Upload a QSAR model

Uploading Linear Models to the OCHEM Website

Models can be saved to and uploaded from excel sheets. The model upload is only supported for linear (MLRA) models.

Uploading a linear model

Sample file

Sample file for the linear model upload

Description

In order to upload a model, one needs to make sure first that have the compounds and their respective properties used to develop the model are stored in the database. If you have not done so already, use the "Database->Batch upload" feature to upload the compound and property data. Clicking on the batch upload title will lead you to the help on how to use it.

It is strongly suggested that you verify the uploaded data against the database. One can use the 'Select similar compounds" feature to check for duplicates, deviant measurements and typing errors.

In order to upload a model, although it is not strictly necessary, i suggest that one creates a linear model by the modeling capabilities of the database. To do so, follow the Models->Create a model help. Having successfully create such a model, use the 'export to xls' function (Models->Apply a model->Export to XLS) to create an excel sheet with containing the model data.

In this file, there will be a sheet named "Linear regression data", which will also be searched for and used for model upload. If there is no such sheet the first sheet will be used. The first column will hold the descriptor names and be titled 'DESCRIPTORS', the second column holds the respective linear regression coefficients. The reserved descriptor name 'BIAS' is used for the model bias.

If you want to upload your model coefficients, preferably just replace them in this excel file. It is, of course, also possible to create a similar excel file 'from scratch'.

After the data is filled in, proceed to (Models->Upload your model). Select the training and validation set which will be used to display model statistics only, then select all the descriptor types that you have used. The descriptors will be recalculated for your compounds to ensure compatibility with the other compounds in the database and therefore applicability of your model to compounds other than used during model development. Note that using 3D/Structure optimization parameters will require you to wait for the structural optimization and its verification of the molecules used for modeling which is usually a matter of days, not minutes.

You will be asked to verify the model parameters read from the excel sheet. After that, you can verify your model statistics and save your model.