

version 1.1.4

USER'S GUIDE

Table of Contents

1.Introduction.....2
2.Installation.....4
3.How to Use.....5
4.1 Insert.....5
4.2 Select.....6
4.3 Export.....6
4.4 Predict.....7
4.What's new.....8
5.Acknowledgements and contacts.....9

1. Introduction

VEGA is an application that allows to apply several QSAR models to the given chemicals provided by the user, returning as output complete reports (in plain text and in PDF format) containing the predictions and all needed supporting information. It is a JAVA standalone application and it does not require any network communication (chemicals are processed on the local machine, no information is shared across the Internet).

Currently, there are **33 models** implemented in VEGA:

Five models from the CAESAR project (www.caesar-project.eu):

- BCF
- Mutagenicity (Ames test)
- Carcinogenicity
- Developmental toxicity
- Skin sensitization

Two models from EPA's T.E.S.T. Software (www.epa.gov/nrmrl/std/cppb/qsar/):

- Fathead Minnow LC50 (96h)
- Daphnia Magna LC50 (48h)

One model from DEMETRA project (www.demetra-tox.net/):

- Daphnia Magna LC50 (48h)

Three models from EPISuite (www.epa.gov/oppt/exposure/pubs/episuite.htm):

- BCF (Meylan model)
- LogP (Meylan/Kowwin)
- kM/Half-life (Arnot)

Eighteen models developed by Istituto di Ricerche Farmacologiche Mario Negri:

- Mutagenicity Consensus
- Mutagenicity (Ames test) Sarpy rule set
- Mutagenicity (Ames test) KNN/Read-Across
- Carcinogenicity (based on the Antares dataset)
- Carcinogenicity (based on the ISSCAN-CGX dataset)
- ER Relative Binding Affinity (RBA)
- ER-mediated effect (CERAPP)
- Hepatotoxicity model

- Fish acute (LC50) toxicity classification Sarpy rule set
- Fish acute (LC50) toxicity KNN/Read-Across
- Bee toxicity model
- BCF KNN/Read-Across
- Ready Biodegradability Sarpy rule set
- Persistence in sediment
- Persistence in soil
- Persistence in water
- LogP (AlogP)
- LogP (MlogP)

Two models from ISS - ToxTree (toxtree.sourceforge.net):

- Benigni/Bossa rule set for Carcinogenicity
- Benigni/Bossa rule set for Mutagenicity (Ames test)

One model from National Institute of Chemistry Slovenia:

- Fish acute (LC50) toxicity (NIC)

One model from Procter&Gamble:

- Developmental/Reproductive toxicity virtual library

2.Installation

VEGA is a JAVA application that works on every operating system (Windows/Linux/Mac) with **JAVA Runtime Environment (JRE) version 7** installed. Before proceeding please check that the latest JRE (freely available at http://www.java.com) is installed on your computer.

Download and unpack the zipped file. To start the application, just move to the VEGA folder and run the file *Vega-launcher-WIN.bat* (on Windows platforms) or *Vega-launcher-LINUX.sh* (on Linux platforms).

If you experience problems with the above mentioned launchers, you can just run the file *Vega.jar* with the Java Runtime Environment (JRE). On most systems, it is enough to double-click it. If you are not able to directly run it , open a command line window (like Command Prompt on Windows systems or BASH shell on Linux), move to the Vega folder and type:

java -jar Vega-1.1.4.jar

Note: Vega may encounter problems when processing large datasets and/or using a high number of models at the same time. You can modify and use the launcher files (Vega-launcher-WIN.bat and Vega-launcher-LINUX.sh) to assign more memory to the application. In the launcher file modify the parameter "-Xmx1024m", where 1024m means that Vega uses 1,024 MB (1 GB), setting a higher number (for instance -Xmx2048m for running Vega with 2 GB of memory).

3. How to Use

VEGA has a user-friendly interface, organized in three tabs corresponding to the steps needed to perform calculations. In order to run the models, all the parameters of the three tabs must be correctly set. In the first tab,INSERT, you have to build the compounds dataset. In the second tab, SELECT, you have to select the models to be calculated. In the third tab, EXPORT, you have to select the type and the destination folder for saving the generated reports.

4.1 Insert

In this tab, you have to insert one or more structure(s) in order to build the dataset that will be used for the calculations. You can insert the structure by directly typing it, using the SMILES format, in the "Load" text field; it is also possible to copy-and-paste single SMILES in the text field. Then, click the "Load" button, or just press enter to insert the structure. If the SMILES format of the structure is correctly typed, then it is imported in the table below the text field.

Also, you can import one or more structure(s) from SDF or SMILES file formats. SDF (structure-data file) is a chemical data file format developed by MDL. This kind of text-based file is specifically formatted to include multiple structures supplied with data (e.g., name, CAS, molecular mass, activity etc.). SMILES files are text files containing multiple SMILES strings (one molecule for each line of the file). For this kind of files, also tab-separated text files are accepted: for each line, other information for the given SMILES molecule (like its id or CAS number) can be reported, separated by a tab character.

To import a file, click the "Import file" button and select the file to be loaded. Depending on the type of the file, further information could be required:
- For SDF files, if some tags are available a form will give you the chance to choose one of the tags as the id (name) of the molecules that will be loaded.
- For SMILES files, if the file contains tab-separated fields a form will give you the chance to choose what field contains the SMILES string, and to optionally choose another field as the id (name) of the molecules that will be loaded.
In both cases, it will be possible to select the option for parsing the chosen molecule's id as a CAS number: the read string will be normalized so that CAS numbers written with separation characters other than "-", or CAS written without any separation character will be reported in a default style (like 12345-67-8).

After successfully importing the structures using one or both of the ways above, you can visualize the structure by clicking on it.

By selecting some of the imported structure(s) and clicking the "Delete" button it is possible to erase them. To clean the whole table just click "Delete All" button.

4.2 Select

In the second tab, you have to select the models to be calculated. The currently available models are organized in four categories based on the end-point type: Toxicological (Tox), Eco-toxicological (Ecotox), Environmental (Environ), Physical-Chemical properties (Phys-Chem). Just click to select or de-select the models.
By clicking the question mark icon near each model, you can access to the details of the model and to the list of available files for that model: its user's guide (PDF document); the original dataset used for building the model (plain text file) containing the CAS number, the SMILES, the set (training or test set) and the experimental value for each compound; the external link (opening in the default web browser) for the QMRF on the JRC database if available.

4.3 Export

In the third tab, you have to select the type of the reports to be generated, and the destination folder where the reports will be saved. Reports can be saved in two different formats, PDF and plain text. The supported output choices are the below:

- Multiple PDF documents – Results are saved in PDF format, one file for each calculated model
- PDF single document (ordered by model) – Results are saved in a single PDF report, ordered by model
- PDF single document (ordered by molecule) – Results are saved in a single PDF report, ordered by molecule
- Summary as tab-separated text file – Results are saved in a single text file, where for each model is reported only the final assessment

- Multiple tab-separated text files – Results are saved in text files, one for each model

For each choice, you have also to select the destination folder where the reports will be saved after the calculation. It is possible to select one or more export choices.

4.4 Predict

When the parameters of the three tabs described above are correctly filled, just click “Predict” button to start calculations. Please note that the calculation and report generation task may take long time, depending on the number of the imported substances and the selected models. In particular, note that the task of creating PDF reports can be particularly time-consuming when calculation are done for high numbers of compounds

4. What's new

Version **1.1.4**: this release includes two new models (kM/Half-life and Hepatotoxicity).

Version **1.1.3**: this release includes one new model (Bee acute toxicity).

Version **1.1.2**: this release includes three new models (fish acute toxicity, ER-mediated effect and mutagenicity consensus).

Version **1.1.1**: this release includes five new models (two for carcinogenicity endpoint, three for persistence endpoint). Some minor bugs has been fixed (in particular: problem with the selection of the output directory for PDF reports; problem in the output of the mutagenicity Sarpy/IRFMN model).

Version **1.1.0**: this release includes a major update of the calculation core. Due to this update, predictions and applicability domain assessment could result slightly different from older versions.
Several new models have been added, and exporting of training set has been improved.

5. Acknowledgements and contacts

VEGA application has been developed by Istituto di Ricerche Farmacologiche Mario Negri (Laboratory of Environmental Chemistry and Toxicology) and Kode srl. The official website of the project is http://vega-qsar.eu

Please check each model's guide for acknowledgements on its development.

Please check the official website for acknowledgements on the EU projects that have supported the development of VEGA.

For questions regarding the software and its use please contact:
a.manganaro@kode-solutions.net