

Rapid Property Profiling and Similarity Calculations in Large Virtual Libraries

J. Barnard¹

G. Downs¹

¹*Digital Chemistry Ltd., Leeds, UK*

Extremely large virtual compound libraries, containing up to 10^{12} or even 10^{15} different molecules, may be used in drug discovery, and challenge even the fastest virtual high-throughput screening (VHTS) analyses. Even using Markush structure-based property-calculation techniques¹, production of property profiles for simple drug-like characteristics such as the Lipinski properties can be prohibitively slow. Appropriate sampling of library members can permit much faster analyses, and factors affecting the accuracy of property distribution profiles based on such sampling are discussed. Direct analysis of the Markush structure can also be used to calculate upper and lower bounds on the range of values for a particular property, without the need to enumerate individual values.

Exhaustive comparisons of individual library members against a target molecule, to identify the most similar, are also too time-consuming to be used on very large libraries. An approximate similarity search algorithm has been developed, which allows selection of molecules from a library that are highly similar to a specified target, though not guaranteed to be the most similar. The sets of molecules selected by this algorithm are compared to those identified by exhaustive similarity search.

1. Barnard, J. M.; Downs, G. M.; von Scholley-Pfab, A.; Brown, R.D. Use of Markush structure analysis techniques for descriptor generation and clustering of large combinatorial libraries. *J. Mol. Graph. Modelling*, **2000**, *18*, 452-463.