PURSUING LEADLIKENESS IN PHARMACEUTICAL RESEARCH

Tudor I. Oprea

Division of Biocomputing, University of New Mexico School of Medicine,
MSC08 4560, 1 University of New Mexico, Albuquerque NM 87131-0001, U.S.A.

Exhaustive enumeration of unique compounds with molecular weight (MW) <= 300 a.m.u., having only C, H and (in the future) O, N, S, P, and X is an on-going project between UNM and Daylight [1]. Based on Daylights GENSMI [1], all unique canonical non-isomeric SMILES are currently generated and stored in a database that can effectively become a resource for mining the entire virtual space of small molecules. The complexity of this problem at a computational and logistic level (over 10^{10} molecules are anticipated) is justified by recent trends in the pharmaceutical industry to move toward a fragment-based drug discovery approach [2], rooted in the concept of leadlikeness [3]. Existing chemicals significantly under-sample chemical space at MW > 300 [3]. The degree of overlap between the current enumeration effort and WDI, the World Drug Index is discussed based on descriptors related to branching, cyclization and molecular complexity [4]. Over 44 million unique SMILES meet the WDI criteria. We anticipate that these structures can become the basis for exploring novel chemistry spaces.

Our undestanding of the quality of leads rests on mining known biological actives. Such a source is the WOMBAT 2005.1 database [5], which contains over 104,000 unique chemicals and 230,000 biological activities. A derivative database related to clinical pharmacokinetics is the WOMBAT-PK (WB-PK) database [6]. WB-PK 2005.1 contains 656 drugs with multiple human ADME/Tox endpoints: > 600 oral bioavailability and half-life data, > 500 plasma protein binding and volume of distribution (steady state) values, > 400 total clearance, non-renal clearance and maximum recommended therapeutic daily dose values, etc. Matching clinical data with calculated properties, one can gain better insights for lead discovery. In particular, the relationship between the Maximum Recommended Therapeutic (daily) Dose, MRTD, and the partition coefficient (clogP and LogD74) will be discussed. Selection criteria that rely on ChemGPS [7], a principal components analysis-based model for, e.g., PK prediction [8], will also be highlighted.

[1] Kappler MA, Allu TK, Oprea TI. Exhaustive Enumeration of Small Molecules. 1. Simple Graphs. J Chem Inf Model 2005; *submitted*. See presentation by Mick Kappler for details on GENSMI: http://www.daylight.com/meetings/mug04/Kappler/GenSmi.html

[2] Fattori D. Molecular recognition: the fragment approach in lead generation. Drug Discov Today 2004; 9:229-238.

[3]Hann MM, Oprea TI. Pursuing the leadlikeness concept in pharmaceutical research. Curr Opin Chem Biol 2004; 8:255-263.

[4] Allu TK, Oprea TI. Rapid Evaluation of Synthetic and Molecular Complexity for In Silico Chemistry. J. Chem. Inf. Model. 2005, *submitted.*

[5] Olah M, et al. *WOMBAT: World of Molecular Bioactivity*, in **Chemoinformatics in Drug Discovery**. Oprea TI (Ed), Wiley-VCH, New York, 2005, pp. 223-239.

[6] WOMBAT and WB-PK are available from http://www.sunsetmolecular.com

[7] Oprea TI, Gottfries JG. Chemography: The art of chemical space navigation. J Comb Chem 2001; 3:157-166

[8] Oprea TI, Zamora I, Ungell AL. Pharmacokinetically based mapping device for chemical space navigation. J Comb Chem 2001; 4:258-266