Separating Drugs from Nondrugs

Nadine Schneider, Christine Jäckels, Claudia Andres and Michael Hutter*

Center for Bioinformatics, Building C7.1
Saarland University, D-66041 Saarbrücken, Germany

Gradual in silico filtering using successively applied decision trees

Virtual screening of large compound libraries requires an efficient strategy, i.e.
- suitable partitioning algorithms
- descriptors that are adapted in terms of computational demand

Here, we use several successively applied decision trees (A, B, and C) and two kinds of descriptors:

simple	**complex**
molecular weight	substructure patterns
logP	accounting for known
molar refractivity	toxic issues expressed as
drug-likeness indices [3,4,5]	SMARTS strings [1,2]
...	
rapidly computable	**time consuming**

Tree A: all descriptors
Tree B: only rapidly computable descriptors
Tree C: remaining descriptors not used in tree B

Results

Up to 90% of all nondrugs can be filtered out by applying two successive decision trees. Hutter's drug-likeliness index is superior to other indices [3,4] accounting for drugs.
Decision trees can derive margins for partitioning where no obvious separation is visible.

Guidelines for potential drugs:
Molecular weight above 220, molar refractivity of 40 or higher, one or more ring systems, at least two functional groups, Hutter's drug-likeliness index above zero [5].

The compounds used as data set will be made available in the upcoming CoEPrA (Comparative Evaluation of Prediction Algorithms) modeling competition. [6]

References

[1] M. Hann et al., *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 897-902.
[2] D. Flower, *J. Mol. Graph. Model.*, **1998**, *16*, 239-253.
[3] A. Ghose, V. Viswanadhan, J. Wendoloski, *J. Comb. Chem.*, **1999**, *1*, 55-68.
[4] T. Oprea, *J. Comput.-Aided Mol. Des.*, **2000**, *14*, 251-264.
[5] M. Hutter, *J. Chem. Inf. Model.*, **2007**, *47*, 186-104.
[6] http://www.coepra.org/