Create a dataset and upload descriptors at https://ochem.eu/descriptorsstorage/show.do

OCHEM allows you to use your own precalculated descriptor values for the development of QSAR models.
You can upload descriptors as an Excel or csv file with columns "MOLECULE (or EXTERNALID for molecules without structures), D
molecule (e.g. M12345) or SMILES string. N.B.!: do not use identifiers starting with "M" as EXTERNALID. They will be erroneously
A couple of examples with and without molecules: Exemplary file 1 Exemplary file 2

Descriptor type name
e.g., Dragon6, MyCustomDescriptors, E-State: 468 descriptors

Excel file with descriptors
first line lists the descriptor names as specified above: Browse... 3_tox24_alogps_oestate+EXTERNAL_sort1_s_leaderboard.xlsx

Start the upload | Cancel

Privately stored descriptors

Descriptor type	User	Entries	Storage size	Operations
468 descriptors		512	3961039 bytes	

Select the training and validation sets:

Training set *(required)*: Copy of tox24_challenge_train_BL1 [details]
Add a validation set

The model will predict this property:
TTR binding activity using unit: %

Choose the learning method:

Suggested modeling methods:

- (•) ASNN: ASsociative Neural Networks doi:10.1007/978-1-60327-101-1_10
- ( ) Transformer-CNF model
- ( ) Consensus model (based on models developed for the same set)
- ( ) DEEPCHEM: several methods from DeepChem (GPU) arXiv:1703.00564
- ( ) DNN: Deep Neural Network (GPU) doi:10.1021/acs.jcim.8b00685
- ( ) FSMLR: Fast Stagewise Multiple Linear Regression doi:10.1134/S0012500807120026
- ( ) Image based QSAR doi:10.1021/acs.jcim.0c01288
- ( ) *KGCNN collection of models (e.g., AttFP, PAiNN, HamNet, etc.) from https://github.com/aimat-lab/gcnn_keras
- ( ) KNN: k - Nearest Neighbors

New set will be available at the descriptor page

Recommended descriptor types (2D)

- ☐ OEState
- ☐ ALogPS (2)
- ☐ Mold2 (777)
- ☐ CDDD
- ☐ JPlogP
- ☐ SIRMS
- ☐ ISIDA fragments
- ☐ The in Hashed Atom Pair fingerprint (MAP4)
- ☐ GSFragment (1138)
- ☐ QNPR
- ☐ Multilevel Neighborhoods of Atoms (MNA)
- ☐ Structural alerts (about 3.5k ToxAlerts and Functional Groups)

Recommended descriptor types (3D)

- ☐ alvaDesc v.2.0.16 (5666/3D)
- ☐ Dragon v. 7 (5270/3D)
- ☐ CDK 2.8 descriptors (256/3D)
- ☐ RDKit descriptors *(3D)*
- ☐ MORDRED descriptors (1826/3D)
- ☐ MOPAC2016 descriptors *(35/3D)*
- ☐ KrakenX descriptors (MOPAC2016 derived)*(124/3D)*
- ☐ PyDescriptor descriptors *(16251/3D)*
- ☐ MERA descriptors *(529/3D)*
- ☐ MERSY descriptors *(42/3D)*
- ☐ 'Inductive' descriptors *(54/3D)*
- ☐ Spectrophores (144/3D)

Stored descriptors (explore the descriptors storage)

☑ 468 descriptors

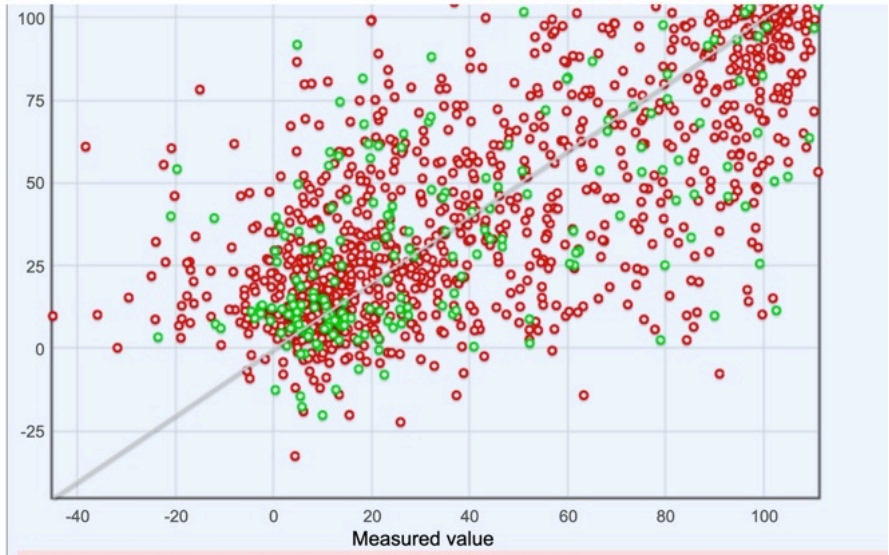
Model creator

Select model template and training set

Save model and add your custom descriptors at https://ochem.eu/multiplemodels/create.do as shown on other pages

Save the model

Please enter your model's name: TTR binding activity_ASNN



Number of compounds ignored because of errors in original model = 7

[Exclude duplicated records]

Export this model | Create a copy of this model | View configuration XML | Export

Save | Discard

https://ochem.eu/multiplemodels/create.do

The comprehensive modeling feature allows you to simultaneously run multiple models with different
Please note that running multiple models may require significant computational resources and time.

Select the training and validation sets:

Training set *(required)*: [...]
Add a validation set

Select the methods you want to use for the modeling:

Method

[all] [none]
- [x] FSMLR (to cache descriptors)
- [x] ASNN (MTL, bias correction)
- [ ] KNN
- [ ] MLRA
- [ ] PLS
- [ ] WEKA-RF (classification only)
- [ ] WEKA-J48 (classification only)
- [ ] LSSVMG (MTL, GPU)
- [ ] DNN (MTL, GPU)
- [ ] XGBOOST
- [ ] RFR
- [ ] DeepChem TexCNN (no descr, MTL, GPU)
- [x] TRANSNNI 10/10 (no descr, MTL, GPU)
- [x] Transformer CNF2 (no descr, MTL, GPU)
- [x] KGCNN AttFP (no desc, MTL, GPU)
- [x] KGCNN ChemProp (no desc, MTL, GPU)

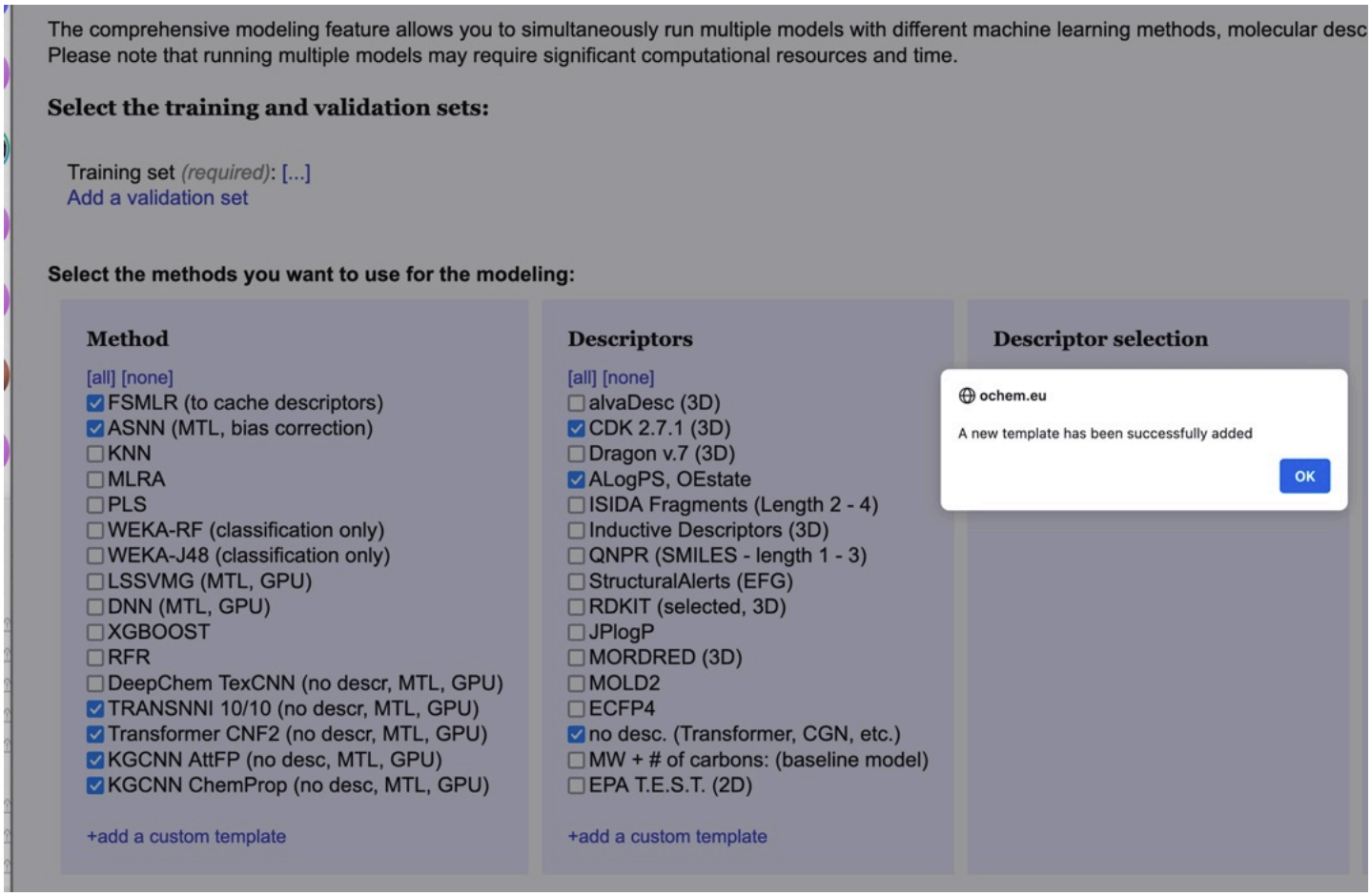
+add a custom template

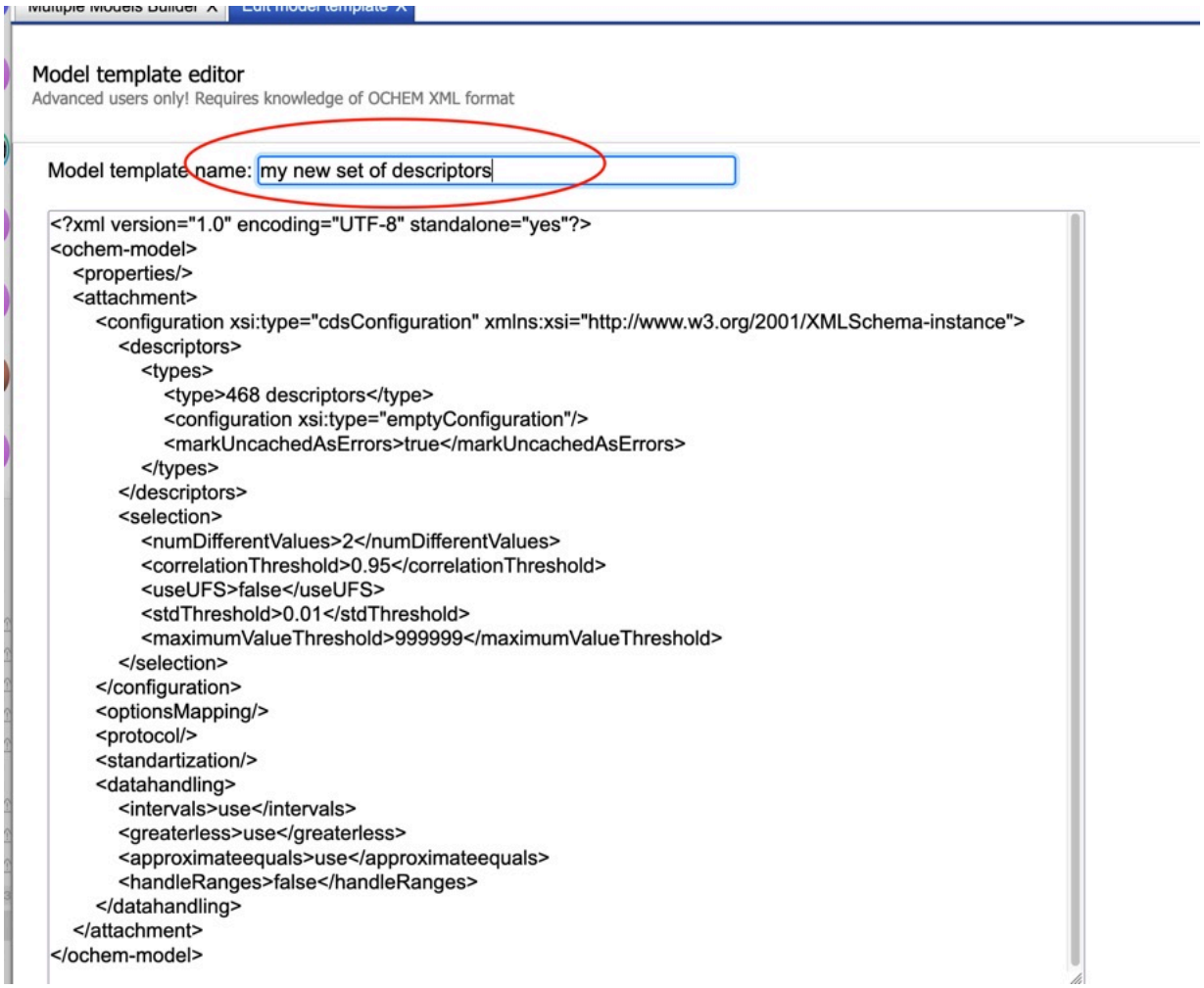
Descriptors

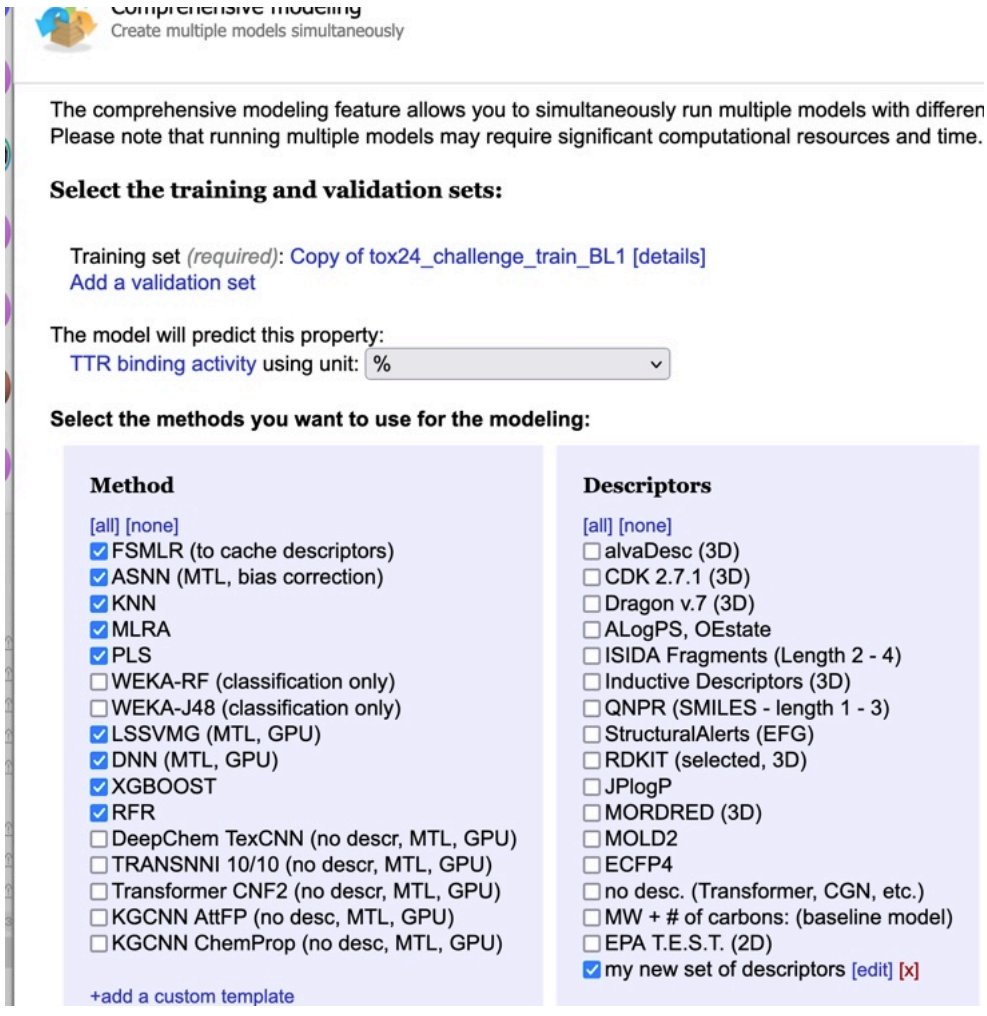
[all] [none]
- [ ] alvaDesc (3D)
- [x] CDK 2.7.1 (3D)
- [ ] Dragon v.7 (3D)
- [x] ALogPS, OEstate
- [ ] ISIDA Fragments (Length 2 - 4)
- [ ] Inductive Descriptors (3D)
- [ ] QNPR (SMILES - length 1 - 3)
- [ ] StructuralAlerts (EFG)
- [ ] RDKIT (selected, 3D)
- [ ] JPlogP
- [ ] MORDRED (3D)
- [ ] MOLD2
- [ ] ECFP4
- [x] no desc. (Transformer, CGN, etc.)
- [ ] MW + # of carbons: (baseline model)
- [ ] EPA T.E.S.T. (2D)

+add a custom template

Show advanced options>>




Model template editor
Advanced users only! Requires knowledge of OCHEM XML format
Model template name: my new set of descriptors
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ochem-model>
<properties/>
<attachment>
<configuration xsi:type="cdsConfiguration" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<descriptors>
<types>
<type>468 descriptors</type>
<configuration xsi:type="emptyConfiguration"/>
<markUncachedAsErrors>true</markUncachedAsErrors>
</types>
</descriptors>
<selection>
<numDifferentValues>2</numDifferentValues>
<correlationThreshold>0.95</correlationThreshold>
<useUFS>false</useUFS>
<stdThreshold>0.01</stdThreshold>
<maximumValueThreshold>999999</maximumValueThreshold>
</selection>
</configuration>
<optionsMapping/>
<protocol/>
<standartization/>
<datahandling>
<intervals>use</intervals>
<greaterless>use</greaterless>
<approximateequals>use</approximateequals>
<handleRanges>false</handleRanges>
</datahandling>
</attachment>
</ochem-model>


Create multiple models simultaneously
The comprehensive modeling feature allows you to simultaneously run multiple models with differen
Please note that running multiple models may require significant computational resources and time.
Select the training and validation sets:
Training set (required): Copy of tox24_challenge_train_BL1 [details]
Add a validation set
The model will predict this property:
TTR binding activity using unit: %
Select the methods you want to use for the modeling:
Method
[all] [none]
FSMLR (to cache descriptors)
ASNN (MTL, bias correction)
KNN
MLRA
PLS
WEKA-RF (classification only)
WEKA-J48 (classification only)
LSSVMG (MTL, GPU)
DNN (MTL, GPU)
XGBOOST
RFR
DeepChem TexCNN (no descr, MTL, GPU)
TRANSNNI 10/10 (no descr, MTL, GPU)
Transformer CNF2 (no descr, MTL, GPU)
KGCNN AttFP (no desc, MTL, GPU)
KGCNN ChemProp (no desc, MTL, GPU)
+add a custom template
Descriptors
[all] [none]
alvaDesc (3D)
CDK 2.7.1 (3D)
Dragon v.7 (3D)
ALogPS, OEstate
ISIDA Fragments (Length 2 - 4)
Inductive Descriptors (3D)
QNPR (SMILES - length 1 - 3)
StructuralAlerts (EFG)
RDKIT (selected, 3D)
JPlogP
MORDRED (3D)
MOLD2
ECFP4
no desc. (Transformer, CGN, etc.)
MW + # of carbons: (baseline model)
EPA T.E.S.T. (2D)
my new set of descriptors [edit] [x]