On- and Off-Target Prediction using 2D and 3D Molecular Similarity

Alexander Steudle[1], Emmanuel R. Yera[2], Ann E. Cleves[2], and Ajay N. Jain[2]

[1] Tripos – A Certara Company, Martin-Kollar-Straße 17, 81829 München

[2] University of California, San Francisco, Department of Bioengineering and Therapeutic Sciences, Helen Diller Family Comprehensive Cancer Center, San Francisco, California 94158, United States

Drug structures may be quantitatively compared based on 2D topological structural considerations and based on 3D characteristics directly related to binding. The present study establishes a framework in which 2D and 3D similarity computations can be directly compared and also combined. Given this framework, we studied the similarity patterns exhibited by 358 marketed small molecule drugs linked through partially shared molecular pharmacology and addressed two broad questions. The degree to which primary and secondary targets could be predicted was quantified using 2D similarity, 3D similarity, or a combination of both by making use of sets of drugs whose targets were known. The specific methods used were Surflex-Sim, and the 2D GSIM computation implemented within the Surflex platform.

The results were expected, but striking as to degree. The performance of the methods for predicting target annotations was 2D < 3D < 2D + 3D. Consistent with prior observations, 3D similarity did not yield a dramatic gain over 2D for primary targets due to the historical design bias problem: molecules designed to hit target X are often made specifically to look (in 2D) very much like molecules that are already marketed to modulate X. However, for off-targets, a dramatic improvement over 2D in the ability of 3D molecular similarity to identify relevant pharmacological effects was observed.

A broad observation is that a drug that shares high 2D and 3D structural similarity with another drug is likely to have indistinguishable pharmacological effects at the level of biochemically characterized modulation of protein targets. If, on the other hand, a drug shares little 2D similarity to existing drugs for the same cognate target but has high 3D similarity, there is greater likelihood to obtain a novel pharmacological effect. Specifically, drug pairs with high 3D and high 2D similarity showed identical biochemical targets four times more frequently than did pairs with high 3D similarity but low 2D similarity.

The work reported here introduces a new methodological approach for data fusion, demonstrated with 2D and 3D molecular similarity. Given other recent reports of methods for data fusion and off-target prediction, the differentiating features of what can be concluded based upon 2D and 3D molecular similarity is important to understand.