DEVELOPMENT OF A QSAR MODEL AND SYNTHESIS OF A CNS-FOCUSED LIBRARY

Rafael Gozalbes[a], Boryeu Mao[b], Frédérique Barbosa[a], Eric Nicolaï[a], Dragos Horvath[a,c] and N. Froloff[a]

[a]Cerep, 128 rue Danton, 92500 Rueil-Malmaison, France
[b]Cerep, 15318 N.E 95th street Redmond, WA, 98052 USA
[c]Current address : Institut de Biologie, UMR8525 CNRS, 1, rue Calmette, 59800 Lille, France

We have designed a diverse CNS-focused library based on the structure/CNS-oriented activity relationships of marketed drugs, clinical candidates and reference compounds and derived on hand of the data from our BioPrint® database (an overall of 2500 compounds profiled in house in more than 170 in vitro assays: receptors, enzymes, ion channels, cellular functional tests and in vitro ADMET assays) [1, 2].

A total of 978 compounds from BioPrint were classified such as "CNS-active" (193 compounds) or "CNS-inactive" (785 compounds), and a Linear Discriminant Analysis (LDA) was performed on a randomly chosen training set of 665 compounds in order to generate a CNS-QSAR model. LDA is a pattern recognition method providing a classification model based on the combination of variables that best predicts the category or group to which a given compound belongs. Independent variables in this study were Cerep 3-D pharmacophoric descriptors (Fuzzy Bipolar Pharmacophoric Autocorrelograms). Compounds with positive and negative LDA values correspond to "predicted CNS-active" and "predicted CNS-inactive" respectively.

The resulting QSAR model was applied to the test set of 313 structures, and it was able to correctly classify 80% of both CNS-active and inactive compounds. Furthermore, when the model was applied to an external set of 545 structures from the Merck Index, the ratio of CNS-active compounds witnessed a significant enrichment among the predicted actives, since 79.8 % of them (170 from 213 predicted actives) were correctly classified.

We have used this model to select and synthesize a library of CNS-focused compounds. In order to further guarantee the ability of these compounds to pass the blood-brain barrier, several important physicochemical parameters have been predicted as well using our predictive QSAR models derived from BioPrint® data, such as LogD at pH=7.4 and Caco-2 apical-to-basolateral permeability. Empirical filters (Lipinski rule-of-5, PSA) were also applied. Finally a set of more than 2000 compounds with predicted LDA > 0 and optimal physicochemical parameters were selected for synthesis. Amongst the final set of compounds, several original chemotypes have been identified.

[1] Krejsa, C.M. *et al*: "Predicting ADME properties and side effects: the BioPrint approach". *Curr. Opin. Drug Discov. Devel.* **2003**, *6 (4)*, 470-480.
[2] Gozalbes, R. *et al*: "The BioPrint approach for the evaluation of ADME-T properties: application to the prediction of cytochrome P450 2D6 inhibition" (*proceedings of the LogP symposium*, **2004**, to be published).