ADMEWORKS

Science is life.

FUJITSU

ADMEWORKS ModelBuilder

ADMEWORKS ModelBuilder is a Windows application dedicated for building QSAR/QSPR models that can later be used for predicting various chemical and biological properties of compounds.

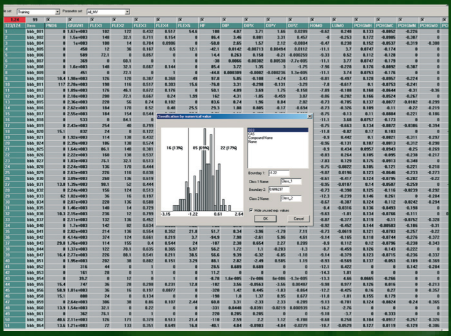
A set of data on molecular structures and their respective experimental values of the property of interest is a prerequisite for every model building. Two classes of models (Qualitative and Quantitative) can be built using various algorithms.

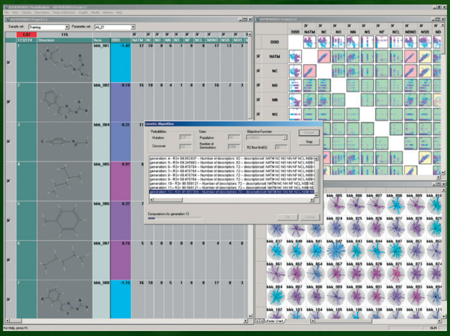
The models are based on values of physico-chemical, topological, geometrical, and electronic properties derived from the molecular structure – called descriptors. ModelBuilder includes a large number of topological, quantum, geometric and substructure-related descriptors for a large area of applications ranging from physical properties to biological activities.

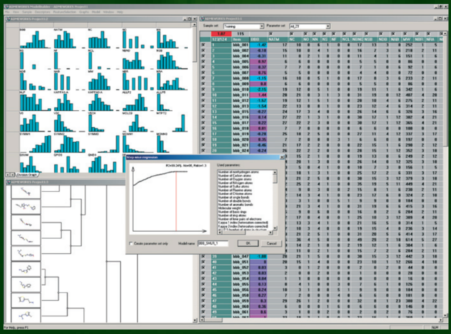
Models created in ModelBuilder are easily imported into ADMEWORKS Predictor (optional product) which is a high-speed virtual (*in silico*) screening system intended for simultaneous evaluation of the ADMET properties of compounds. Simultaneous evaluation of the pharmacological as well as the ADMET properties of compounds is useful in the discovery phase to produce balanced quality hits, and also in the lead optimization phase to lessen the occurrence of faulty leads.

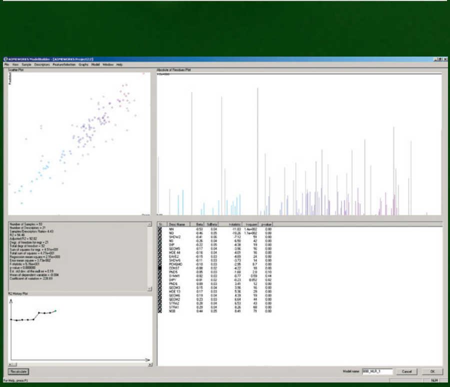
System Specifications

- Quantitative/qualitative structure-activity relation's analysis.
- Multiple statistical methods for generating predictive models: Perceptron, Iterative Least Squares Method, Multiple Linear Regression, Stepwise Regression, Leap-and-Bounds Regression, Genetic Algorithm, Fuzzy Adaptive Least Squares, k Nearest Neighbors, ADA Boost and Support Vector Machine.
- Customizable cross-validation of models.
- Multivariate/pattern recognition data analysis.
- ADMET data analysis.
- Over 400 predefined descriptors and an unlimited number of substructure-related descriptors.
- Interactive graphs that display property distribution, predicted vs. actual activity, property correlations, clustering of samples and properties and others.
- Interactive graphical feature and outlier selection tools.
- Automated statistical tools for feature and sample selection, including evolutionary algorithms
- SDF and CSV file import/export.









ADMEWORKS Predictor

ADMEWORKS Predictor is a highspeed virtual (*in silico*) screening system intended for simultaneous evaluation of the ADMET properties of compounds. It complements existing *in silico* technologies for evaluating pharmacological properties. Simultaneous evaluation of the pharmacological as well as the ADMET properties of compounds is useful in the discovery phase to produce balanced quality hits, and also in the lead optimization phase to lessen the occurrence of faulty leads. ADMEWORKS Predictor also makes it possible to prioritize while simultaneously evaluating these properties.

By freely ranking the properties according to their relative importance, ADMEWORKS Predictor allows for a more focused screening of compounds, stressing only the properties that are of highest interest.

ADMEWORKS Predictor seamlessly integrates with ADMEWORKS ModelBuilder (optional product), which allows creation of customized models to be used for the prediction of compound properties.

ADMEWORKS Predictor also provides support for legacy systems and models through a comprehensive interface, and is highly scalable as it allows integration with other third party computational tools.

Main Features

- Store molecular structures and properties in a central database.
- Manage and control via a web browser.
- Divide work into convenient worksheets.
- Handle many users in a corporate environment.
- Run all-at-once predictions via a web browser.
- View structures using a highly functional 3D structure viewer.
- Import/export MOL, SDF files.
- Filter and sort results.

www.fqs.pl/ADMEWORKS_ModelBuilder

Models available together with ADMEWORKS Predictor:

CYP3A4 Inhibitor

This model was developed using data for the Human CYP3A4. A training set of 370 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 Test

Three mutagenicity Ames tests based on bacteria Salmonella strains were developed with the guided help of Japan's National Institute of Health Sciences and Japan's Nationals Toxicology Program. Models were created using Linear Vector Machine and gave the 100% correct classification rate.

Carcinogenicity

Carcinogenicity 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.

Skin Sensitization

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. Cross-validation of the model using the leave-one-out method confirmed an accuracy of 80% (98/122). External validation performed on an independent test set of 63 compounds 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

The Biodegradation Model, developed by Fraunhofer Institute for Molecular Biology and Applied Ecology (Germany), predicts if a compound is "easily degradable". It was trained from a set of 413 compounds, of which 214 are classified as "Low" and 199 as "High" biodegradability. All data were taken from the OECD Report on QSAR-models for biodegradation. External validation on 147 test compounds showed a correct prediction for 120 compounds.

hERG

Model predicts potassium channel protein inhibition and was developed using pIC50 values found in literature. 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

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

This model predicts the logarithmic value of water solubility (logS). It was developed using the Physprop database from Syracuse Corp. 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

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

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. 96 molecules (drugs and organic compounds) were used in the training set giving the following statistical results: $R^2 = 0.82$, CV R2 = 0.75.

ABCB1(Pgp) Transporter Substrate

Model predicts potential substrates of Pgp (P-Glycoprotein). 60 molecules were used in the training set giving $R^2 = 0.87$. 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

Model was developed using Km values measured on baculovirus-infected insect cells expressing human CYP3A4. The model was trained on a set of 57 compounds (R^2=0.92, MSE=0.14, LOO=0.83). Validation on a separate test set of 17 compounds shows R^2=0.55, MSE=0.19.

CYP2D6 Km

Model was developed using Km values measured on baculovirus-infected insect cells expressing human CYP2D6. The model was trained on a set of 53 compounds and gives $R^2 = 0.93$.

CYP3A4 Ki

Model was developed using Ki values measured on baculovirus-infected insect cells expressing human CYP3A4. The model was trained on a set of 32 compounds and gives $R^2 = 0.83$.

Bioconcentration (logBCF)

The LOGBCF Model, developed by Fraunhofer Institute for Molecular Biology and Applied Ecology (Germany), predicts a compound's bioconcentration factor (BCF) for fish. It was trained from 117 compounds with BCF values based on literature. Cross-validation by leave-one-out method showed a correlation of R^2=0.8 between the predicted and observed values.

CASE STUDY

Development of QSAR model for high-speed *in silico* identification of potentially phototoxic organic compounds

The phototoxic effects of a chemical compound are of concern in numerous areas of chemistry-related industry. Pharmaceuticals, cosmetics, food additives, cleaning agents are just few examples of products that come into frequent contact with the human organism and may cause harm by means of light assisted toxicity. The objective of this study was to create a Quantitative Structure--Activity Relationship (QSAR) computer model that could be used for rapid *in silico* assessment of chemicals' potential to cause harmful phototoxic effects, given its structure.

The structures of 114 compounds known to be phototoxic to humans were retrieved from the literature. The same literature also yielded 36 compounds that did not exhibit noticeable phototoxic effects. Additionally, 78 compounds routinely used in cosmetic products were added to the non-phototoxic part of the training set, yielding a balanced set and increasing its chemical diversity.

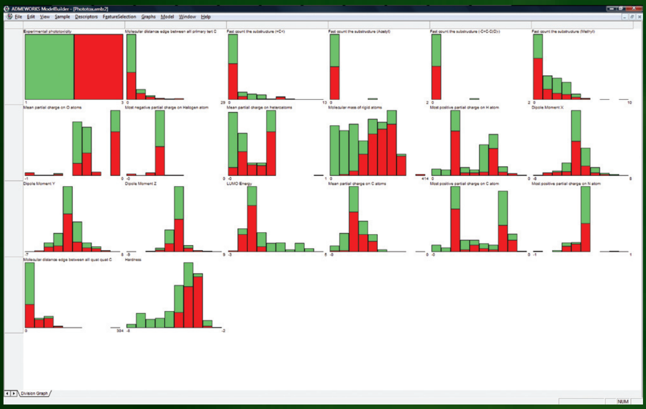

ADMEWORKS ModelBuilder - [Phototox.omb2]
File Edit View Sample Descriptors FeatureSelection Graphs Model Window Help
Experimental phototoxicity
Division Graph

Figure 1. Populations of descriptors for the training set phototoxic compounds – red, non-phototoxic – green.

The QSAR analysis aims to explain the observed (experimental) property by a mathematical expression of "descriptors" – numerical values that may be calculated for a given compound straight from its chemical structure.

The initial choice of descriptors depends upon the assumed mechanism of the reaction. As photoxicity was the objective of this study, it was assumed that the descriptors related to the molecules' quantum properties, shape, charge distribution and existence of specific structural parts may be of importance to the observed activity.

The structures of all compounds were put in an industry standard Structure Data File and all further analyses were performed in the ADMEWORKS ModelBuilder software. 152 descriptors were calculated for the whole training set. The quantum and charge descriptors were calculated using a fast and robust AM1 semiempirical method.

A subsequent data set analysis was performed using the Particle Swarm Optimization algorithm for the feature selection. Next, the 19 descriptors (3 topological, 4 substructure-count and 12 quantum/charge) with the highest potential for explaining the experimentally determined photoxicity were selected. A Linear Discriminant Function (LDA) model was built using the Stochastic Gradient Perceptron algorithm.

The Figures 1-3 illustrate the relationships between descriptors' values and the observed phototoxicity. The population analysis yields no conclusive results as to the significance of any single descriptor to the photoxicity. However, both the clustering and principal component analysis show clearly noticeable tendencies in the training set – the samples with the same phototoxic properties tend to form continuous regions, which is an indication of an existing order in the training set, making it suitable for the creation of a QSAR model.

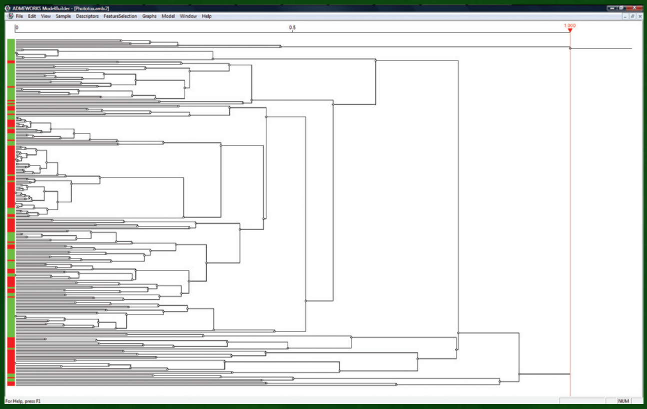

ADMEWORKS ModelBuilder - [Phototox.omb2]
File Edit View Sample Descriptors FeatureSelection Graphs Model Window Help
0
0.5
For Help, press F1
NUM

Figure 2. Clustering of compounds in multidimensional descriptor space.

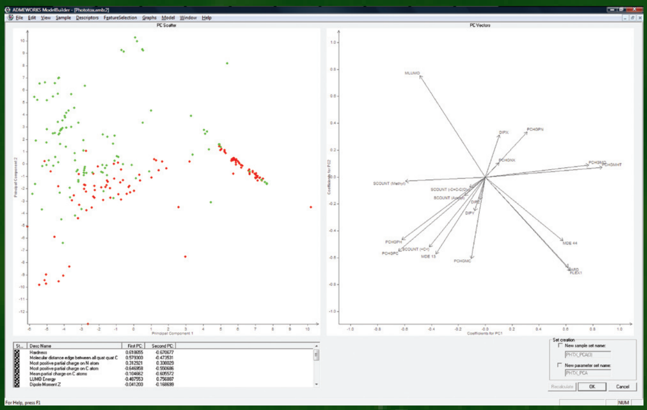

ADMEWORKS ModelBuilder - [Phototox.omb2]
File Edit View Sample Descriptors FeatureSelection Graphs Model Window Help
PC Scatter
PC Vectors
Principal Component 1
Principal Component 2
OK
Cancel
For Help, press F1
NUM

Figure 3. Principal component analysis of descriptor values.

The statistical parameters of the final LDA model created are as follows:

Overall classification rate: 96.05 %

Phototoxic compound classification rate: 100%

Non-Phototoxic compound classification rate: 92.11%

Leave-1-out internal validation rate: 92.54%

Very high overall classification rate (by Leave-1-out cross-validation) as well as 100% classification rate of the phototoxic compounds show the potential of the model for practical use in filtering compounds with unwanted phototoxic effects.

Copyright ©2011 FUJITSU
All rights reserved.