Prediction of Biological Properties of Drug-like Molecules

Michael C. Hutter*, Claudia Andres and Michael Gepp

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

INTRODUCTION

During the process of commercial as well as academic research for new drugs, rapid discovery of prediction of **blood-brain barrier** permeability and binding affinity to the **hERG channel**. Only drugs targeting the central nervous system (CNS) should pass the barrier of the endothelial cells (Figure 1). A severe adverse effect of some drugs is the blocking of the delayed rectifier current potassium channel (hERG-channel) that causes fatal arrhythmia (Figure 2).

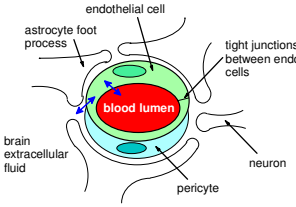
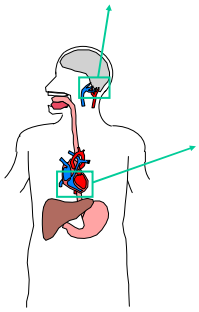


Figure 1: Blood-brain barrier
To reach the cerebrospinal fluid in the brain and the central nervous system, substances must pass the endothelial cells.

MATERIALS AND METHODS

Substance data sets
Experimental data for CNS permeabilities (CNS+/–), blood-brain distribution (logBB), binding affinity to the hERG channel, and QT-interval prolongation were collected from the literature and reported elsewhere. [1-3] The derived linear QSAR equation for the prediction of logBB is based on 90 substances.[1] In the decision tree approaches 224 compounds were used for prediction of CNS permeability and likewise 236 agents for prediction of QT-interval prolongation [2, 3]

Calculation of descriptors and molecular properties
Descriptors used were computed with a modified version of the semi-empirical program package VAMP and HYPERCHEM.[4, 5] Statistical analysis was carried out using the OpenStat2 program appling default settings.[6] The applied binary decision tree was developped in our lab.[2]



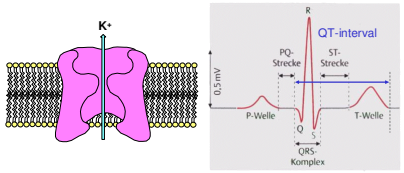
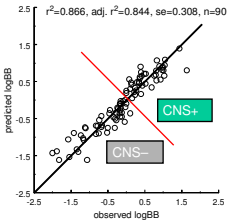


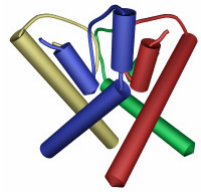
Figure 2: QT-interval prolongation as a result of blocking of the hERG potassium channel
Substances with high affinity to the delayed rectifier current potassium channel lead to changes in the electrocardiogramm that can be life threatening.

Plot of linear QSAR equation for the prediction of logBB (blood-brain distribution) [1]



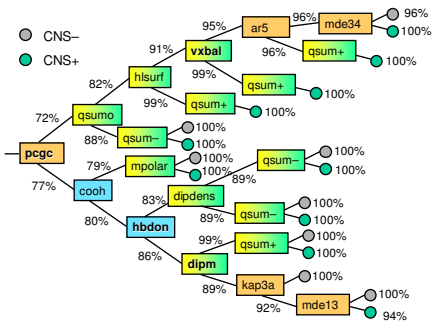
The red line indicates the boarder between substances that are readily availably to the central nervous system (CNS+) and those that do not cross the blood-brain barrier (CNS–).

Homology model of the hERG channel
The tetrameric form of the human $K_v11.1$



Docking of several compounds with high binding affinity gave rise to a **pharmacophoric pattern** that was used as descriptor.

Prediction of CNS permeability (CNS+/-) by a decision tree [2]

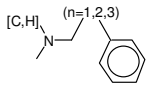


Descriptors are colour coded corresponding to their nature

- Size & shape
- electrostatic
- H-bonds

Descriptors also used in the linear QSAR equation for logBB are marked bold.
The level of accuracy for each branch and leaf is given in %.

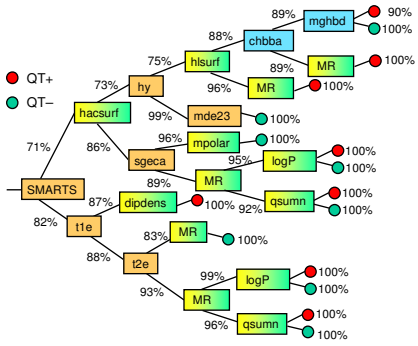
The pharmacophoric SMARTS string used as selection criteria for hERG binding



N(C)([H,C])C[!O]*~*~c

Using this SMARTS string we also identified a series of additional drugs in the PubChem database [9] that exhibit QT-interval prolonging activity: fentanyl, fluphenazine, melperone, clobutionol, and amodiaquine.

In silico filtering of QT-prolonging drugs by a decision tree [3]



REFERENCES

[1] M.C. Hutter *J. Comput.-Aided Mol. Des.* **2003**, *17*, 415.
[2] C. Andres & M. C. Hutter *submitted for publication.*
[3] M. Gepp & M. C. Hutter *unpublished results*.
[4] T. Clark et al., VAMP version 6.5, Oxford Molecular, Erlangen, **1997**.
[5] HYPERCHEM version 6.02, Hypercube Inc., Gainsville FL, **1999**.
[6] W.G. Miller, OpenStat2 version 6.1, West DeMoines, Iowa 50265.
[7] N.Guex, M. C. Peitsch, Electrophoresis, **1997**, *18*, 2714.
[8] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, Nucl. Acid Res., **2003**, *31*, 2281.
[9] http://pubchem.ncbi.nlm.nih.gov/

RESULTS AND CONCLUSIONS

Using a decision tree, the CNS permeability of compounds can be estimated with 94% accuracy. The corresponding strategy for the prediction of QT-interval prolonging drugs benefits from the presence of a pharmocophoric SMARTS string. The rate of false positives is below 3%.