ADMET Predictor™ release 8.5

By Simulations Plus, Inc. 11/6/2017

Dear ADMET Predictor User,

Many major enhancements have been made throughout ADMET Predictor for this release. The biggest is that version 8.5 includes a next-generation HTPK Simulation Module for estimating fraction absorbed, fraction bioavailable and the dose required to achieve a target plasma concentration. There have also been significant improvements to the interface, ADMET property models, MedChem Studio™ module, MedChem Designer™ and ADMET Modeler™, as well as numerous bug fixes.

HTPK Simulation Module

The HTPK (high-throughput pharmacokinetic) Simulation Module allows you to estimate fraction absorbed (Fa) and relative bioavailability (Fb) using the basic ACAT™ intestinal absorption simulation scheme used in GastroPlus™. It now includes paracellular permeability, bile salt effects and precipitation phenomena. Fb is calculated by applying estimated hepatic clearance to the fraction absorbed.

Estimation of the dose needed to attain a given steady-state plasma concentration (OptDose) can still make use of Vd, our ANNE model for human volume of distribution at steady state. Now, however, the mechanistic estimates used in GastroPlus™ are also available for use. These estimates – either the human value (S+hVd_PBPK) or rat value (S+rVd_PBPK) – are displayed in the spreadsheet after any Simulation run.

A model for predicting the percentage of unbound drug in rat plasma (rat_fup%) has been added to complement the corresponding model for human plasma. A model for blood-plasma ratio in rats (RBP_rat) is also now available, as is a model for CYP intrinsic clearance in rat liver microsomes (CYP_RLM_CLint). Access to the relevant human and rat models is enabled by a HTPK Simulation Module license.

Interface & Infrastructure Changes

ADMET Predictor now incorporates many more interface features from the MedChem Studio™ software. The main compound spreadsheet, for example, can now be toggled between the default Table display and a new Tile display in which each spreadsheet cell contains a structure image. In addition, a new control panel is available with dynamic filter bars to hide compounds outside of a desired property range.

Exporting to Excel files has been significantly improved. Perhaps most importantly, the time required to save Excel files is now greatly reduced, often by a factor of 100 or more. The newer XML-based *.xlsx file format is now supported, in addition to the older *.xls format. The newer format allows the number of exported columns to be larger than 256. Embedded images will now sort properly along with the data in the other spreadsheet columns. ADMET property predictions that are "out of scope" will now be displayed using red font, just as in the ADMET Predictor interface.

A new stratified sampling feature is available to select a subset of compounds in the spreadsheet. As with the similar ADMET Modeler feature for test-set selection, the data is sorted using a specified attribute and partitioned as evenly as possible into quantiles so that a subset of specified size can be selected. This subset can be used in conjunction with ADMET Modeler as an external test set, or for other purposes within ADMET Predictor. See the ADMET Modeler manual for details on the implementation.

Additional improvements include:

- Heteroatoms can be colored by atom type. The default convention, where N is blue and O is red, etc., can be customized by the user.
- ADMET Modeler can be launched without sending it the data in the spreadsheet. This is useful when you want to open a Modeler project that you saved previously.
- Predictor can directly open a *.dat file saved with ADMET Modeler, giving you access to the chemical structures and modeling descriptors it contains.
- Predictor can import attributes from a Modeler *.som file. This file contains information about which compounds were contained in the training set and which were contained in the test set.
- A new "confusion matrix" chart type is now available. These charts are similar to ADMET Modeler's performance plots for classification models, showing true positives, false positives, etc.
- Previously generated molecular and atomic descriptors are reused rather than having to be recalculated when new ADMET predictions are requested. This can save a lot of time, particularly when you have compounds with many ionizable groups.
- Cancelling an ADMET property calculation now returns values for all completed rows. This allows you to capture partial results so you can save them and run calculations on the remaining rows at a later time.
- Confidence values can now be hidden in the spreadsheet. This is enabled via the new Display setting called "Display confidence scores in the spreadsheet."
- Large data sets (N>10000) are processed much more quickly, particularly in the MedChem module features.
- Hiding and unhiding spreadsheet rows are now eligible operations for Undo and Redo.
- The pKa microstate window can be saved as an image file (e.g., *.png).

MedChem Studio Module

The MedChem module in ADMET Predictor 8.5 now includes more functionality from the MedChem Studio™ software. There is greater support for subclasses, including a number of new options under the **Classes >> Subclasses** menu. A new **Classes >> Compare Classes** menu item has features to compute various class-versus-class similarity scores. Classes can be annotated with arbitrary user-defined attributes. ADMET Predictor 8.5 now also supports the command-line and Pipeline Pilot functionality from MedChem Studio™.

There is a new feature to estimate the difficulty of chemical synthesis for one or more compounds. Based on work by Ertl and Schuffenhauer (see manual), this new feature assigns compounds a score between 0 (easy) and 10 (hard). The paper reports that these scores are in good agreement with independent difficulty estimates by experienced synthetic chemists.

The R Tables tab has a new feature called **Find Subsets Using R Groups**. This feature is used to discover interesting groups of compounds by requiring that certain R positions have specific substituents. For example, if an R table has five R groups, there might exist a subset of compounds for which three of these groups have fixed substituents (e.g., R1 = CH3, R4 = H, R5 = Cl), with only two groups being variable. Creating a new R table using just this compound subset can greatly facilitate subsequent analysis, since the new R table will have only two R groups instead of five.

The R Group Analyzer feature includes a number of improvements, including support for both tile and circle plots, text inside of spreadsheet cells, colored filter sliders, and more.

MedChem Designer

When you copy a chemical structure to the clipboard, you can pass along up to three predicted ADMET properties. These will be pasted along with the structure image into applications such as Microsoft PowerPoint. The three ADMET properties are selected using the Format tab of the main settings dialog.

The embedded spreadsheet displaying predicted ADMET properties can be limited to a user-specified subset of properties. The list of specified properties is saved to a file so that only these properties are displayed each time Designer is used.

Just as in the ADMET Predictor spreadsheet, Designer has a new option to color heteroatoms by type.

ADMET Modeler

Multi-threading has been implemented to parallelize model building, with the submodels that contribute to an ensemble model now each being built in its own thread. Memory use by Modeler has been significantly reduced for large datasets, especially when data comes from ADMET Predictor.

Confidence Analysis has been improved by allowing separate ("split") beta binomial fits to the positive and negative predictions rather than to a consolidated ("unified") error pool. The "split" uncertainty profiles that result tend to improve reliability for unbalanced datasets over what is seen for the unified alternative. The program tries both unified and split beta binomial fits and chooses the one that works best for each architecture. The addition of Min Confidence to the Model Performance Grid as a new metric complements this innovation, allowing users to assess CFA quality across architectures at a glance. See the manual for further details.

ADMET Modeler includes many other significant changes:

- Antilogit has been added as a new Transform type for models. It can be useful when building models for an endpoint represented by a fraction between 0 and 1.
- An Import Architecture command has been added to facilitate (re)building a new model from exactly the same descriptor set as found in an existing model.
- A Save Model command has been added to the Performance Grid. You can right-click on a single specific model and save it along with its key support files rather than having to save all of the models in the Performance Grid.
- The stratified sampling algorithm has been improved.
- A Sync Selection option has been added to the performance display (pno) windows so that selecting a point in one window simultaneously selects the corresponding point in all open plot performance windows .
- Confusion matrix performance plots have been changed for classification models so that the relative position of points for all models of the grid will always be the same unless the random number seed is changed.
- The location of a new model within Model Editor can now be specified when exporting a model from Modeler to Predictor, thereby controlling its position within the spreadsheet.
- Models can be reordered in Model Editor either via the Edit Model dialog or by simply dragging/dropping row headers in the Model Editor table.
- Some of the GUI controls in the Add Model dialog have been rearranged to better reflect a typical user's workflow.
- The weight initialization schemes used by the Kohonen algorithm for SVM, KPLS, and MLR models have been improved.

ADMET Property Models

- The name of the model for the percentage of unbound drug in plasma has been changed from PrUnbnd to hum_fup% to be more consistent with general usage.
- The model for total intrinsic CYP clearance by rat liver microsomes (CYP_RLM_CLint),

created to support the HTPK Simulation Module, is also available if you have licensed the Metabolism Module.

- The models for rat plasma protein binding (rat_fup%) and for rat blood-plasma ratio (RBP_rat) are also available if you have licensed the Physicochemical and Biopharmaceutical Module.
- The Ames (MUT) models have been rebuilt to take advantage of the new split beta binomial confidence estimation available in Modeler. This has increased sensitivity while providing an improved confidence measure to make it easy to identify those positives most likely to be false alarms.

Bug Fixes

Some of the significant bug fixes are described below.

- Multiple command-line runs can now occur simultaneously.
- Integer attributes are now searchable using **Edit >> Find Text.**
- User-defined equations such as 1 / exp(x) now work properly.
- Addressed freezing behavior when scrolling through large data sets containing hidden rows.
- Addressed a crash that could occur when pinning a chart.
- Addressed a crash that could occur after defining a custom template containing a stereo center.
- Addressed a crash that could occur when using the Combinatorial Transform feature on a compound with a double-bond stereo center.
- Addressed a crash that could occur when starting ADMET Predictor in viewer mode.
- Fixed Modeler's Input Gradient algorithm to correctly set the maximum number of iterations to use when building networks.