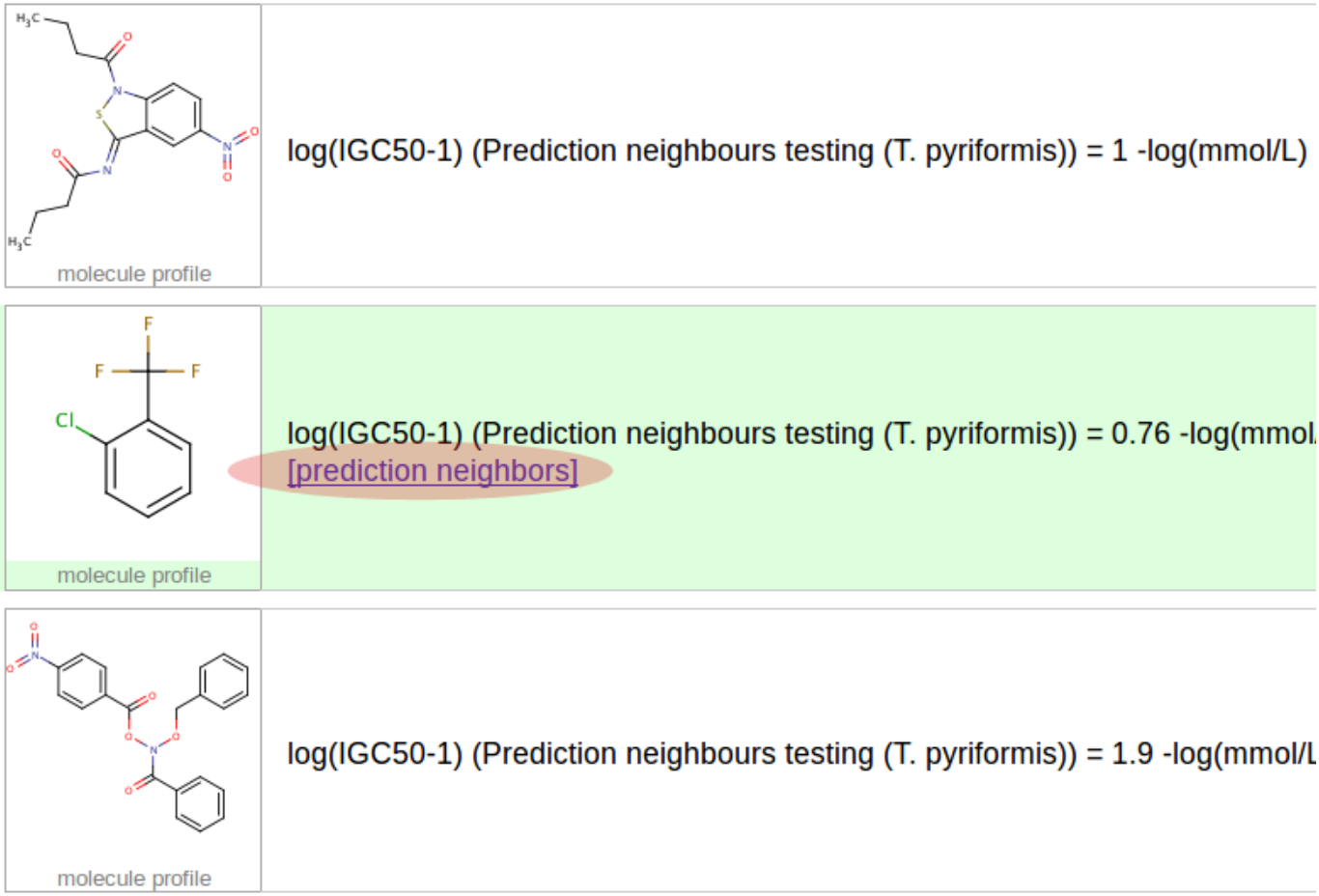
Prediction neighbors

OCHEM allows exploring the training set compounds that are in some way similar to the predicted compound (so called "*prediction neighbors*").

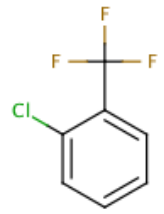
The **prediction neighbors** feature might be helpful for interpretation of a particular prediction.
Are there similar compounds in the training set? Are they active or inactive? These questions are addressed with the prediction neighbors feature.

To browse the neighbors of a particular prediction, simply click the "prediction neighbors" link as on the screenshot below:



This will open the prediction neighbors explorer, which will show the similar structures from the training set and the detailed info (predictions, experimental values, etc.).

The predicted compound



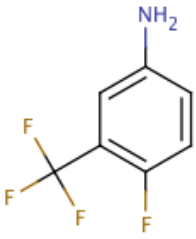
molecule profile

log(IGC50-1) (Prediction neighbours testing (T. pyriformis)) = 0.76 -log(mmol/L)

Nearest training set neighbours

Similarity measure: Structural similarity

Similarity: 0.75



molecule profile

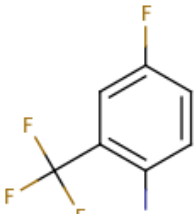
● log(IGC50-1) = 0.77 (in -log(mmol/L))
Predicted value: 0.11 (in -log(mmol/L))
BAGGING-STD: 0.14

Zhu, H
Combinatorial QSAR modeling of chemical toxicants tested aga...
N: 109
J. Chem. Inf. Model. **2008**; 48 (4) 766-84

2357-47-3

[open in browser] [prediction neighbors]

Similarity: 0.72



● log(IGC50-1) = -0.02 (in -log(mmol/L))
Predicted value: 0.55 (in -log(mmol/L))
BAGGING-STD: 0.20

Zhu, H
Combinatorial QSAR modeling of chemical toxicants tested aga...
N: 244
J. Chem. Inf. Model. **2008**; 48 (4) 766-84

393-39-5

Similarity types

There are several ways to define the "nearest" neighbors from the training set. The most obvious one is structural similarity (e.g., using Tanimoto similarity). This is the default option.

If the model was built as an ensemble (that is, using bagging validation protocol, configured to store the individual ensemble predictions), the two other similarity measures are available:

- Correlation in the prediction space
- Euclidian distance in the prediction space