

Tutoriál

OCHEM

Pokročilá chemoinformatika
cvičení 3

OChem – databáze a modelovací nástroj

Online Chemical Modeling Environment

https://ochem.eu/home/show.do

Online chemical database
with modeling environment

v.2.3.3

log in create account

Home Database Models A+ a-

Welcome to OCHEM! Your possible actions

- Explore OCHEM data**
Search chemical and biological data: experimentally measured, published and exposed to public access by our users. You can also [upload your data](#).
- Create QSAR models**
Build QSAR models for predictions of chemical properties. The models can be based on the experimental data published in our database.
- Run predictions**
Apply one of the available models to predict property you are interested in for your set of compounds.
- Screen compounds with ToxAlerts**
Screen your compound libraries against structural alerts for such endpoints as mutagenicity, skin sensitization, aqueous toxicity, etc.
- Tutorials**
Check our video tutorials to know more about the OCHEM features.
- Our acknowledgements**

Feedback and help

Check out the properties available on OCHEM

OCHEM contains 1151032 experimental records for about 490 properties collected from 12539 sources

Melting Point **logPow** **logBB**
LogL(water) Cbrain/Cblood **LogD** Cblood/Cair Cbrain/Cair Cfat/Cair

C liver/Cair C muscle/Cair SIF solubility logPI(+) logPI(-) **logS**

LogL(blood) LogL(brain) LogL(fat) LogL(heart) LogL(kidney)

LogL(liver) LogL(lungs) LogL(muscle) LogL(oil) LogL(plasma)

LogBPR LogCSFPR ER fu(brain) P/Papp Biodistribution(kidney)

Biodistribution(liver) Biodistribution(lungs) Biodistribution(muscle) Biodistribution(heart)

Cbrain/Cplasma IC 50 Papp(Caco-2) Papp(MDCK)

P(brain) Oral absorption LIC 50 pK(1/logK) Cliver/Cplasma

Clung/Cplasma Cheart/Cplasma CKidney/Cplasma Cbrain/Cserum

Cfat/Cplasma Cmuscle/Cplasma Cskin/Cplasma Papp ratio(Caco-2)

Papp(MBUA) **Plasma protein binding** Papp(HPBEC)

Pendothelial(HPBEC) Papp(BBEC) Pendothelial(BBEC) Papp ratio(HPBEC)

Pendothelial ratio(HPBEC) Papp(SV-ARBEC) Pendothelial(SV-ARBEC)

Papp(MBEC4) Papp ratio(MDCKATCC) Pendothelial ratio(SV-ARBEC)

Latest active users

- burunduk:** Dr. Sofya Lushchekina about 1 hours ago
- nizamibilal1064:** Mr. Bilal Nizami about a day ago
- amaziz:** Mr. Ahmed Abdelaziz about a day ago
- MarcZimmermann:** Dr. Marc Zimmermann about a day ago
- novserj:** Mr. Sergii Novotarskyi about a day ago
- itetko:** Dr. Igor Tetko about a day ago

Latest published models

- LEL model** published by **novserj** several weeks ago
- LEL model** published by **itetko** about a month ago
- Ready biodegradability model** published by **svorberg** 7 months ago

OCHEM – přihlášení

Online Chemical Modeling Environment

https://ochem.eu/login/show.do

Online chemical database
with modeling environment

v.2.3.3

log in create account

Home Database Models A+ a-

Please, login

Instant login

In order to access OCHEM, you must login. If you do not wish to register now, you can login as a guest. Guest users have access to less features than registered users. You can also use your Facebook account to login.

[LOGIN AS A GUEST](#) [LOGIN WITH facebook](#)

Already have an account?

If you already have an account, please enter you login and password below:

Login ID

Password

[LOGIN](#) [PASSWORD REMINDER](#)

POSITIVE SSL SECURED WEBSITE
POSITIVE SSL
SECURED BY COMODO

Join OCHEM - register a new user!

Create a free account to upload data, create and apply QSAR models, screen chemical libraries and many more. Registered users can correct data uploaded by other registered users publish models. As a registered user, you can configure flexible access policies for your data and models.

OChem – výběr vlastnosti a tréninkové sady

Online Chemical Modeling Environment

https://ochem.eu//properties/show.do?render-mode=full

Online chemical database
with modeling environment

v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾

Area of your interest: [no tags selected \[change\]](#)

Compound properties

- Molecules
- Properties**
- Conditions

Units

Articles/Books

Journals

ToxAlerts

MatchedPairs

Baskets

Tags

Set area of interest...

User-related changes

My data exports

Batch data upload

Trash

search Properties from other users: Only approved properties ▾ Calculate counts

15 items on page 1 of 33

(Temperature / °C)	Show records	Melting Point corresponds to the phase transition temperaturu ...	itetko (moderator) ✉ / mojca ✉
(Dimensionless / Log unit)	Show records	The partition coefficient is a ratio of concentrations of u ...	itetko (moderator) ✉ / igor ✉
(Dimensionless / Log unit)	Show records	Measures blood-brain barrier (BBB) permeability values as l ...	mojca ✉ / moderated by Alessandra ✉
(Dimensionless / Log unit)	Show records	Measures partition of gases and vapors between the gas phas ...	mojca ✉ / unmoderated
(Dimensionless / Log unit)	Show records	Brain/blood concentration ratio	mojca ✉ / unmoderated
(Dimensionless / Log unit)	Show records	Kpu(brain) - Brain-to-unbound plasma partition coefficient. ...	mojca ✉ / unmoderated
(Dimensionless / Log unit)	Show records	The distribution coefficient of octanol/water measured at s ...	mojca ✉ / unmoderated
P(BMEC)	Show records	The in vitro BBB permeability was reported as the permeabil ...	mojca ✉ / unmoderated
Kpu(adipose)	Show records	Kpu(adipose) - Adipose-to-unbound plasma partition coeffici ...	mojca ✉ / unmoderated
Kpu(bone)	Show records	Kpu(bone)- Bone-to-unbound plasma partition coefficient. ...	mojca ✉ / unmoderated
Kpu(heart)	Show records	Kpu(heart) - Heart-to-unbound plasma partition coefficient. ...	mojca ✉ / unmoderated
Kpu(kidney)	Show records	Kpu(kidney) - Kidney-to-unbound plasma partition coefficien ...	mojca ✉ / unmoderated

OChem – výběr vlastnosti a tréninkové sady

Online Chemical Modeling Environment

https://ochem.eu//properties/show.do?render-mode=full

Online chemical database
with modeling environment

Welcome, Guest! [Logout](#)

Home Database Models

Property browser X Filtered records X

Compounds properties browser

Search for numerical compounds properties linked to scientific articles

Area of your interest: no tags selected [\[change\]](#)

FILTERS

▼ SOURCE
Article/Source [select]
Page Table

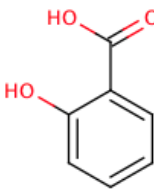
▼ PROPERTY
Activity/Property [select]
LogD
 Hide records without property

► CONDITIONS

▼ MOLECULE FILTERS
Name / OCHEM ID / Inchi-Key
Similarity/substructure search
Draw a structure and search all the molecules containing it or similar to it

CLICK TO DRAW

Basket Records 1 - 5 of 667 5 items on page 1 of 134

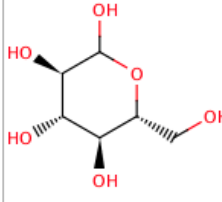

molecule profile

LogD = -2.14 (in Log unit)
pH = 7.4 Log unit

Ungell, AL et al
Membrane transport of drugs in different regions of the inte...
P: 361 T: 1
J Pharm Sci 1998; 87 (3) 360-6

Salicylic acid
MoleculeID: M13520
Public record

RecordID: R1969300
17:04, 31 Oct 11 / 21:14, 5 Nov 11
charochkina


molecule profile


LogD = -3.0 (in Log unit)
pH = 7.4 Log unit

Ungell, AL et al
Membrane transport of drugs in different regions of the inte...
P: 361 T: 1
J Pharm Sci 1998; 87 (3) 360-6

D-glucose
MoleculeID: M12533
Public record

RecordID: R1969299
17:04, 31 Oct 11 / 21:14, 5 Nov 11
charochkina

OChem – příprava modelu



Online chemical databases
with modeling environment

Home ▾ Database ▾ **Models ▾**

Create a model
Apply a model
Create
Open p

Create a model ⓘ
Select the training and validation sets, the machine learning method and the validation protocol


Create a model ⓘ
Select the training and validation sets, the machine learning method and the validation protocol

Select the training and validation sets:




Training set (required): [...]

[Add a validation set](#)

Basket browser ⓘ
Browse, Compare or Join molecule set

Filter by name: [Create new ] Show public sets

1 - 1 of 1

  	Selected records created by	667 records
---	------------------------------------	-------------

1 - 1 of 1

OChem – výběr metody

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment

v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Training set (required): Selected records [details]
[Add a validation set](#)

The model will predict this property:
LogD using unit:

Choose the learning method: ⓘ

Suggested modeling methods:

- ASNN (ASsociative Neural Networks)
- FSMLR (Fast Stagewise Multiple Linear Regression)
- KNN (K-Nearest Neighbors)
- Library model (A model based on another ASNN model enriched with new compounds data)
- LibSVM wrapper with grid-search parameter optimisation
- MLR (Multiple Linear Regression)
- PLS (Partial Least Square)
- WEKA-J48 (Weka-based implementation of C4.5 decision tree)
- KNN (Weka implementation)
- LADTree (Weka implementation)
- Naive Bayes (Weka implementation)
- REPTree (Weka implementation)
- WEKA-RF (Weka-based implementation of Random Forest)

Models under development. (Do not use unless you are sure how to use):

- Consensus model (experimental)

Model validation

Validation method:

Number of folds:

Stratified cross-validation

You can create a model from template: [import an XML model template](#) or [use another model as a template](#)

OChem – preprocessing

The screenshot shows a web browser window titled "Online Chemical Modeling Environment". The address bar displays "https://ochem.eu/modelconfigurator/choose.do". The page header features the "Online chemical database with modeling environment" logo and version "v.2.3.3". A navigation menu includes "Home", "Database", and "Models". The main content area is titled "Model creator" and contains a section for "Select the preferred data preprocessing options". Under "Preprocessing of molecules (Chemaxon)", four options are checked: Standardization, Neutralize, Remove salts, and Clean structure. A section for "Records with ranges" includes an unchecked option to "Include 'greater' and 'less' records (12 records)". Navigation buttons for "<<Back" and "Next>>" are at the bottom.

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment

v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Model creator
Select model template and training set

Select the preferred data preprocessing options

Preprocessing of molecules (Chemaxon) ⓘ

- Standardization
- Neutralize
- Remove salts
- Clean structure

Records with ranges
Include following records:

- Include "greater" and "less" records (12 records)

<<Back Next>>

OChem – výběr deskriptorů

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

v.2.3.3

Model creator

Select model template and training set

Select the molecular descriptors ?

Recommended descriptor types

E-state

E-State types:

<input checked="" type="checkbox"/> Atom indices	<input type="checkbox"/> Atom counts
<input type="checkbox"/> Bonds indices	<input type="checkbox"/> Bonds counts
<input type="checkbox"/> Extended indices (experimental)	

Aromatize structures:

ALogPS (2)

GSFfragment (1138)

Dragon v. 6.0 (4885/3D) *Not supported by your installation*

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)

Predictions by OCHEM's featured models ?

ALogPS 3.0

Melting Point - 3D (Dragon 6 + Corina)

Melting Point - 2D (ALOGPS 2.01 +OESTATE)

Toxicity against T. Pyriformis

Ames levenberg

CYP3A4 Estate+ALogPS

CYP2D6 Estate+ALogPS

CYP2C19 Estate+ALogPS

CYP2C9 Estate+ALogPS

CYP1A2 Estate+ALogPS

Outputs of other OCHEM models

Additional descriptor types

OESate

MolPrint

Dragon v. 5.4 (1630/3D)

Dragon v. 5.5 (3190/3D)

Structural alerts (ToxAlerts)

MOPAC descriptors (21/3D) *Not supported by your installation*

OChem – výběr deskriptorů 2

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment

v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

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

- AMBIT Descriptors
- ISIDA fragments (2011) *Not supported by your installation*
- Scaffold Hunter Descriptors
- ECFP Fingerprints
- Chemaxon Scaffolds
- Silicos-It Scaffolds
- Experimental values of other properties *Not supported by your installation*

Conditions of experiments

- pH
- logPow Method
- logPow Buffer
- Temperature
- Ionic strength

Forbid NaN and Infinite descriptor values


Force recalculation of descriptors in cache (use with caution)

<<Back Next>>

OChem – optimalizace struktury

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

 **Online chemical database**
with modeling environment v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Model creator
Select model template and training set

Select a tool to optimize molecule structures

- No optimisation
- Optimise with Corina
- Optimise with MOPAC without stereochemistry
- Optimise with MOPAC with stereochemistry

OChem – filtry deskriptorů

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment

v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Model creator

Select model template and training set

Select filters of descriptors

- Eliminate descriptors with less than unique values
- Delete descriptors that have absolute values larger than
- Delete descriptors that have variance smaller than
- Group descriptors, that have pair-wise correlations Pearson's correlation coefficient R larger than
- Use Unsupervised Forward Selection to delete variables using the above value of multiple correlation coefficient R
- Perform principal component analysis
- After filtering, I want to select necessary descriptors myself (advanced)


OChem – spuštění výpočtu



Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Reader

 **Online chemical database**
with modeling environment v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Model creator
Select model template and training set

Start calculation of the model

Now we are ready to start calculation.
Please provide the name for your model:

Save models

Task priority:

High priority (please, use for fast tasks only)

Normal priority

Low priority

Large task priority (for long tasks)

OChem – ...

The screenshot shows a web browser window titled "Online Chemical Modeling Environment". The address bar contains the URL "https://ochem.eu/modelconfigurator/choose.do". The page header features the "Online chemical database with modeling environment" logo and text, along with the version "v.2.3.3" and a "Welcome, Guest! Logout" message. A navigation menu includes "Home", "Database", and "Models". The main content area is titled "Model creator" with the instruction "Select model template and training set". Below this, a section titled "Run model builder" displays a loading spinner and the text "Model training - Task started", with links for "[cancel]" and "[fetch result later]". At the bottom, there are two buttons: "<<Back" and "Next>>".

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment


v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Model creator
Select model template and training set

Run model builder


Model training - Task started
[\[cancel\]](#) [\[fetch result later\]](#)

<<Back Next>>

OChem – výsledky

Online Chemical Modeling Environment

https://ochem.eu/modelconfigurator/choose.do

Online chemical database
with modeling environment

v.2.3.3

Welcome, Guest! [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a-

Save the model

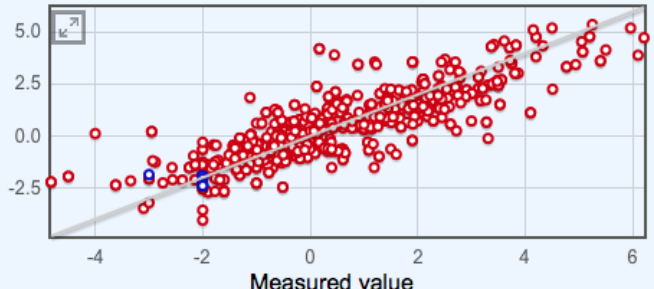
Please enter your model's name:

Overview | Applicability domain

Model name: LogD_ASNN_[Adriana, ChemaxonDescriptors (7.4), ALogPS, Dragon55 (blocks: 1-22)],
213859 [\[rename\]](#)
Private ID is 31481910

Predicted property: **LogD**
Training method: ASNN

Data Set	#	R2	q2	RMSE	MAE
Training set: Selected records	646 records	0.71 ± 0.02	0.7 ± 0.02	1.01 ± 0.04	0.77 ± 0.02
Excluded from training set	12 records	0.2 ± 0.2	0 ± 0	0.4 ± 0.1	0.27 ± 0.09



Number of compounds ignored because of errors in original model = 9

[Adriana, ChemaxonDescriptors (7.4), ALogPS, Dragon55 (blocks: 1-22)]
Correl. limit: 0.95 Variance threshold: 0.01,
Maximum value: 999999,
Supersab, 1000 iterations, 3 neurons
ensemble=64 additional param
PARTITION=3, SELECTION=2
5-fold cross-validation

1221 pre-filtered descriptors
Supersab, 1000 iterations, 3 neurons
ensemble=64 k=32 additional param
PARTITION=3, SELECTION=2

Calculated in 3186 seconds
Size: 659 Kb

Úkoly

① Vyberte si jednu metodu přípravy QSAR/QSPR modelu:

① ASNN

② MLR

③ KNN

a pomocí OChemu připravte podle této metody model predikující teplotu bodu tání. Porovnejte s uloženými modely.

② Připravte model predikující hodnotu pKa a diskutujte kvality modelu pomocí R², RMSE a MAE.

③ Porovnejte výsledky validace pomocí tréninkové + testovací sady a „cross“ validace.