also somewhat when the Tanimoto similarity (TS) is replaced by the Tversky similarity [1-5]. However, for a subset of such drugs, amounting to some 10% of the set and a Tanimoto similarity of ~0.8 or greater, the similarity coefficient is relatively robust to the encoding used. This leads to a metric that, while arbitrary, suggests that a Tanimoto similarity of 0.75-0.8 or greater genuinely does imply a considerable structural similarity of two molecules in the drug-endogenite space. Although comparatively few (<10% of) marketed drugs are, in this sense, robustly similar to an endogenite, there is often at least one encoding with which they are genuinely similar (e.g. TS > 0.75). This is referred to as the Take Your Pick Improved Cheminformatic Analytical Likeness or TYPICAL encoding, and on this basis some 66% of drugs are within a TS of 0.75 to an endogenite [6].

We next explicitly recognise that natural evolution will have selected for the ability to transport dietary substances, including plant, animal and microbial ‘secondary’ metabolites, that are of benefit to the host. These should also be explored in terms of their closeness to marketed drugs. We thus compared the TS of marketed drugs with the contents of various databases of natural products. When this is done, we find that some 80% of marketed drugs are within a TS of 0.7 to a natural product, even using just the MACCS encoding. For patterned and TYPICAL encodings, 80% and 98% of drugs are within a TS of 0.8 to (an endogenite or) an exogenous natural product. This implies strongly that it is these exogeneous (dietary and medicinal) natural products that are more to be seen as the ‘natural’ substrates of drug transporters (as is recognised, for instance, for the solute carrier SLC22A4 and ergothioneine). This novel analysis [6] casts an entirely different light on the kinds of natural molecules that are to be seen as most like marketed drugs, and hence potential transporter substrates, and further suggests that a renewed exploitation of natural products as drug scaffolds would be amply rewarded.

- [1] O’Hagan, S., Swainston, N., Handl, J. & Kell, D. B. (2015). A ‘rule of 0.5’ for the metabolite-likeness of approved pharmaceutical drugs. *Metabolomics* **11**, 323-339.
- [2] O’Hagan, S. & Kell, D. B. (2015). Understanding the foundations of the structural similarities between marketed drugs and endogenous human metabolites. *Front Pharmacol* **6**, 105.
- [3] O’Hagan, S. & Kell, D. B. (2015). The apparent permeabilities of Caco-2 cells to marketed drugs: magnitude, and independence from both biophysical properties and endogenite similarities *PeerJ* **3**, e1405.
- [4] O’Hagan, S. & Kell, D. B. (2016). MetMaxStruct: a Tversky-similarity-based strategy for analysing the (sub)structural similarities of drugs and endogenous metabolites. *Front Pharmacol* **7**, 266.
- [5] O’Hagan, S. & Kell, D. B. (2017). Analysis of drug-endogenous human metabolite similarities in terms of their maximum common substructures. *J Cheminform*, in press.
- [6] O’Hagan, S. & Kell, D. B. (2017). Consensus rank orderings of molecular fingerprints illustrate the ‘most genuine’ similarities between marketed drugs and small endogenous human metabolites, but highlight exogenous natural products as the most important ‘natural’ drug transporter substrates. bioRxiv version. *bioRxiv*, 110437.