ADMEWORKS Predictor Models

CYP3A4 Inhibitor (commercial model)
This model was developed using the Human P450 database from Professor Slobodan Rendic of Zagreb University. A training set of 370 3A4 inhibitors and 120 non-inhibitors were used. The model was validated on an external set of 884 molecules and achieved a sensitivity of 96.9%.

AMES-TA98 (commercial model)
Mutagenicity AMES-TA98 model was developed with the guided help of Japan's National Institute of Health Sciences. The model is based on the bacterial strain TA98. AMES-TA98 was created using Linear Learning Machine. The training set of 31 positive and 69 negative compounds gives the 100% correct classification rate.

AMES-TA100 (commercial model)
Mutagenicity AMES-TA100 model was developed with the guided help of Japan's National Institute of Health Sciences. This model is based on the bacterial strain TA100. AMES-TA100 was created using Linear Learning Machine. The training set of 42 positive and 79 negative compounds gives a 92% correct classification rate.

NTP Ames (commercial model)
Mutagenicity NTP Ames model was developed with the guided help of Japan's Nationals Toxicology Program. This model was conducted using the bacteria Salmonella. The training set of 142 positive and 98 negative compounds gives the 100% correct classification rate.

Carcinogenicity-FN (commercial model)
Carcinogenicity-FN model was developed using data from the National Toxicology Program (NTP) for long-term study of male rats. The model was developed using a training set of 111 positive and 167 negative compounds. The Carcinogenicity-FN model was created to increase correct classification of positive compounds and thus reducing false negative (FN) predictions.

Carcinogenicity-FP (commercial model)
Carcinogenicity-FP model was developed using data from the National Toxicology Program (NTP) for long-term study of male rats. The model was developed using a training set of 111 positive and 167 negative compounds. Carcinogenicity-FP model was created to increase correct classification of negative compounds and thus reduce false positive (FP) predictions.

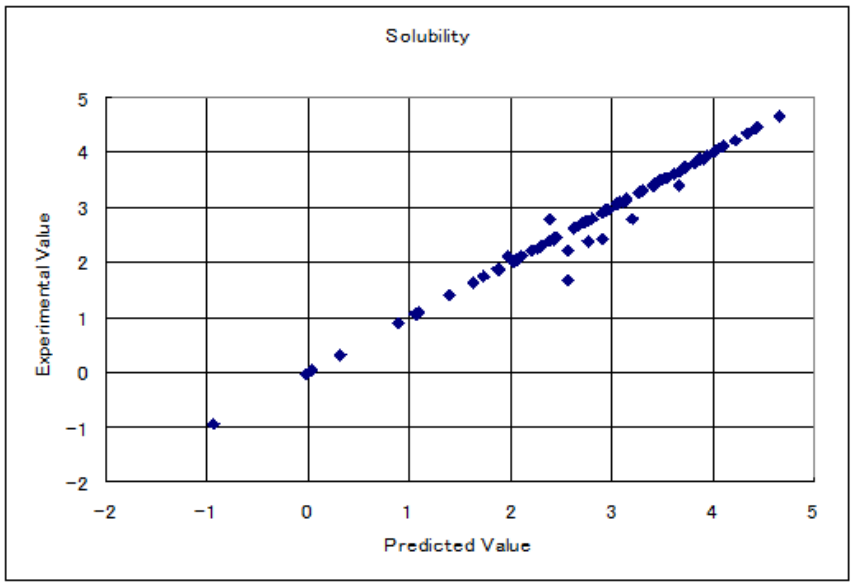
Skin Sensitization (commercial model)
The Skin Sensitization model is based on the in vivo mouse test "murine local lymph node assay (LLNA)", which is developed to determine the potential of pure chemicals or mixtures for inducing allergic contact dermatitis (ACD) in humans. The LLNA measures lymphocyte proliferation using incorporation of radioactive thymidine or iododeoxyuridine in the draining lymph nodes of mice topically exposed to the test article. The results are expressed as a ratio (SI = simulation index) of the mean number of disintegrations per minute (dpm) for treated mice compared to controls. A test substance is considered positive in this assay if the SI is greater than or equal to 3.0 (meaning a dose of the test article is capable of inducing a threefold or greater increase in the SI). It is considered negative otherwise. Cross-validation of the Skin Sensitization Model using the leave-1-out method confirmed an accuracy of 80% (98/122). External validation performed on an independent test set of 63 compounds (44+, 19-) not found in the training set successfully predicted the ACD potential of 57 compounds (concordance of 90%, specificity of 89%, and sensitivity of 91%).

Biodegradation (commercial model)
The Biodegradation Model predicts if a compound is "easily degradable" based on its structure. It was trained from a set of 413 compounds, of which 214 are classified as "Low" and 199 as "High" biodegradability. Compounds showing more than 60% mineralisation within 28 days are classified as "High", otherwise classified as "Low". All data were taken from the OECD Report on QSAR-models for biodegradation (1993). External validation on 147 test compounds (not included in the training set) showed a correct prediction for 120 compounds. The model was developed by Fraunhofer Institute for Molecular Biology and Applied Ecology (Germany).

hERG (commercial model)
Model predicts potassium channel protein inhibition and was developed using pIC50 values found in literature. A threshold of pIC50 greater than 6.0 was used as criteria for classifying positive compounds. The model was trained on a set of 73 compounds (45 negatives and 28 positives) using RFSBoost algorithm which attained a 100% correct classification. Validation on a separate test set (8 negatives and 9 positives) predicted 1 false negative and 3 false positive compounds.

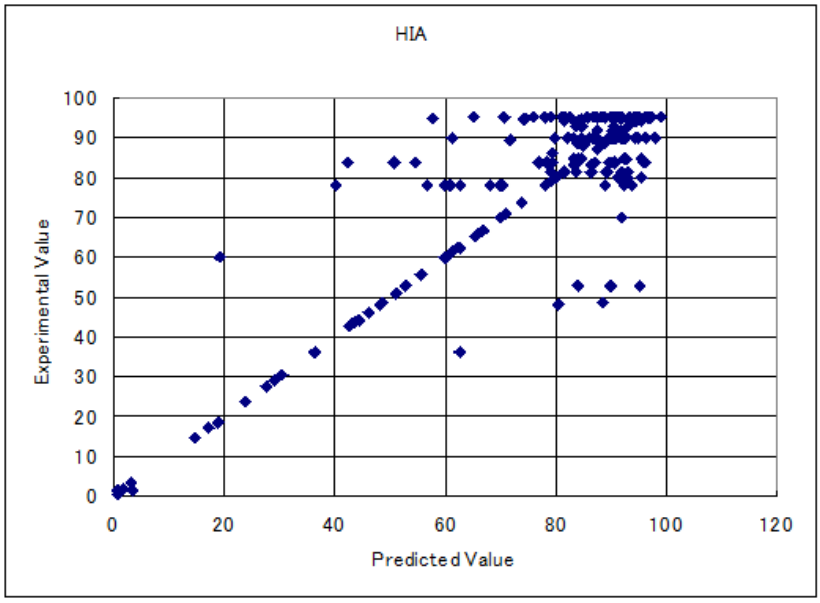
Chromosomal aberration (commercial model)
Model predicts whether a chromosomal abnormality occurs. The model was trained on a set of 395 compounds (195 positives and 200 negatives). Validation on a separate test set (61 negatives and 28 positives) predicted 10 false negative and 12 false positive compounds

Water solubility (commercial model)
This model predicts the logarithmic value of water solubility (logS). It was developed using the Physprop database from Syracuse Corp., extracting only compounds that have experimental water solubility (mg/L) values at 25 degrees Celsius. Converting all solubility values to logarithmic scale (base 10), 215 molecules were used in the training set giving the following statistical results: RSQ=0.96 and RMSE=0.38 (internal set) and RSQ=0.93 and RMSE=0.54 (external set).



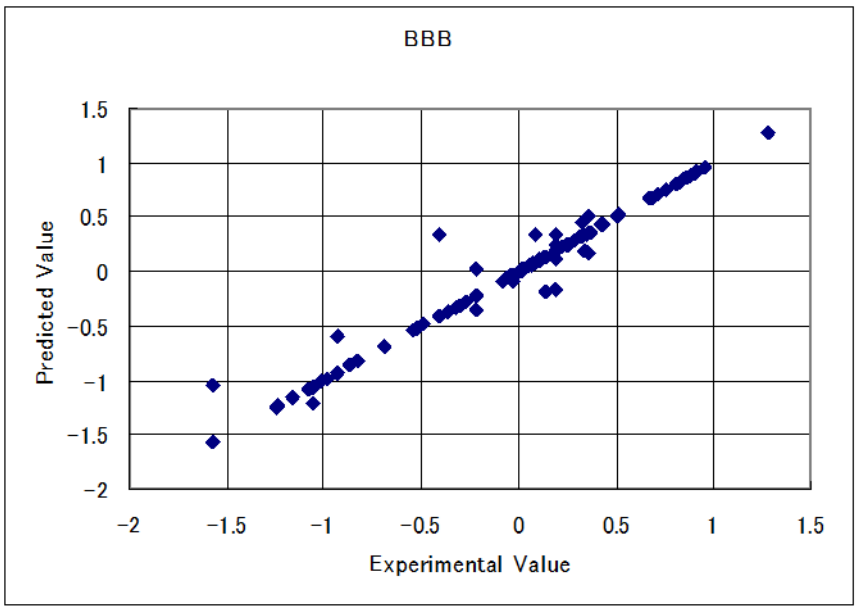
HIA (commercial model)
Human Intestinal Absorption (HIA) - the intestinal epithelium forms a permeability barrier for absorption of orally administered compounds such as food, drugs and toxicants. Model will allow for

rapid screening of absorption and mechanistic studies on transport and metabolism. 200 molecules were used in the training set giving the following statistical results: RSQ=0.70.



BBB (commercial model)

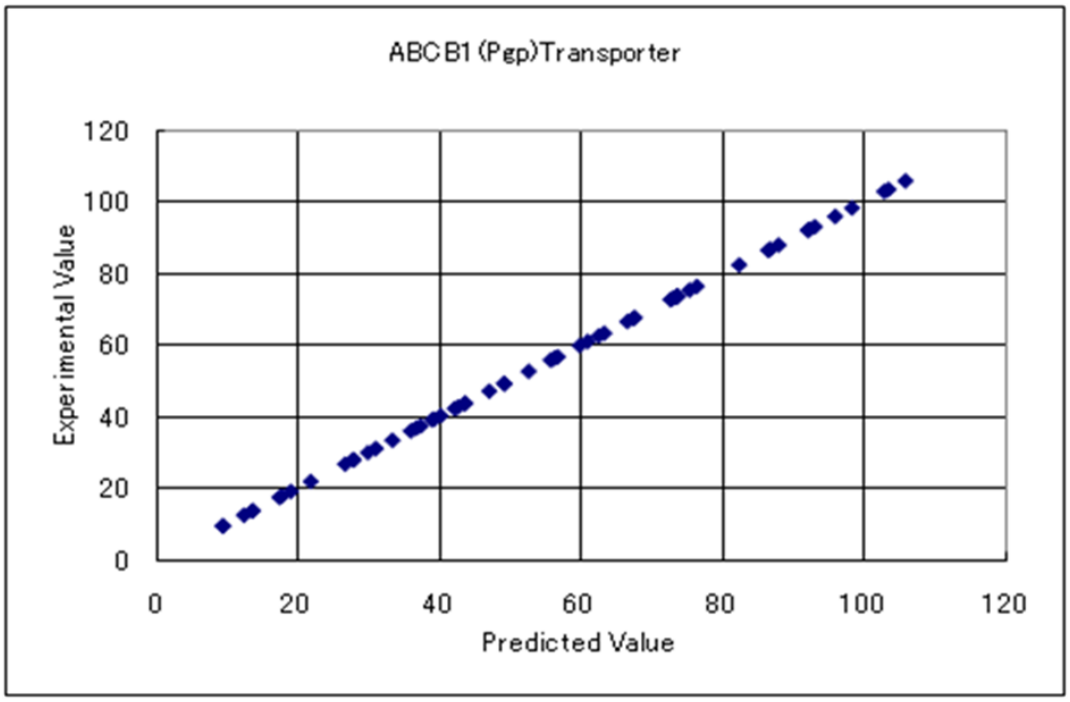
This MLR model predicts the Blood-Brain Barrier (BBB, expressed as the logarithmic value of the ratio of drug concentrations in brain and blood) - the barrier that excludes many molecules and substances from freely diffusing or being transported into the brain tissues from the blood stream. The higher the value of BBB, the better is the ability of the molecule to cross the barrier (higher concentration of compound in brain). 96 molecules (drugs and organic compounds) were used in the training set giving the following statistical results: R2 = 0.82, CV R2 = 0.75.



ABCB1(Pgp) Transporter Substrate (commercial model)

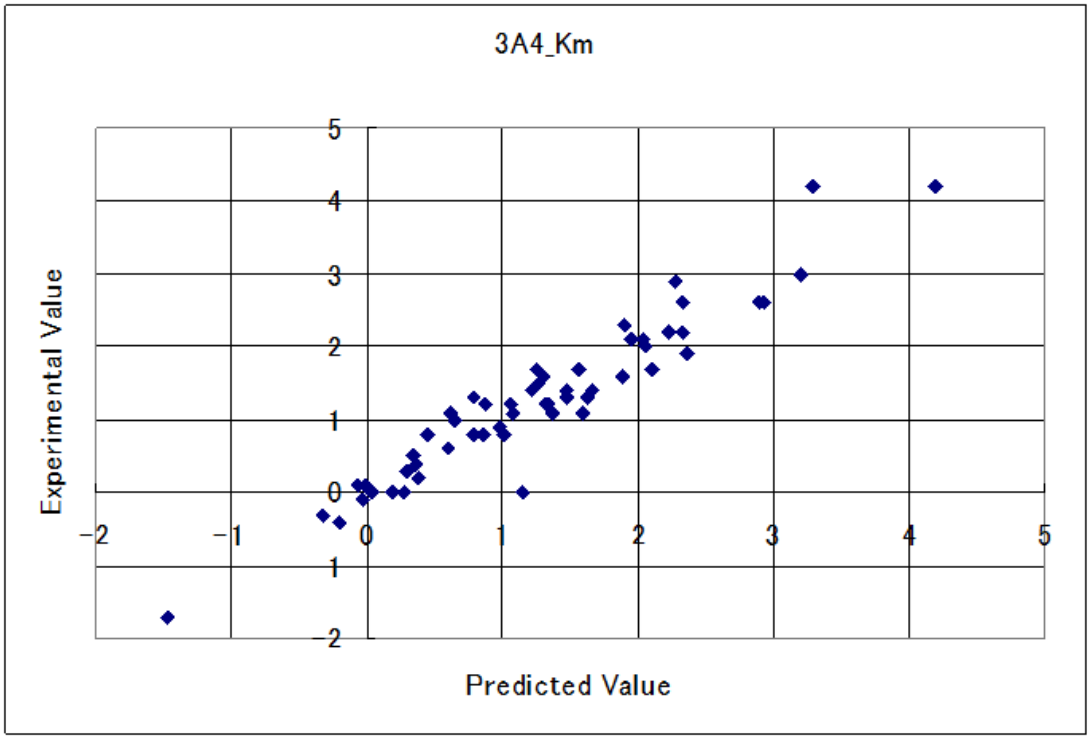
Model predicts potential substrates of Pgp (P-Glycoprotein). 60 molecules were used in the training set giving R2 = 0.87. Model was developed based on relative ATPase activity of the training molecules,

with relative ATPase activity for 10µM of Verapamil set as 100%. The higher the value, the greater the tendency to become a Pgp substrate.



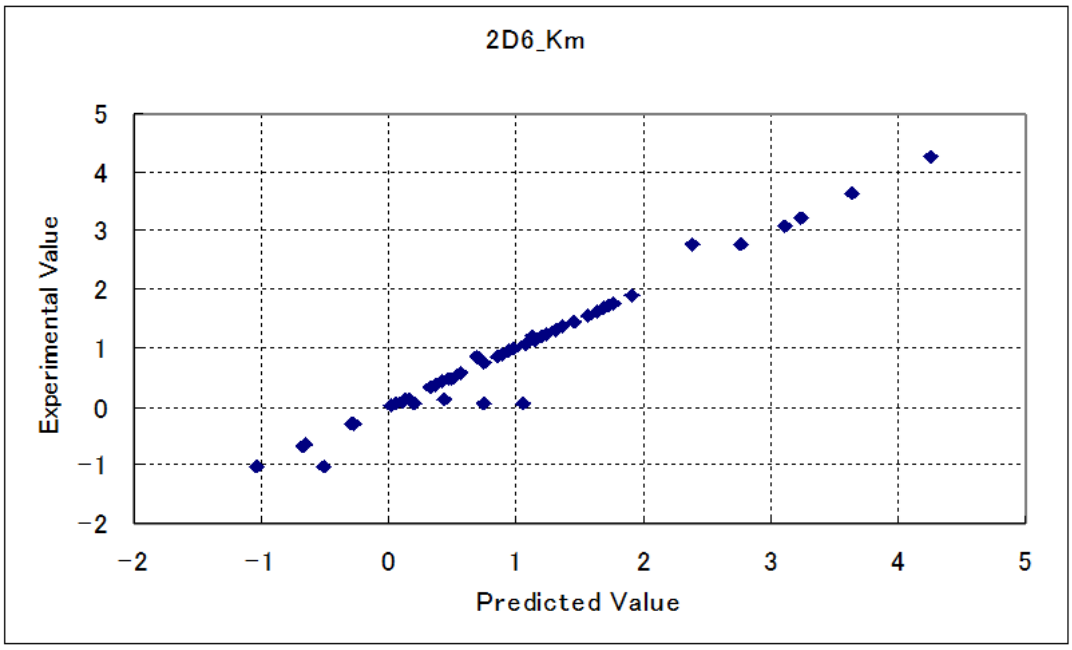
CYP3A4 Km (commercial model)

Model was developed using Km values measured on baculovirus-infected insect cells expressing human CYP3A4 (Supersomes). Prediction results are expressed in the logarithmic scale (from 2uM to 50uM). The model was trained on a set of 57 compounds (R2=0.92, MSE=0.14, LOO=0.83). Validation on a separate test set of 17 compounds shows R2=0.55, MSE=0.19.



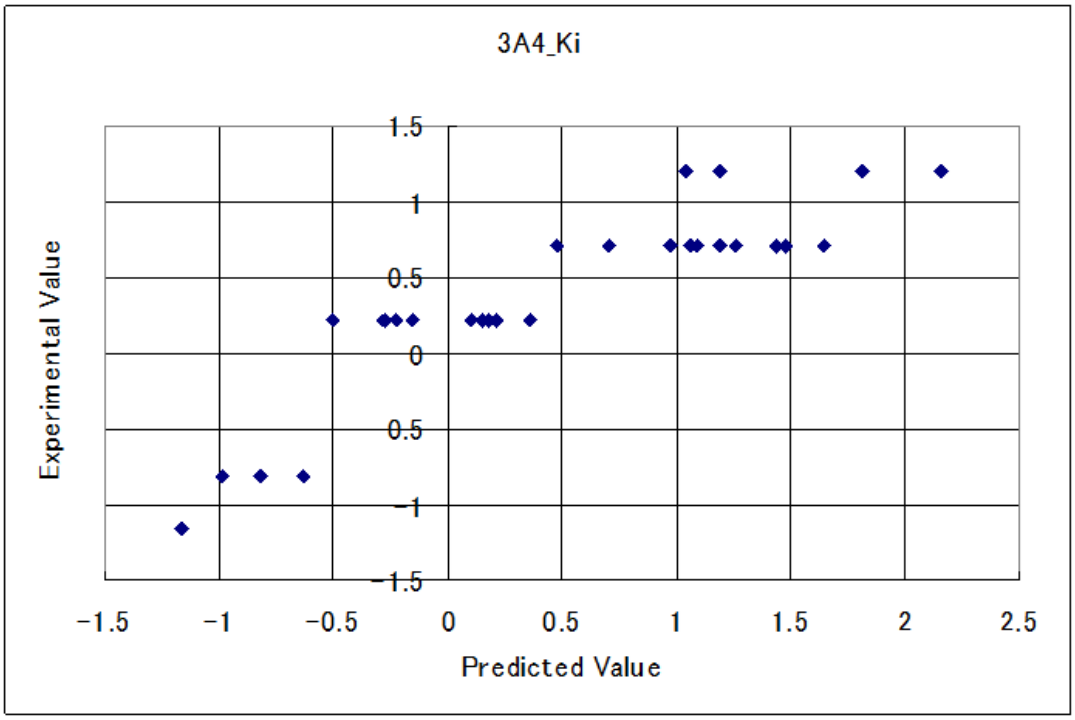
CYP2D6 Km (commercial model)

Model was developed using Km values measured on baculovirus-infected insect cells expressing human CYP2D6 (Supersomes). Prediction results are expressed in the logarithmic scale (from 2uM to 50uM). The model was trained on a set of 53 compounds and gives R2 = 0.93.

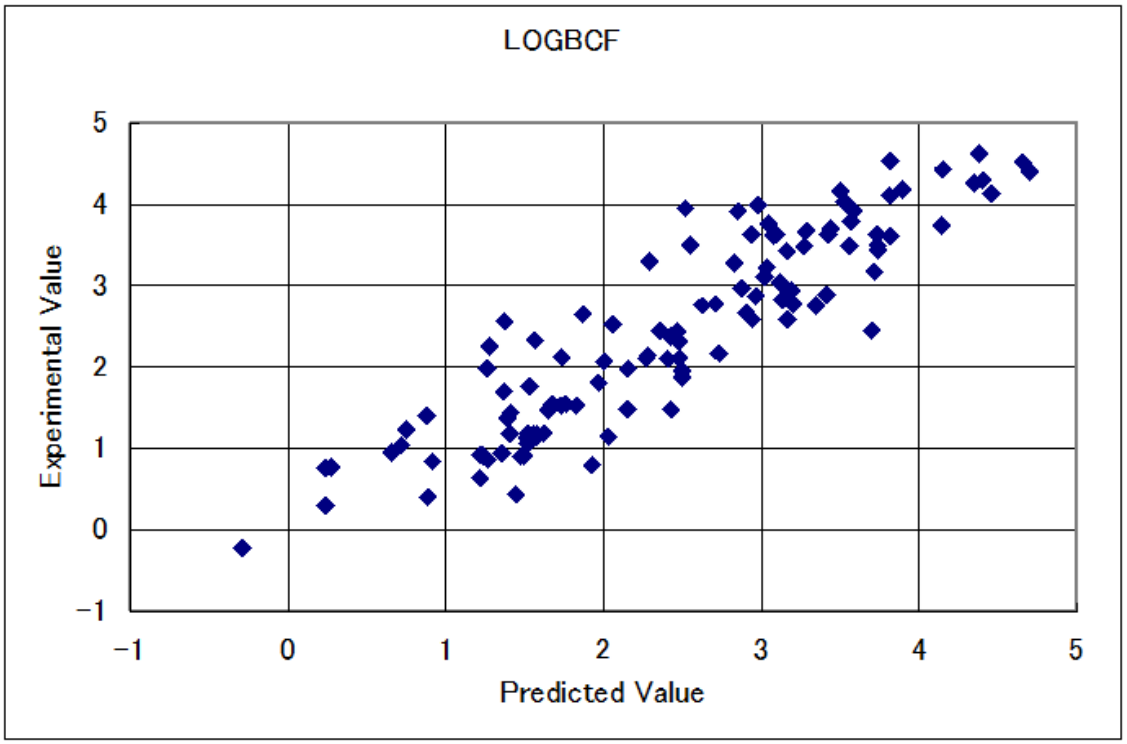


CYP3A4 Ki (commercial model)

Model was developed using Ki values measured on baculovirus-infected insect cells expressing human CYP3A4 (Supersomes). Prediction results are expressed in the logarithmic scale (from 1uM to 30uM). The model was trained on a set of 32 compounds and gives R2 = 0.83.



Bioconcentration (logBCF) (commercial model)
The LOGBCF Model predicts a compound's BCF (Bioconcentration Factor) for fish based on its structure. It was trained from 117 compounds with BCF values based on literature. The training set was restricted to compounds with logP less than 6. Cross-validation by Leave-1-Out method showed a correlation of R2=0.8 between the predicted and observed values. All predicted values are in the logarithmic scale. The model was developed by Fraunhofer Institute for Molecular Biology and Applied Ecology (Germany).



AndrewsBinding (free model)
AVERAGE (Average Energy Resulting from All Group Energies) calculates average bind energies as a sum of contributions of 10 common functional groups.
References: Andrews P.R., Craik D.J., and Martin J.L.; J. Med. Chem. 1984, 27, 1648-1657

Number of H-Bonds Donors (free model)
Number of H-Bond Donors, calculated using the LEADLIKENESS descriptor generator.

LEADLIKENESS (free model)
Physiochemical properties typical of good lead compounds. This model is based on Lipinski Rule of Five and can be used as a rule of thumb to indicate whether a molecule is likely to be bioactive. In general, an orally active drug has: 1. not more than 5 hydrogen bond donors (expressed as a sum of O-Hs and N-Hs groups) - HDO descriptor 2. not more than 10 hydrogen bond acceptors (expressed as a sum of Ns and Os) - HAC descriptor 3. a molecular weight under 500 - Mass descriptor 4. a LogP value under 5. In our model we also use the Number of Rings (Rings count descriptor), Number of Rotatable Bonds (Count of rotatable bond descriptor) and FQLogS descriptor for predicting the value of water solubility of organic compounds in water environment.
References: 1. Lipinski A., Lombardo F., Dominy B., Feeney P.; Adv. Drug. Deliv. Rev. 2001, 46, 3-26. 2. *Oprea T.; J. Comput. Aided Mol. Des., 2000, 14, 251-264*

MlogP (free model)
A model for predicting compound's logP value. The value is derived from compound's attribute-value. The octanol-water partition coefficient is calculated using the algorithm by Moriguchi et al.
References: Moriguchi I., Hirono S., Liu Q., Nakagome I., Matsushita Y.; Chemical & Pharmaceutical Bulletin, 1992, 40, 127-130

Number of H-Bond Acceptors (free model)
Number of H-Bond Acceptors, calculated using the LEADLIKENESS descriptor generator.

Rotatable bonds (free model)
The number of rotatable bonds (RTB) is formulated in the following equation: RTB=N(nt)+SUM(n(i)-4-RGB(i)-ShB(i)) where: N(nt) is the number of non-terminal freely rotatable bonds (but single bonds observed on groups like, e.g., sulfonamides (N-S) or esters (C-O),are excluded); n(i) is the number of single bonds in any non-aromatic ring i with 6 or more bonds; RGB(i) is the number of rigid bonds in ring i; ShB(i) is the number of bonds shared by ring i with any other ring. The number of rigid bonds (RGB) is defined as the difference between the total number of bonds and the total number of rotatable bonds (including terminal single bond).
References: Oprea T.; J. Comput.Aided Mol. Des., 2000, 14, 251-264

Rule of Five (free model)
Lipinski's Rule Of Five. A model for predicting properties and structural features that make molecules more or less drug-like. The value is derived from compound's attribute-value. Compounds evaluated as negative have poor absorption or permeation.
References: Lipinski CA, Lombardo F, Dominy BW and Feeney PJ.; Adv Drug Del. Rev, 1997, 23, 3-25

Rule of Five (CNS) (free model)
Lipinski's Rule Of Five. This model is dedicated for CNS (Central Nervous System) related drugs. It uses the standard Rule Of Five descriptors with special cut-off parameters optimized for CNS drugs. Compounds evaluated as negative have poor absorption or permeation.
References: Lipinski CA, Lombardo F, Dominy BW and Feeney PJ.; Adv Drug Del. Rev, 1997, 23, 3-25

TPSA (free model)
The method, termed topological PSA (TPSA), provides results which are practically identical with the 3D PSA (the correlation coefficient between 3D PSA and fragment-based TPSA for 34 810 molecules from the World Drug Index is 0.99), while the computation speed is 2-3 orders of magnitude faster. The new methodology may, therefore, be used for fast bioavailability screening of virtual libraries having millions of molecules.
References: Ertl E., Rohde B., and Selze P.; Cheminformatics, Novartis Pharma AG, WKL-490.4.35, CH-4002 Basel, Switzerland, 2000