ADMET Predictor® release 11.0

By Simulations Plus, Inc. June 2023

Dear ADMET Predictor User,

This new release of ADMET Predictor includes significant enhancements throughout the program. A brief overview of the changes is provided in this document. For detailed descriptions of all changes, please consult the relevant sections of the ADMET Predictor user manual.

Important Notes

This release of ADMET Predictor involves a major change to molecular and atomic descriptors. If you have custom predictive models developed at your organization, these models should be retrained prior to use in this new version.

Improved pKa Model

The pKa model has been greatly improved through careful retraining on a much larger data set of experimental ionization constants. The data set incorporates proprietary data from three new industrial partners: approximately 19000 compounds from Partner 1, 2400 from Partner 2 and 4100 from Partner 3. Overall the number of ionization constants has increased from 33640 in version 10.4 to 70810 in the current version. The expansion of model applicability domain resulted in much more accurate predictivity on the vast majority of our and our partners' external test sets.

Note that due to the resulting change in our ionization descriptors, most other models in ADMET Predictor were retrained for this version. You should expect small differences in predicted values.

3D Virtual Screening

This version includes new functionality to perform 3D virtual screening based on a combination of shape and pharmacophore-feature similarity. Users specify one or more reference 3D structures, then the software computes 3D similarity scores between the reference compounds and those contained in a user-generated 3D conformer database. The creation of such databases is also provided as new functionality in this version. An option is available to run the screen on a compatible Nvidia GPU, which accelerates computations substantially.

Metabolism Module

New regression models have been added to predict inhibition constants (Ki) for seven CYP isoforms: 1A2, 2A6, 2C8, 2C9, 2C19, 2D6 and 3A4.

The classification models for CYP inhibition have also been expanded. The existing models for the major CYPs 1A2, 2C9, 2C19, 2D6 and 3A4 have been improved through the addition of new data, while four new models have been added for the minor CYPs 2A6, 2B6, 2C8 and 2E1.

AIDD

A new objective based on 3D similarity is available. Users specify one or more reference 3D structures, and AIDD calculates for each virtual compound an objective equal to the largest of the 3D similarity scores between the virtual compound and each of the references. The use of this new objective is controlled by a script file called AIDD_Sim3D_Params.txt. See the notes in that file, as well as the user manual, for detailed instructions.

A new run setting allows users to request 3D ADMET models, where available. When this option is invoked, all seed compounds must already have 3D coordinates defined.

HTPK

A new parameter PeffScalingOn has been added to control whether or not input permeability values should be scaled when the selected species is rat or mouse. In prior versions this scaling occurred only when input permeability was set to the name of ADMET Predictor's human Peff model (S+Peff), meaning that scaling did not occur if the permeability was specified as a number. The default value of PeffScalingOn is 1, and this should be changed only if the input permeability has already been scaled to reflect the selected nonhuman species.

ADMET Predictor Service (REST API)

Several script-file workflows can now be executed using the /run_cmd request introduced in ADMET Predictor 10.4. Supported operations include metabolite prediction, logD and solubility versus pH profile generation, 3D conformer generation, compound transformation (e.g., standardization), tautomer generation and the creation of images depicting atomic properties.

A new configuration parameter FEATURE_UPFRONT_CHECKOUT allows users to control whether licenses are checked out immediately upon service startup (default) or only as needed for each individual request.

The /getstatus request can now be submitted without a job ID, in which case the response will indicate which jobs, if any, are complete, in the queue or currently running.

The /predict_admet request supports a new parameter **use3d**. When set to true, predictions are based on 3D models, where available. This feature requires that input structures have 3D coordinates.

Command Line

Several new script-file workflows have been added:

- ms_create_3ddb.txt. Create a 3D conformer database.
- ms_export_3ddb.txt. Export an existing 3D conformer database to an SD file.
- ms_find_3dmatches.txt. Compute 3D similarity scores between one or more 3D reference conformers and those contained in a 3D conformer database.

3D Conformer Generation

Considerable improvements have been made to handling macrocycles containing stereochemistry around double bonds.

New functionality has been introduced to add or remove explicit hydrogen atoms.

A new option was added to increase the number of distance geometry trials when generating coordinates for complex ring systems.

A new advanced option allows users to augment the torsion angle preferences using a file containing custom SMARTS patterns.

Additional Changes

- Opening large SDF and SMILES files has been made significantly faster on machines with multi-core CPUs.
- A new named-property query MaxRCFRingCount returns the largest number of rings in any single ring system.
- SMILES files can now be opened even when the SMILES strings are enclosed by double quotes.
- A new spreadsheet column **3D Alerts** has been added to flag issues related to 3D structures, such as missing hydrogen coordinates.
- A new script-file workflow parameter **-p** writes to the console all available input parameters.

- Double clicking on the **Averaged single proton acidities** and **Averaged site protonations** graphs in the pKa Microstates Display shows additional information about pK50 values.
- Principal Components Analysis (PCA) includes a new option to use all descriptors. Previously users were forced to choose descriptor subcategories.
- New transforms have been added to moleculeTransforms.crf, which is used by the AIDD and Combinatorial Transforms features.
- A new advanced option controls whether the Enter key can be used to input new lines in dialogs for specifying queries and reactions (e.g., for substructure search).
- The script file workflow for tautomer generation (ms_generate_tautomers.txt) supports a new parameter to write input compounds even if they do not yield additional tautomers.
- Two new modeling descriptors – QAtomPos and QAtomNeg – have been added.
- The following models have been removed: MHSw-MP, MHpH-MP, MHIS-MP, MHSF-MP, MHSp-MP, MHSw-NoMP, MHpH-NoMP, MHIS-NoMP, MHSF-NoMP, MHSp-NoMP, HIVI-ST, HIVI-TC, CYP3A4_Inh_midaz, CYP3A4_Inh_testo, CYP3A4_Ki_midaz and CYP3A4_Ki_testo.

Bug fixes

The most significant bug fixes are described below.

- Several small memory leaks in the REST API have been addressed.
- The command-line option **-path** could not handle paths enclosed by double quotes.
- The command-line version of AIDD could not process scaffolds contained in *.mol files.
- A minor bug was addressed in the BFGS optimization algorithm used within ADMET Modeler and the HTPK module.
- Regression uncertainty scores could not be exported for models that supported text values (e.g., Estro_RBA and OATP1B1_Km).
- Problems interpreting complex SMILES stereochemistry have been addressed. The following is an example of a compound that is now processed correctly: [C@@H]12C[C@H]3C[C@@H](C1)C[C@@H](C2)C3
- Structures output using the command-line arguments “-w CC” (Computational Chemist) and “-w PC” (Physical Chemist) could contain incorrect bond types in rare cases.