How does Cheminformatics Contribute to Lead and Drug Discovery

Tudor I. Oprea

University of New Mexico School of Medicine and Sunset Molecular Discovery LLC

This will be a 3-parts talk. Issues in cheminformatics will be discussed within that context after each part.

1. Our undestanding of the quality of leads for drug discovery rests on mining the known bioactivity and medchem spaces. Such sources are the WOMBAT database [http://www.sunsetmolecular.com], which contains over 136,000 unique chemicals and 307,000 biological activities. A derivative database related to clinical pharmacokinetics is WOMBAT-PK (WB-PK). WOMBAT-PK 2006 contains 935 drugs with multiple human ADME/Tox endpoints: > 750 oral bioavailability and half-life data, > 700 plasma protein binding and volume of distribution (steady state) values, > 500 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 coefficients (clogP and LogD74) will be discussed. These two parameters appear to provide a physical basis to the Lipinski "Rule of five". *Note: Scott Boyer (AstraZeneca) and Igor Tetko (GSF) contributed to this work*

2. The University of New Mexico is one of 10 national centers for biomolecular screening, funded through the NIH Roadmap [http://nihroadmap.nih.gov/molecularlibraries/fundedresearch.asp]. The complementarity between virtual and biomolecular screening will be illustrated on two G-protein coupled receptor (GPCR) targets: the formyl peptide receptor, FPR [http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=362] and the estrogen-binding GPR30 [http://www.nature.com/nchembio/journal/v2/n4/abs/nchembio775.html]. The ability of ligand-based virtual screening to support the rapid identification of chemical probes will be discussed, in the context of single and multiple target profiling. N*ote: Cristian Bologa, Chetana Revankar, Susan Young, Bruce Edwards, Eric Prossnitz and Larry Sklar (UNM), Jeff Arterburn (NMSU) and Alex Kiselyov, Matt Parker, Sergey Tkachenko, Konstantin Balakin and Nikolay Savchuk (ChemDiv) contributed to this work*

3. The industrial sector and the FDA have accumulated significant data related to chemical structures in various stages of biological and clinical investigation. Our ability to mine that data is limited by the distinct possibility of reverse-engineering of chemical structures. The "Safe Exchange of Chemical Information" initiative and the ChemMask challenge [http://biocomp.health.unm.edu/chemmask.shtml] summarize recent efforts from a number of investigators in this area. Our work towards designing "confused" chemical descriptors [http://biocomp.health.unm.edu/people/tudor/papers/Bologa_DescriptorCollision_JCAMD2005.pdf] and the use of surrogate data [http://biocomp.health.unm.edu/people/tudor/papers/Tetko_SurrogateData_JCAMD2005.pdf] will briefly be discussed. *Note: Cristian Bologa, Marius Olah and Tharun Allu (UNM), Mick Kappler (Daylight), and Igor Tetko (GSF) and Ruben Abagyan (Scripps) contributed to this work*