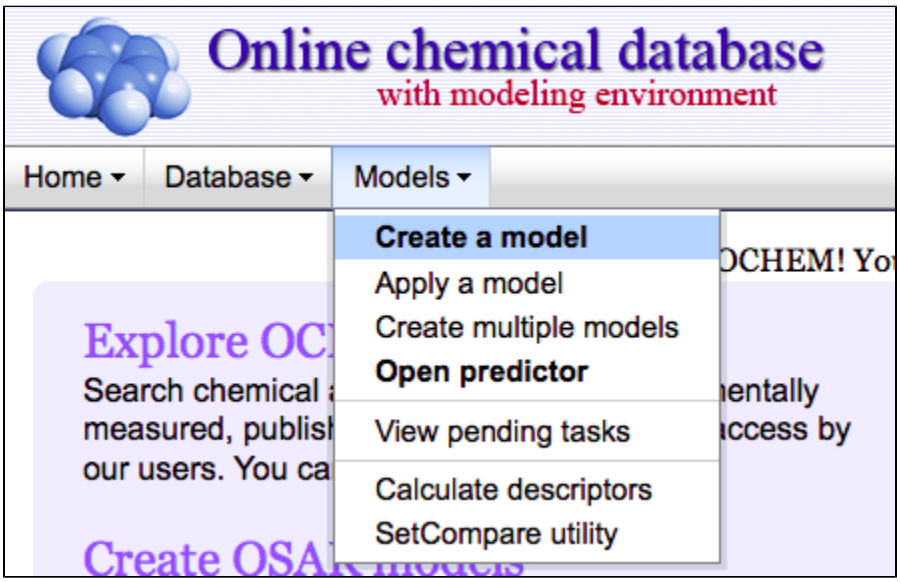
Creation of a single model

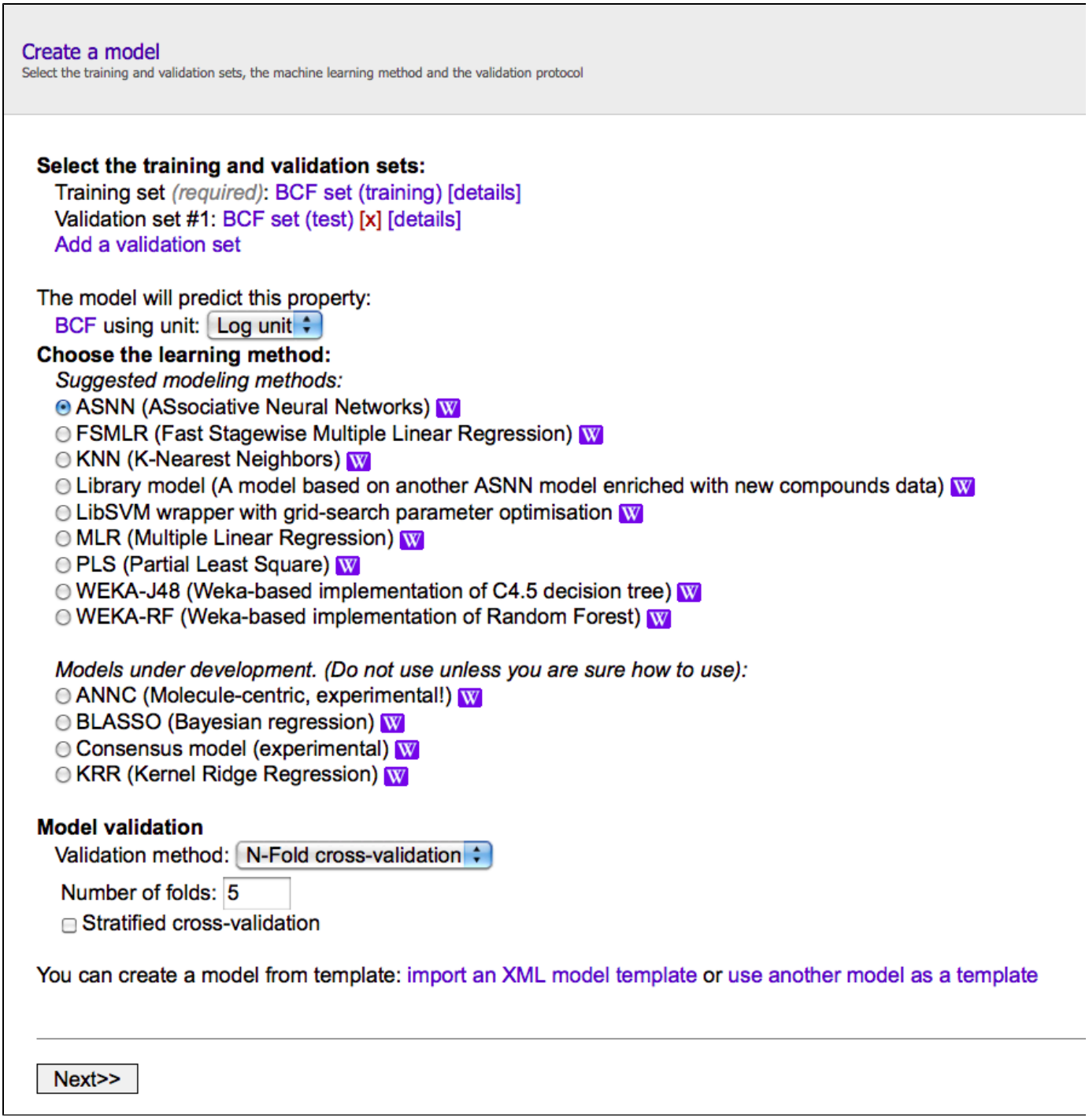
Model creation wizard

You can create a single model using web interface by accessing the "Models > Create a model" menu as shown below, which will open a wizard-like dialogue with a series of the necessary configuration steps.



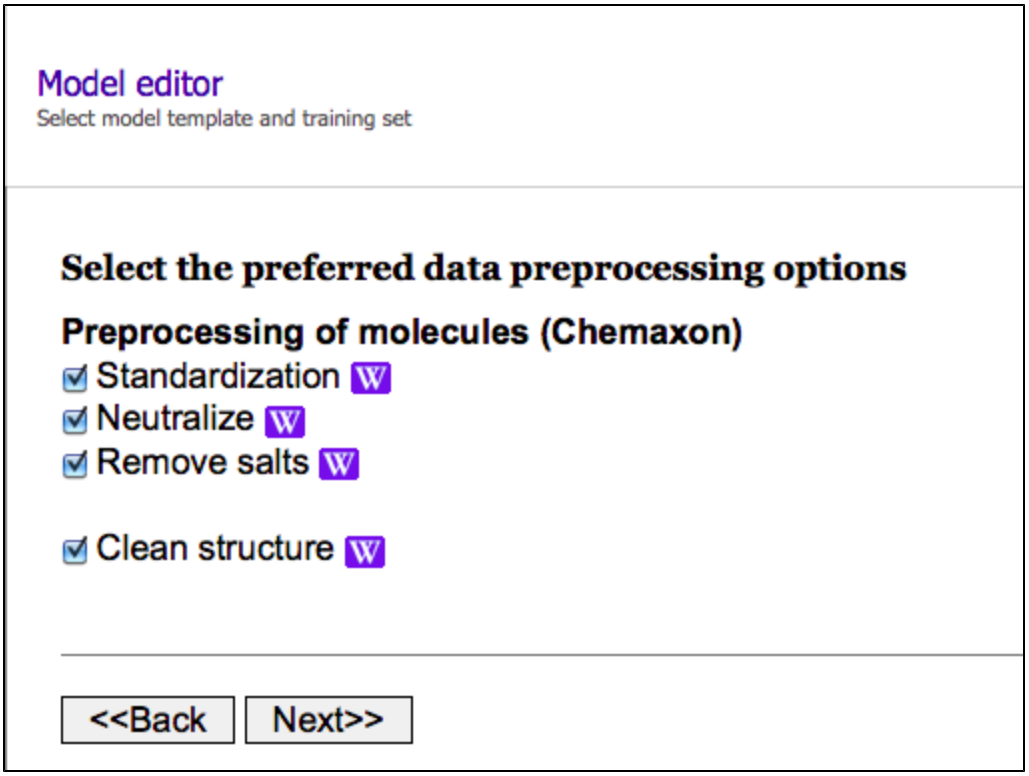
First, you have to select a training and (optionally) a validation set. It is possible to have multiple validation sets. We assume that you have already prepared the sets.
On the same screen, it is required to select:

- the unit of measurement for the model
- the desired machine learning method
- the desired validation protocol (cross validation, bagging or no validation)



The next screen allows to configure the preprocessing of molecular structures. The options include:

- Standardization (e.g., of nitro-groups)
- Neutralisation of ions
- Removal of salts (by keeping the counter-ion)
- Cleaning of structures (by converting it to SMILES and back to SD-file)



Molecular descriptors

The next important step is the choice of molecular descriptors. OCHEM supports more than 20 descriptor packages provided by different vendors.
OCHEM policy is to integrate state-of-the art descriptors rather than to develop own solutions.

Select the molecular descriptors:

Suggested descriptors:

- [x] E-state

E-State types:

- [x] Atom indices
- [x] Bonds indices
- [ ] Atom counts
- [ ] Bonds counts

Aromatize structures: Chemaxon Basic

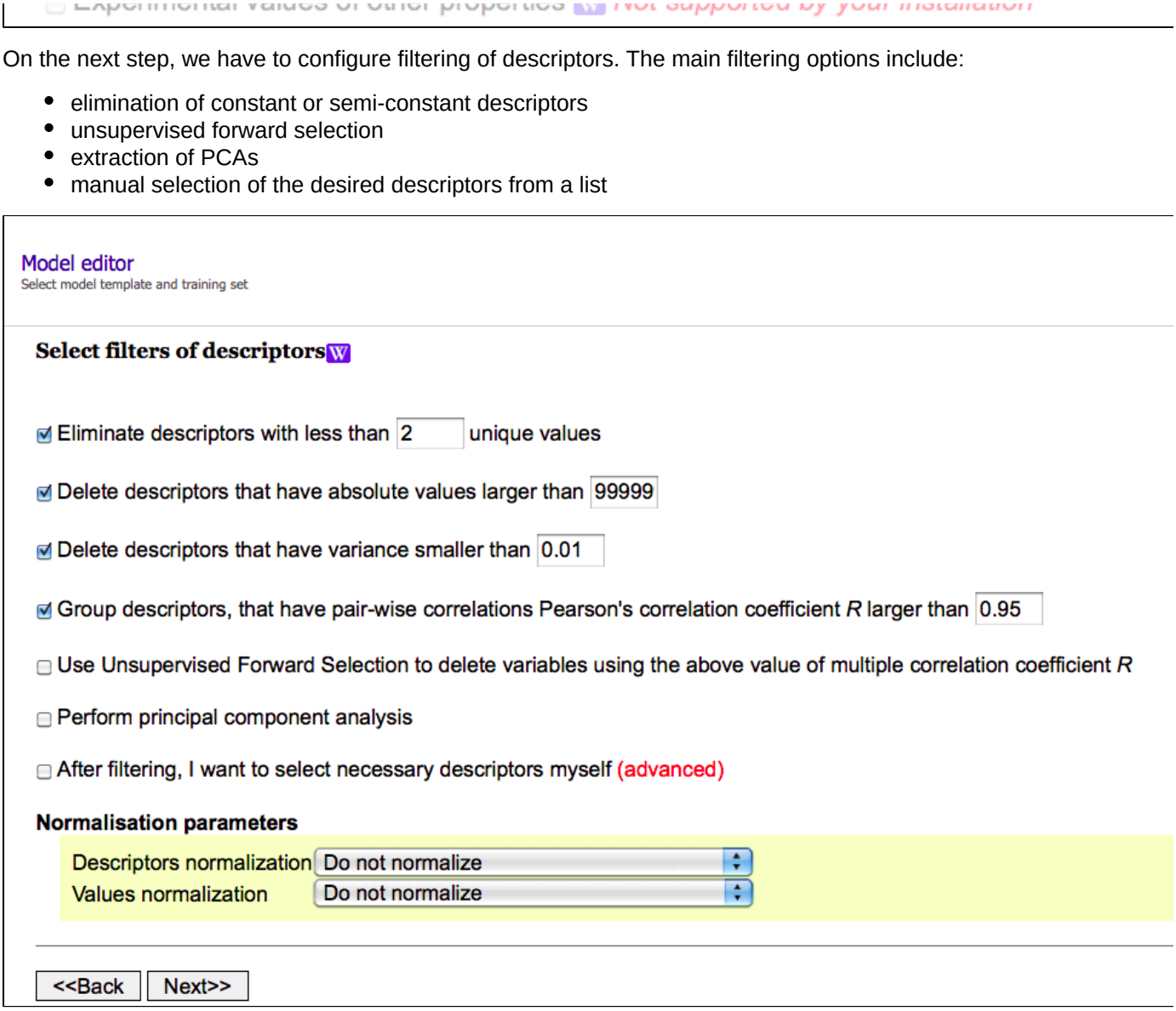
- [x] ALogPS (2)
- [ ] GSFragment (1138)
- [ ] Dragon v. 6.0 (4885/3D)
- [ ] ISIDA fragments
- [ ] ADRIANA.Code *(211/3D)*
- [ ] CDK descriptors (246/3D)
- [ ] 'Inductive' descriptors *(54/3D)*
- [ ] MERA descriptors *(529/3D)*
- [ ] MERSY descriptors *(42/3D)*
- [ ] Chemaxon descriptors (499/3D)
- [ ] QNPR
- [ ] Spectrophores (144/3D)

Additional or obsolete descriptors:

- [ ] OEState
- [ ] MolPrint
- [ ] Dragon v. 5.4 (1630/3D)
- [ ] Dragon v. 5.5 (3190/3D)
- [ ] Structural alerts (ToxAlerts)
- [ ] MOPAC descriptors *(21/3D)*
- [ ] ShapeSignatures *(3D)*

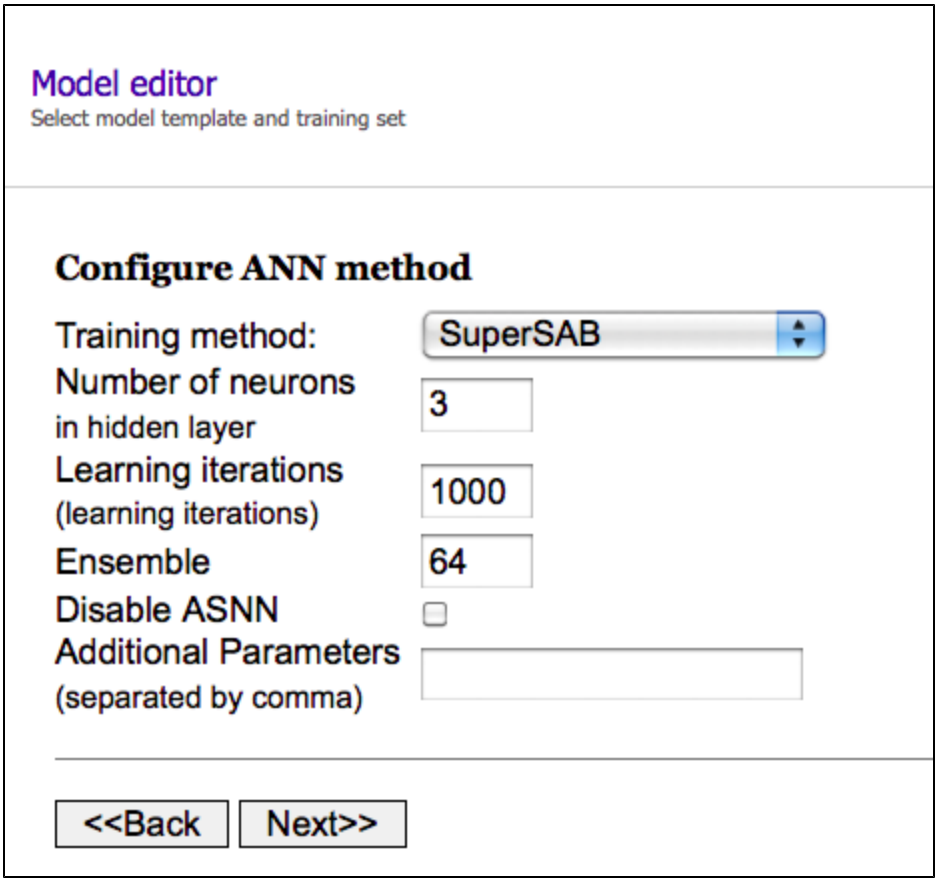
Experimental descriptors (use only if you know how to use them):

- [ ] Custom descriptors from a file
- [ ] AMBIT Descriptors
- [ ] ISIDA fragments (2011)
- [ ] Chiral Descriptors (/3D)
- [ ] Scaffold Hunter Descriptors
- [ ] Functional Groups
- [ ] ScrambledDragon (tmp) *Not supported by your installation*
- [ ] ACD pKa (tmp) *Not supported by your installation*
- [ ] Random descriptors (for testing) *Not supported by your installation*
- [ ] ETM descriptors
- [ ] DockingDescriptors (pre-pre-alfa) *Not supported by your installation*
- [ ] Experimental values of other properties *Not supported by your installation*



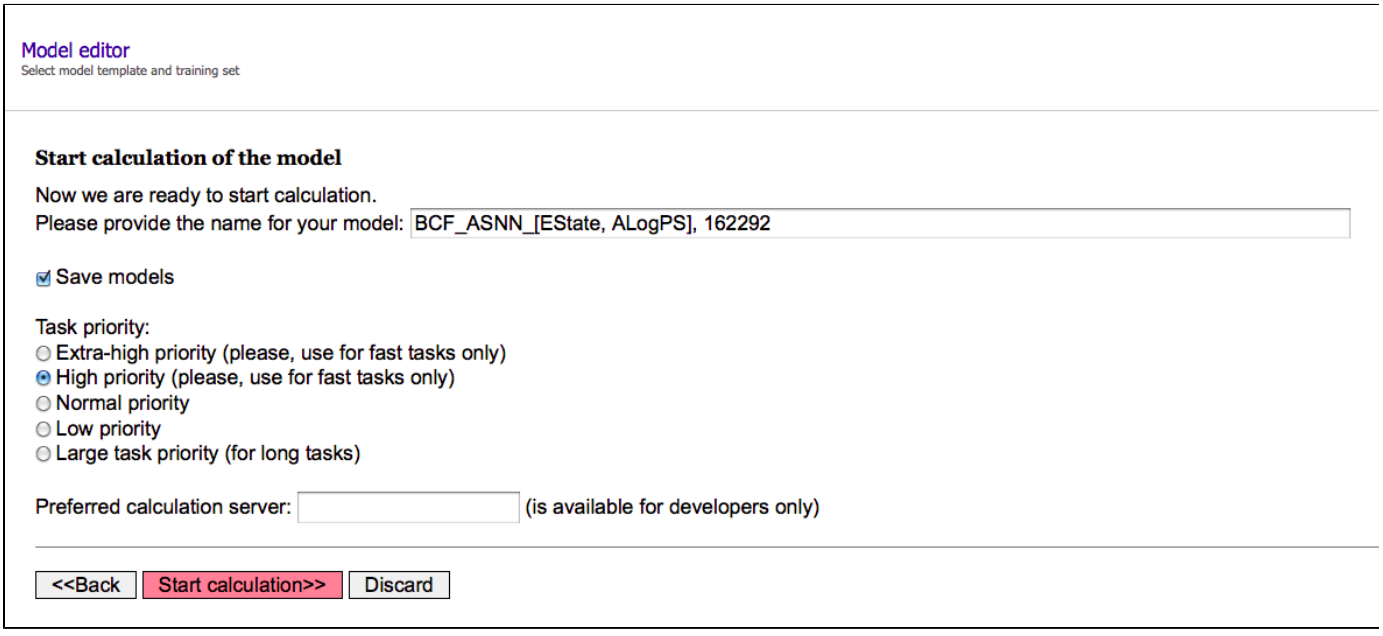
Machine learning method

Next, we have to configure the machine learning method. This step is method-specific. The screenshot below shows the configuration options for a neural network model.
Naturally, the other machine learning methods (e.g., linear regression, PLS, random forest, etc) will have different options.

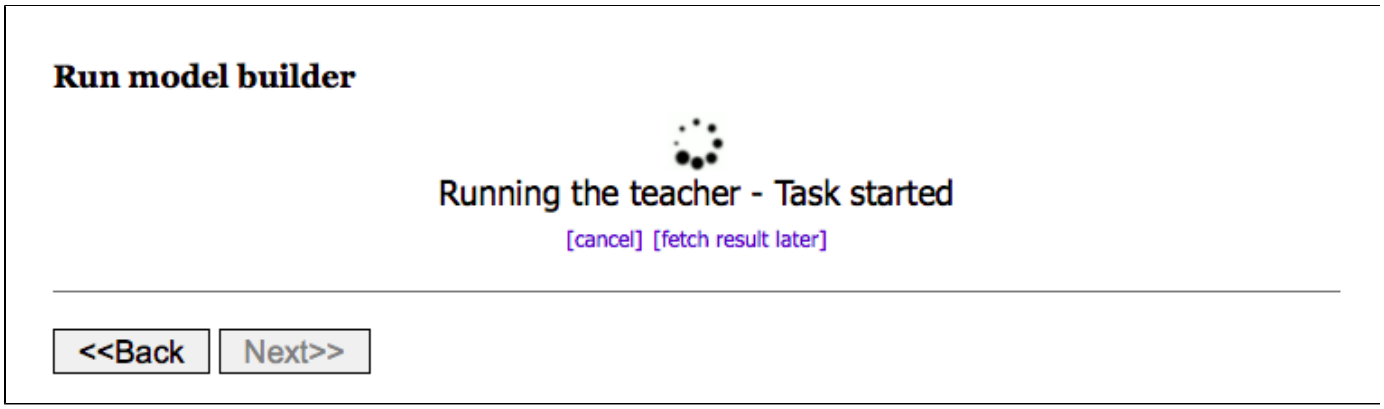


Initiating the calculations

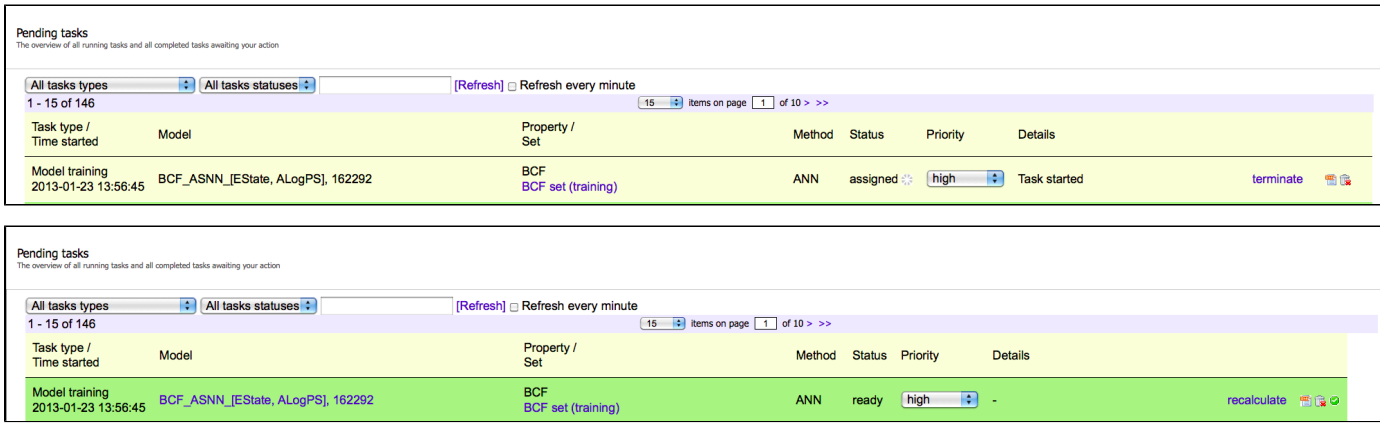
Now, we are ready to start calculations. In the dialog below, we have to give our model a name, define the priority of this task and, finally, initiate the calculations.
We recommend using high and extra-high priorities only for fast tasks (e.g., models with less than 300-500 molecules in the training set).



The next screen shows the progress of calculations. It is possible to fetch results later (in one hour, one day or more) from the registry of pending tasks.

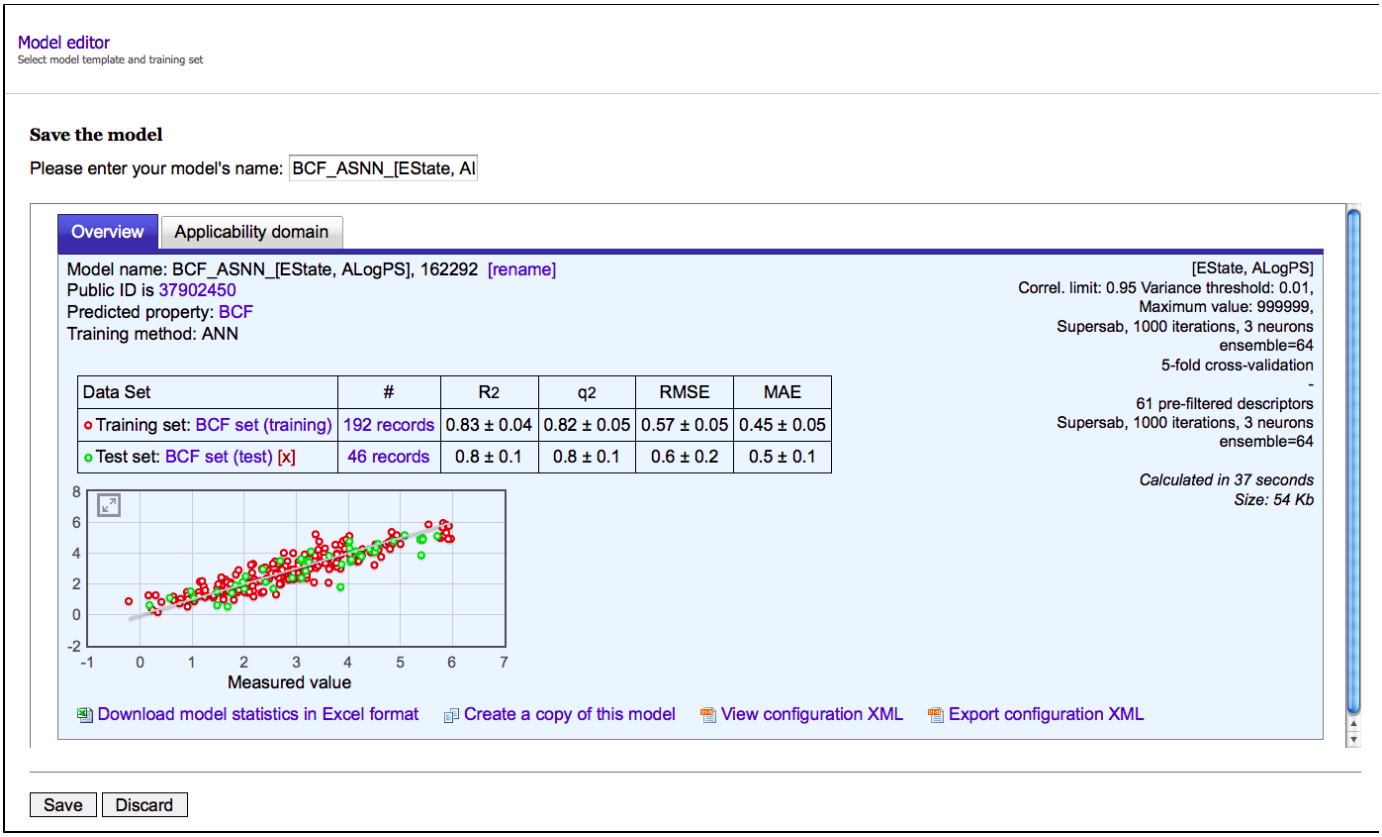


The registry of pending tasks displays the status of all currently running calculation tasks. Once the task is ready, it will become "green", and it will be possible to fetch it by clicking a green checkbox icon on the right.



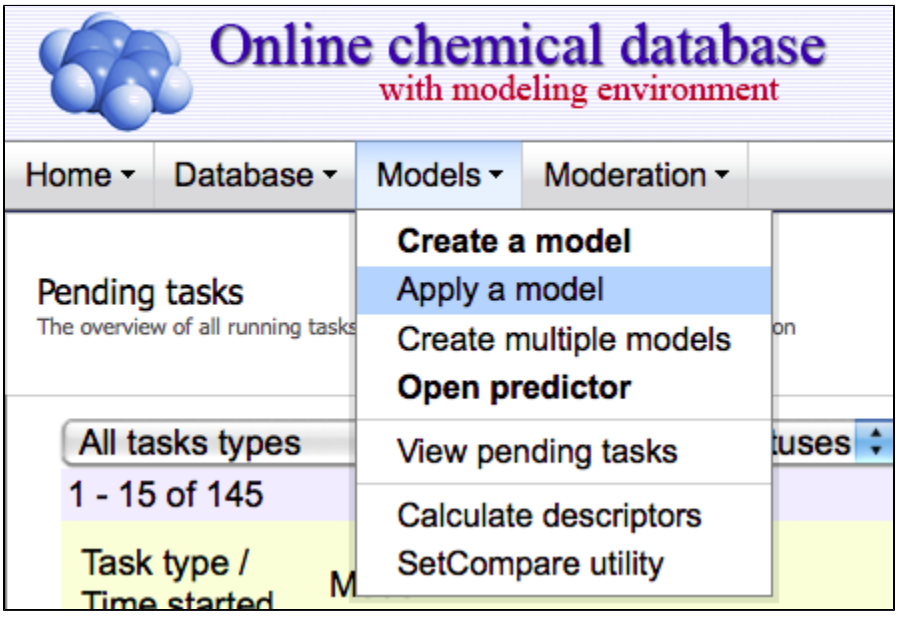
After you fetch the calculation task you will be presented with a model profile.

The model profile displays basic model statistics (RMSE, R2, etc), scatter plot, applicability domain plots, etc. On this step, you can either save your model or discard it in case if the model performance is unsatisfactory.



Accessing your model

After you save your model, it will become available from the browser of models:



The browser of models displays all the publicly available models developed by other OCHEM users as well as your own private models. Here, it is possible to search for models by various criteria.
The displayed information includes:

- model name
- training and validation sets

- machine learning method

Each available model has a model profile described above, which contains the detailed statistics, the scatter plots and confusion matrices for the classification models.