

Pharmaceutical Chemistry

QSAR through the OCHEM portal



SAPIENZA
UNIVERSITÀ DI ROMA

Introduction

J Comput Aided Mol Des (2011) 25:533–554
DOI 10.1007/s10822-011-9440-2

Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information

The Online Chemical Modeling Environment is a web-based platform that aims to automate and simplify the typical steps required for QSAR modeling. The platform consists of two major subsystems: the database of experimental measurements and the modeling framework. A user-contributed database contains a set of tools for easy input, search and modification of thousands of records. The OCHEM database is based on the wiki principle and focuses primarily on the quality and verifiability of the data. The database is tightly integrated with the modeling framework, which supports all the steps required to create a predictive model: data search, calculation and selection of a vast variety of molecular descriptors, application of machine learning methods, validation, analysis of the model and assessment of the applicability domain. As compared to other similar systems, OCHEM is not intended to re-implement the existing tools or models but rather to invite the original authors to contribute their results, make them publicly available, share them with other users and to become members of the growing research community. Our intention is to make OCHEM a widely used platform to perform the QSPR/QSAR studies online and share it with other users on the Web. The ultimate goal of OCHEM is collecting all possible chemoinformatics tools within one simple, reliable and user-friendly resource. The OCHEM is free for web users and it is available online at <http://www.ochem.eu>

Pagina 2

In this tutorial you will be guided to use OCHE.eu platform to build QSAR models focusing in the development of PLS-based models

Preparing the dataset

Structure-Based Design of 6-Chloro-4-aminquinazoline-2-carboxamide Derivatives as Potent and Selective p21-Activated Kinase 4 (PAK4) Inhibitors | Journal of Medicinal Chem...

RETURN TO ISSUE | PREVIOUS | ARTICLE | NEXT

Structure-Based Design of 6-Chloro-4-aminquinazoline-2-carboxamide Derivatives as Potent and Selective p21-Activated Kinase 4 (PAK4) Inhibitors

Shenhuo Mao¹, Fan Zhao¹, Hongyan Song¹, Jing Guo¹, Xiaojing Ji¹, Feifei Jiang¹, Fan Huo¹, Shun Song¹, Geqing Zhang¹, Ruying Wang¹, He Wang¹, Yu Zhang¹, Tingting Liu¹, Tang Liu¹, Xianhui Wang¹, Yan Qiu¹, He Liu¹, Zhongguo Mei¹, Huihui Li¹, Feng Li¹, Dongmin Chen¹

View Author Information

© Clive W. J. | ACS Chem. 2018, 8(1), 282-291
Publication Date November 30, 2017
https://doi.org/10.1021/acs.jmedchem.7b01342
Copyright © 2017 American Chemical Society

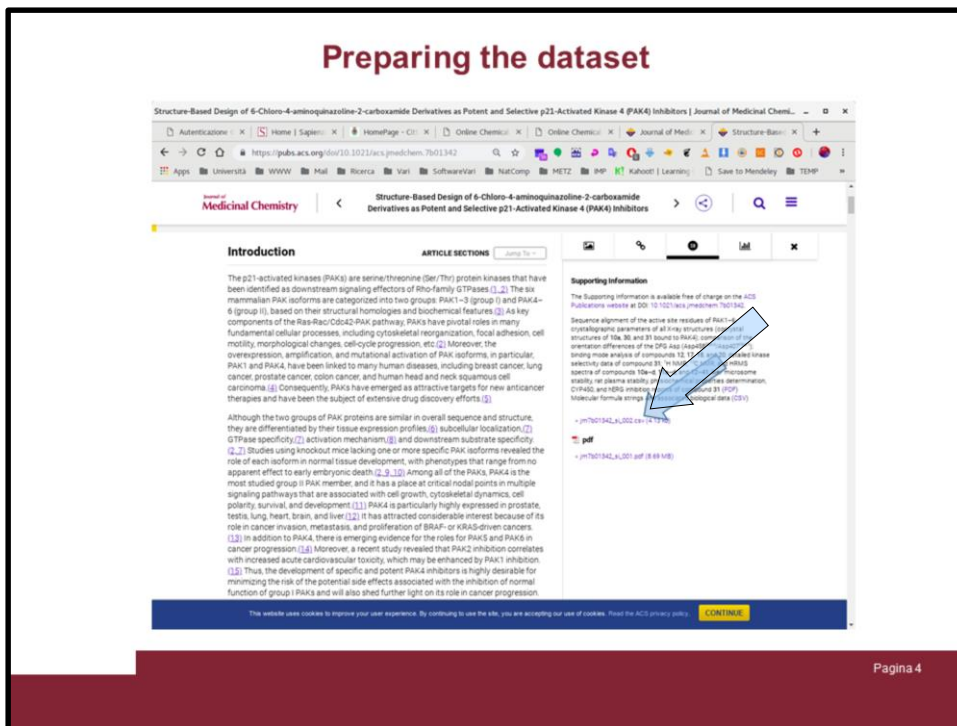
Supporting Information (SI) icon

Abstract

Herein, we report the discovery and characterization of a novel class of PAK4 inhibitors with a quinazoline scaffold. Based on the shape and chemical composition of the ATP-binding pocket of PAK4, we chose a 4-aminquinazoline series of inhibitors as a starting point. Guided by X-ray crystallography and a structure-based drug design (SBDD) approach, a series of novel 4-aminquinazoline-2-carboxamide PAK4 inhibitors were designed and synthesized. The inhibitors' selectivity, therapeutic potency, and pharmaceutical properties were optimized. One of the best compounds, **31** (ZB26), showed remarkable PAK4 selectivity (348-fold vs PAK1) and favorable kinase selectivity profile. Moreover, this compound potently inhibited the migration and invasion of A549 tumor cells by regulating the PAK4-directed downstream signaling pathways *in vitro*. Taken together, these data support the further development of **31** as a lead compound for PAK4-targeted anticancer drug discovery and as a valuable research probe for the further biological investigations of group II PAKs.

First you need a dataset of molecules, as example the molecules available from the J Med Chem publication on PAK4 inhibitor will be retrieved from the article web page.

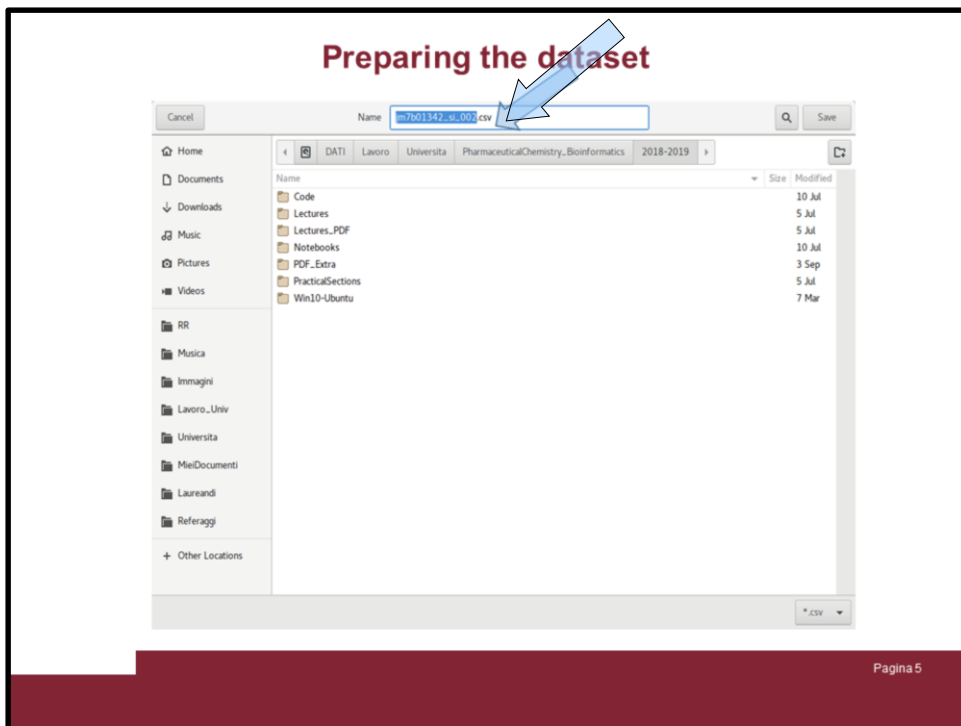
To do this go to pubs.acs.org/doi/abs/10.1021/acs.jmedchem.7b01342 and click on the supporting info "SI" icon



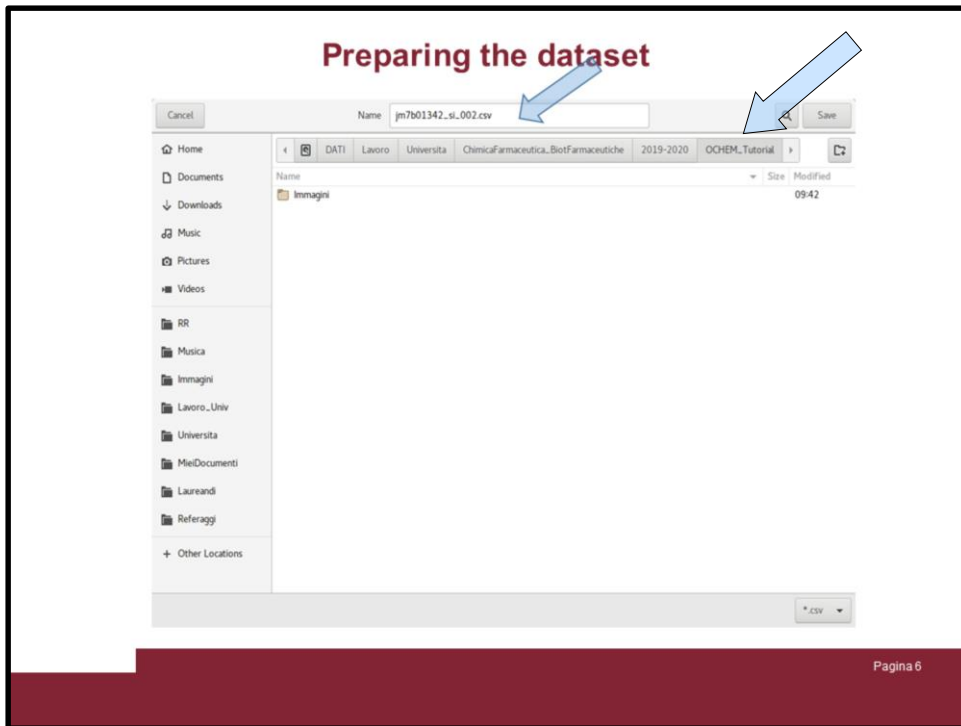
.... the browser will go to the page (in different way depend on the browser you use) and it will present the link to download a comma separated variable (csv) file:

jm7b01342_si_002.csv

Click on it and the file will be downloaded on your PC (Take a note were it is going to be saved)

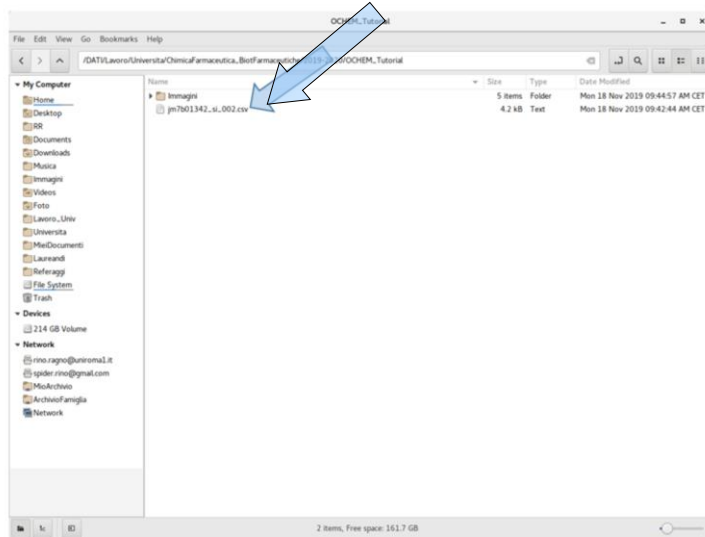


If your browser was set to ask where to save the under downloading files you will be asked for a folder to save the csv file



Here a new folder callse “OCHEM_Tutorial” was prepared to save the csv file

Preparing the dataset

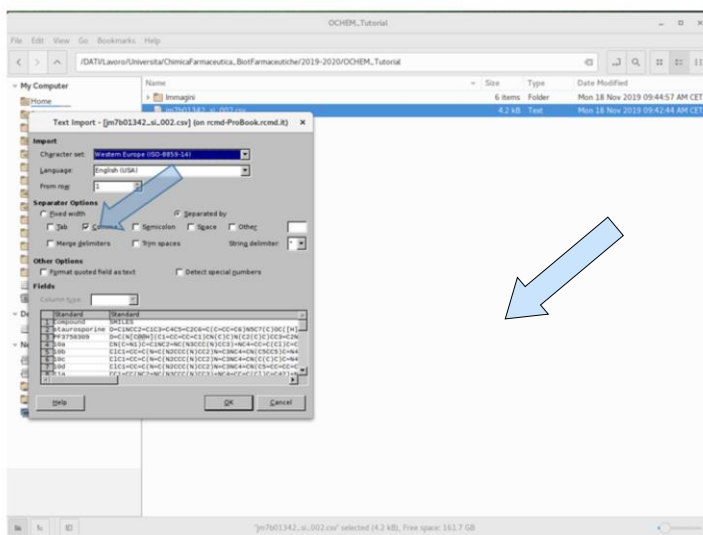


Pagina 7

ALWAYS CHECK THE FILE IS THERE!

Now right click on the csv file and open it with libreoffice or import in excel.
Here libreoffice was used

Preparing the dataset



Pagina 8

A dialog window will pop up asking how the file should be interpreted. Check comma as separator and uncheck all the other options. You have an example on the windows how the data will be read. Pay attention on the way the number will be interpreted (Italian or International)

Preparing the dataset

Compound	SMILES
staurosporine	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC([H])(C[C@H](NC)(C@@H)7O</chem>
PF3758309	<chem>O=C(N[C@@H](C1=CC=CC=C1)C(N(C)C)N(C2(C)C)CC3=C2NN=C3NC4=NC(C)=NC5=C4</chem>
10a	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42</chem>
10b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C1</chem>
10c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C(C)C)C=N4)C3=C1</chem>
10d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N4)C3=C1</chem>
11a	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42)=NN1</chem>
11b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5CC5)=C4)C3=C1</chem>
11c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C(C)C)=C4)C3=C1</chem>
11d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5=CC=CC=C5)C=C4)C3=C1</chem>
12	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
13	<chem>CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
14	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
15	<chem>CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
16	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
17	<chem>CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
18	<chem>CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
19	<chem>CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
20	<chem>C1C1=CC=C(N=C(N2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=C1</chem>
21	<chem>CC1=CC(NC2=NC(C(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
22	<chem>CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>
23	<chem>CC1=CC(NC2=NC(C(N(C)C3CNCC3)=O)=NC4=CC=C(C=C42)C)=NN1</chem>

Pagina 9

Once the file is imported in libreoffice you can check it is correctly read.
Slide on the right ...

Preparing the dataset

	MW	PAK1 Inhibitory Activity (Ki, nM)	PAK4 Inhibitory Activity (Ki, nM)	Tmax-T0 (PAK4, FC)	clogP	PSA
1	466	0.003	0.009ND			3.7 69.8
2	490	0.052	0.026ND			4.3 98.3
3	357.5>10		0.710.50.0			2.6 71.3
4	383.5ND		0.6520.50.0			3 70.9
5	385.5ND		0.8840.80.3			3.2 70.4
6	419.5ND		1.440.00.3			3.8 71
7	357.5	1.891	0.0992.10.0			2.3 86.4
8	383.5	0.288	0.0164.70.3			2.9 85
9	385.5ND		0.0772.30.3			3.1 84.8
10	419.5ND		0.2560.70.3			3.5 85.1
11	385.5ND		0.3540.50.0			1.7 108.9
12	386.5ND		3.8170.10.3			2.1 107
13	399.5ND		0.6740.40.3			2.7 104.3
14	400.5ND		3.89-0.10.0			2.7 101
15	399.5ND		0.6720.30.3			2.7 101
16	371.5>10		0.0452.40.5			1.6 98.7
17	372.5ND		2.1720.20.3			2.2 92.5
18	385.5ND		0.768ND			2 88.9
19	385.5ND		2.408ND			0.8 132.2
20	399.5ND		0.5760.20.3			2.7 95.7
21	385.5ND		0.3250.20.3			2.4 108.8
22	399.5ND		1.630.00.3			2.7 95.6
23	357.5ND		0.7820.70.3			2.8 111.4
24	397.5	2.75	0.0164.10.3			2.2 97.4
25	385.5>4.52		0.2181.50.3			2.4 91.4
26	411.5	0.383	0.0263.60.3			3 90.1
27	385.5>4.52		0.1511.80.3			1.9 92.5
28	411.5		0.0173.60.3			3.6 61.1

... and focus on the PAK4 Ki column

Preparing the dataset

The screenshot shows a spreadsheet with the following data:

MW	PAK1 Inhibitory Activity (Ki, ?M)	Ki	Tmax-T0 (PAK4, FC) clogP	PSA
466	0.003	0.009ND	3.7	69.8
490	0.052	0.026ND	4.3	98.3
357.5-10		0.710.50.0	2.6	71.3
383.5ND		0.6520.50.0	3	70.9
385.5ND		0.8820.80.3	3.2	70.4
419.5ND		1.440.00.3	3.8	71
357.5	1.891	0.0992.10.0	2.3	86.4
383.5	0.288	0.0164.70.3	2.9	85
385.5ND		0.0772.30.3	3.1	84.8
419.5ND		0.2560.70.3	3.5	85.1
385.5ND		0.3540.50.0	1.7	108.9
386.5ND		3.8170.10.3	2.1	107
399.5ND		0.6740.40.3	2	104.3
400.5ND		3.89-0.10.0	2	101
399.5ND		0.6720.30.3	2	97
371.5-10		0.0452.40.5	1.6	98.7
372.5ND		2.1720.20.3	2.2	92.5
385.5ND		0.768ND	2	88.9
385.5ND		2.408ND	0.8	132.2
399.5ND		0.5760.20.3	2.7	95.7
385.5ND		0.3250.20.3	2.4	108.8
399.5ND		1.630.00.3	2.7	95.6
357.5ND		0.7820.70.3	2.8	111.4
397.5	2.75	0.0164.10.3	2.2	97.4
385.5-4.52		0.2181.50.3	2.4	91.4
411.5	0.383	0.0263.60.3	3	90.1
385.5-4.52		0.1511.80.3	1.9	92.5
411.5	0.700	0.0173.60.3	2.6	91.1

Rename the column into just Ki and ...

Preparing the dataset

RECORDID	SMILES	Ki
1	staurosporine	0.009
2	PF3758309	0.026
3	10a	0.71
4	10b	0.652
5	10c	0.884
6	10d	1.44
7	11a	0.099
8	11b	0.016
9	11c	0.07
10	11d	0.256
11	12	0.554
12	13	0.617
13	14	0.674
14	15	3.89
15	16	0.672
16	17	0.045
17	18	2.172
18	19	0.768
19	20	2.408
20	21	0.576
21	22	0.325
22	23	1.63
23	24	0.782
24	25	0.016
25	26	0.218
26	27	0.026
27	28	0.151
28	29	0.117

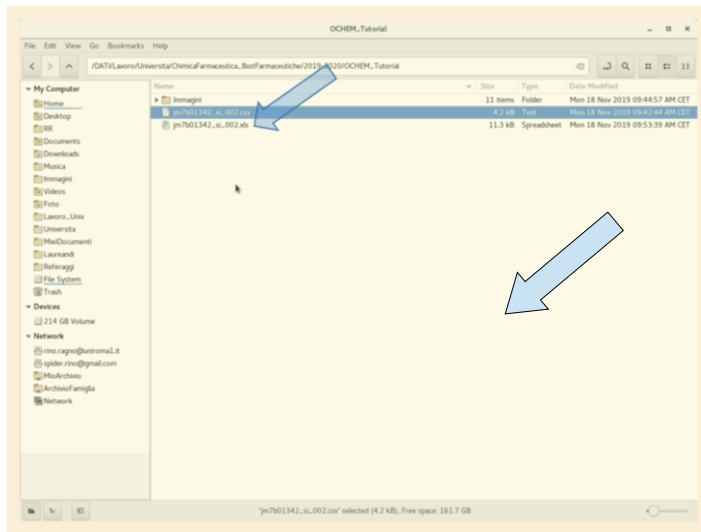
Pagina 12

Delete the other columns. Rename the compound column header into RECORDID.

You need just three columns: RECORDID, SMILES and Ki

Save the file as excel type, the one with .xls extension is only valid

Preparing the dataset



Pagina 13

CHECK THE XLS FILE IS THERE!

Preparing the dataset

The screenshot shows the OChem.eu website interface. The browser address bar displays 'https://ochem.eu/home/show.do'. The page title is 'Online chemical database with modeling environment'. The navigation menu includes 'Home', 'Database', and 'Models'. The main content area is divided into three columns:

- 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.
 - Optimise your molecules:** Optimise different properties for your molecules (e.g., reduce their toxicity or improve their ADME properties) using the state-of-the-art MolOptimizer utility based on matched molecular pairs.
 - Tutorials:**
- Check out the properties available on OCHEM:**
 - OCHEM contains 2854383 records for 638 properties (with at least 50 records) collected from 12973 sources.
 - Melting Point** $\log\text{Pow}$ $\log\text{BB}$ $\log\text{D}$ $\log\text{P}(\text{H}^+)$
 - Water solubility** $\log\text{L}(\text{water})$ $\log\text{L}(\text{air})$ ER
 - Ceramic/Plasma** IC50
 - Papp(Caco-2)** **Papp(MDCK)**
 - Oral absorption** LIC 50
 - Plasma protein binding** $\text{Papp}(\text{rate})(\text{Caco-2})$ $\text{Papp}(\text{rate})(\text{MDCK-mt1})$ pIC50
 - %Human FA Human IA
 - Human FA
 - fraction unbound (fu)
- Latest active users:**
 - rino.ragno: Prof. Rino Ragno seconds ago
 - Xinggguomeng: Miss. guomeng xing seconds ago
 - Ykovalishyn: Dr. Vasyil Kovalishyn seconds ago
 - actimidy: M. Ely Setiawan seconds ago
 - Corde: Mr. Jose Andres Cordero Solano seconds ago
 - zaini1: Mrs. Zaini Rehman about an hour ago
- Latest published models:**
 - IC50 model published by carpoovr 1 hours ago
 - Li50 mouse oral model published by Tenkov_Oleg 9 months ago
 - Drug-Induced Rhabdomyolysis model published by qingshuang5501 9 months ago
 - o_oso_ort_LD model published by

Pagina 14

Go to OChem.eu and click onto "Create account" if you do not have any

Preparing the dataset

Online Chemical Modelling Environment - Google Chrome

https://ochem.eu/user/newuser.do

Online chemical database
with modeling environments

Home Database Models

log in create account
A+ Privacy statement

User account
Details of your personal OCHEM account

Registration Information

Login* (min. 4 characters and max. 20 characters)

e-mail*

Password* Password can contain only letters and numbers.

Confirm password*

Personal Information

Title* Please, select your title

First name*

Last name*

Affiliation*

Form of organization* Please, select your organization type

City

pn7601342...u.d...cw

Pagina 15

Fill in all the data ...

Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

https://ochem.eu/user/newuser.do

Online chemical database
with modeling environment

log in create account

Home Database Models

Country

Zip

Phone

Position

Web

Terms of Service

Online License Agreement and Terms of Use

This is a legal agreement between you (the "USER"), and Helmholtz Zentrum Muenchen - Deutsches Forschungszentrum fuer Gesundheit und Umwelt ("HMGU"), covering your use of the Online Chemical Database & Modeling Environment and the accompanying Software ("OCHEM"). Be sure to read the following agreement before using OCHEM.

HMGU is willing to license the use of OCHEM upon the condition that you accept all of the terms contained in this License Agreement and any supplementary license terms included herewith ("Agreement")

By using OCHEM, you agree to read and be the terms of this Agreement. If you do not agree to the terms of this...

By clicking on "accept below" I acknowledge that I have read and fully understand the foregoing information and agree to abide by License agreement above and the Privacy Policy.

Pagina 16

.... and then click on "I ACCEPT CREATE MY ACCOUNT"

Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Home Database Models

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.

[LOGIN AS A GUEST](#)

Already have an account?
If you already have an account, please input your login and password below:

Login ID

Password

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

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.

[REGISTER A NEW USER](#)

log in create account

Privacy statement

Pagina 17

Once the OCHE account is activated log in in the portal using the username and password you indicated during the registration

Preparing the dataset

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.

Optimise your molecules
Optimise different properties for your molecules (e.g., reduce their toxicity or improve their ADME properties) using the state-of-the-art MolOptimiser utility based on matched molecular pairs

Tutorials

Check out the properties available on OCHEM

OCHEM contains 2854383 records for 638 properties (with at least 50 records) collected from 12973 sources

Melting Point
logPow logBB
Log_(water) LogD log_(H⁺)

Water solubility
Log_(blood) Log_(oil) ER
C_{strat}/C_{plasma} IC50
Papp(Caco-2) Papp(MDCK)
Oral absorption LIC 50
Papp rat(Caco-2)

Plasma protein binding
Papp rat(MDCK-m-d1) pIC50
%Human FA Human IA
Human FA
fraction unbound (fu)

Latest active users

- rino ragno: Prof. Rino Ragno seconds ago
- Vikovalshyn: Dr. Vasyi Kovalishyn seconds ago
- corde: Mr. Jose Andres Cordero Solano seconds ago
- Xingguoming: Miss. guoming xing 2 minutes ago
- schensby: Mr. Ely Seliswan 8 minutes ago
- zain1: Mrs. Zaina Rehman about 1 hours ago

Latest published models

- IC50 model published by carpospy 1 hours ago
- L150 mouse oral model published by Trifkov_Oleg 9 months ago
- Drug-Induced Rhabdomyolysis model published by qingshuang501 9 months ago
- o_oss of: LD model published by

Pagina 18

Now you have logged into OCHEM

Preparing the dataset

The screenshot shows the Online Chemical Modeling Environment (OCHEM) website. The page is titled "Online chemical database with modeling environment". A blue arrow points to the "Database" tab in the top navigation bar. Another blue arrow points to the "Batch data upload" option in the "Compound properties" section on the left. The main content area displays various chemical properties and models, including "Melting Point", "logPow", "logBB", "LogD", "Water solubility", "Oral absorption", and "Plasma protein binding". The right sidebar shows "Latest active users" and "Latest published models".

Pagina 19

Now let's to import the xls file.

Click on the Database tab and select "Batch data upload"

Preparing the dataset

Online Chemical Database
with web/ling environment

Welcome, Dear Prof Ragnon [My account](#) [Logout](#)

Home Database Models

Batch Upload 3.0 - File selection

Select the CSV, SDF or Excel file to upload multiple records to the database.

Instructions

Here you have the possibility to upload data from an excel sheet, sdf or csv .
Backslash \ is used as stereochemistry in cvs, which should not contain \uffff characters.
The format of these data is strict, and can be viewed in [this sample](#) (scientific format) and [this sample](#) (technical format).
For more information, consider the [wiki page](#) that you can access by clicking on the wiki icon next to the title ("Batch upload browser").
If you have difficulty preparing your data, feel free to drop us an e-mail at info@ochem.eu.

Select a file to upload

Upload file
Choose File No file chosen

Settings

- Allow molecule lookup by name on PubChem
- Allow article lookup by PMID on PubMed
- Make the uploaded records hidden

[Upload](#)

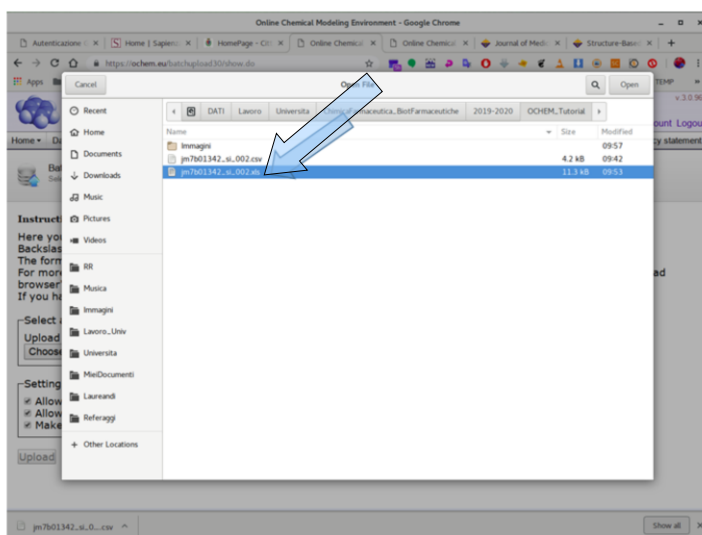
js7601342_u-0_cw

[Show all](#)

Pagina 20

A window will give you some direction. Click on “ChooseFile”

Preparing the dataset



Pagina 21

A window will pop up and ask you to select the file you prepared before. Just click on it twice.

Preparing the dataset

Online Chemical Database
with modeling environments

Welcome, Dear Prof.Ragnor [My account](#) [Logout](#)

Home • Database • Models • [Privacy statement](#)

Batch Upload 3.0 - File selection

Select the CSV, SDF or Excel file to upload multiple records to the database

Instructions
Here you have the possibility to upload data from an excel sheet, sdf or csv .
Backslash \) is used as stereochemistry in cvs, which should not contain \u00ff characters.
The format of these data is strict, and can be viewed in this sample (scientific format) and this sample (technical format).
For more information, consider the wiki page that you can access by clicking on the wiki icon next to the title ("Batch upload browser").
If you have difficulties uploading your data, feel free to drop us an e-mail at info@ochem.eu.

Select a file to upload

Upload file
Choose File | jm7b01342_002.xls

Settings

- Allow molecules to be added by name on PubChem
- Allow molecules to be added by PMID on PubMed
- Make inactive records hidden

Upload

jm7b01342_u-0_csw [Show all](#)

Pagina 22

Now click on "Upload" button

Preparing the dataset

Batch Upload 3.0 - File preview and column remapping

Preview your data, select the sheet and the columns you would like to upload

jm7b01342_sl_002

RECORDID	SMILES	KI
staurosporine	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=C=C6)N5C...</chem>	0.009000000000000001
PF3758309	<chem>O=C(N[C@@H](C1=CC=CC=C1)CN(C)N)(C2(C...</chem>	0.026000000000000002
10a	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=C...</chem>	0.71
10b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=CN(...</chem>	0.652
10c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=CN(...</chem>	0.884
10d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=CN(...</chem>	1.44
11a	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3))=NC4=CC=C(C...</chem>	0.099
11b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=NNC...</chem>	0.016
11c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=NNC...</chem>	0.077
11d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=NNC...</chem>	0.256

The ARTICLE column is missing, the stub unpublished article will be assigned by default

Green titles indicate recognized columns, red titles indicate errors. Please click on the red columns and select whether the column indicates a property, condition or another column type like name, value or molecule, then select the matching entity and confirm your selection by clicking on the green button on the left.

If you have irrelevant columns in your sheet, you can leave them red and they will be ignored in the further process. If you

jm7b01342_sl_0_csw Show all

Pagina 23

You will have a preview on the data under uploading, slide down ...

Preparing the dataset

Online chemical database
with modeling environment

Welcome, Dear Prof Ragnol [My account](#) [Logout](#)

RECORDID	SMILES	KI
staurosporine	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C...</chem>	0.009000000000000001
PF3758309	<chem>O=C(N[C@@H](C1=CC=CC=C1)CN(C)C)N(C2(C...</chem>	0.026000000000000002
10a	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=C...</chem>	0.71
10b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(...</chem>	0.652
10c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(...</chem>	0.884
10d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(...</chem>	1.44
11a	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C...</chem>	0.099
11b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC...</chem>	0.016
11c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC...</chem>	0.077
11d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC...</chem>	0.256

The **ARTICLE** column is missing, the sheet published article will be assigned by default

Green titles indicate recognized columns, red titles indicate errors. Please click on the red columns and select whether the column indicates a property, condition or another column type like name, value or molecule, then select the matching entity and confirm your selection by clicking on the green button on the left.

If you have irrelevant columns in your sheet, you can leave them red and they will be ignored in the further process. If you need help, feel free to drop us an e-mail at info@ochem.eu.

[Upload this sheet](#)

[Cancel Batch Upload](#) [Download Excel file](#)

Pagina 24

And click on “Upload this sheet”

Preparing the dataset

Batch Upload 3.0 - Entity remapping
Review and remap the properties, conditions, articles and baskets involved in the data upload

Database entities remapping

Property: [Ki](#)

Values: [-log\(M\)](#)
Unit: [-log\(M\); min value: 0.006, max value: 3.89](#)

Article: [unpublished](#)

Molecule set: [default](#)

[submit](#)

[Cancel Batch Upload](#) [Download Excel file](#)

Pagina 25

Next the portal will show some info on the data, pay attention on how the activity potency (IC50 or Ki) will be interpreted, here the program suggest Kis as $\log(M)$. You know it is not true so it have to be fixed, therefore click on the “ $\log(M)$ ” link ...

Preparing the dataset

The screenshot shows the 'Online chemical database' website. The page title is 'Online chemical database with modeling environment'. The user is logged in as 'Prof. Ragnol'. The main content is a table of units, with columns for the unit symbol, its description, and a 'Show records' link. A blue arrow points to the 'uM' unit in the list.

Unit	Description	Link
pM	(Concentration) Show records 1 pM (picomolar) corresponds ...	Parika_OCHEM
ppb	(Concentration) Show records parts per billion Denotes one ...	bhuhaha
ppb food	(Concentration) Show records in food - parts per billion De ...	chanchima
pph	(Concentration) Show records parts per hundred, one molecu ...	bhuhaha
ppm	(Concentration) Show records Parts per million (ppm) deno ...	ferko
ppm food	(Concentration) Show records in food - Parts per million (...)	chanchima
ppt	(Concentration) Show records Parts per trillion (ppt) den ...	indylol
ug	(Concentration) Show records ug/ml -- ug stands for micro g ...	Parika_OCHEM
ugbee	(Concentration) Show records The quantity of substance in e ...	meza
ug/g	(Concentration) Show records The unit to measure concentrat ...	chanchima
ug/kg	(Concentration) Show records The unit to measure concentrat ...	chanchima
ug/kg food	(Concentration) Show records The unit to measure concentrat ...	chanchima
ug/L	(Concentration) Show records ug/l -- micrograms per liter, ...	chanchima
ug/Lid	(Concentration) Show records Concentration in micrograms (u ...	chanchima
ugm3	(Concentration) Show records Microgram per cubic meter unit ...	bhuhaha
ug/ml	(Concentration) Show records Micrograms / milliliter = (10 ...	vit3
uL	(Concentration) Show records We assume that 1ul = 1 micro l ...	kamel
uM	(Concentration) Show records Micro mol corresponds to 10 ...	vit3
umol/L	(Concentration) Show records Micromol per liter. The mole i ...	amazz
ug/cm ³	(Concentration) Show records ug/cm ³ -- micrograms per cubi ...	tanzeem
ug/dl	(Concentration) Show records ug/dl -- micrograms per dec ...	kamel
ug/L	(Concentration) Show records ug/l -- micrograms per liter, ...	ferko

Pagina 26

A list of measures will be shown. Slide down till you find uM or the unit you know is your activity potency.

Preparing the dataset

The screenshot shows the 'Batch Upload 3.0 - Entity remapping' interface. Under 'Database entities remapping', the following settings are visible:

- Property: Ki
- Values: [input field]
- Unit: uM, min value: 0.006, max value: 3.89
- Article: unpublished
- Molecule set: default
- Subgrid: [input field]

At the bottom right, there are two buttons: 'Cancel Batch Upload' and 'Download Excel file'. A blue arrow points to the 'Download Excel file' button.

Pagina 27

At this point almost everything is set. Click on "Download excel file"

Preparing the dataset

The screenshot shows the 'Batch upload 3.0 - records preview' interface. The summary indicates 40 rows in the sheet and 40 records. Two rows are visible, both with errors. Row 1 has a Ki value of 0.009000000000000001 (in uM) = 8.05 (in -log(M)) and an error: 'error: Some obligatory conditions for property Ki have not been specified: [Target]'. Row 2 has a Ki value of 0.026000000000000002 (in uM) = 7.59 (in -log(M)) and the same error. Both rows include chemical structures and identifiers like Ragno, R and N: AUTO_1.

Pagina 28

And it can happen the program complain about something, although everything was correctly made.

In this case the complain is about the Ki.

Preparing the dataset

The screenshot shows an Excel spreadsheet with the following columns: SMILES, ki, and pKi. The pKi column contains the formula =0.009*log10(C2). A blue arrow points to the formula bar, and another blue arrow points to the ki column.

SMILES	ki	pKi
O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC(H)(C)C@H		=0.009*log10(C2)
O=C1N[C@@H](C1=CC=CC=C1)CN(C)C(N(C)C)C(C)C3=C2NN=C3N	0.026	
CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42	0.71	
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C	0.652	
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C)C(C)C=N4)C3=C1	0.884	
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N	1.44	
CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42)NN1	0.099	
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NN(C5CC5)=C4)C3=C	0.016	
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NN(C(C)C)C=C4)C3=C1	0.077	
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NN(C5=CC=CC=C5)=	0.256	
CC1=CC(NC2=NC(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)N	0.354	
CC1=CC(NC2=NC(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)C)N	3.817	
CC1=CC(NC2=NC(N3CCC(N)C)CC3)=O)=NC4=CC=C(C=C42)C)N	0.674	
CC1=CC(NC2=NC(N3CCC(O)C)CC3)=O)=NC4=CC=C(C=C42)C)N	3.89	
CC1=CC(NC2=NC(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)N	0.672	
CC1=CC(NC2=NC(N3CCN(C)C)C3)=O)=NC4=CC=C(C=C42)C)N	0.045	
CC1=CC(NC2=NC(N3CCOCC3)=O)=NC4=CC=C(C=C42)C)N	2.172	
CC1=CC(NC2=NC(N3CCN(C)C)C3)=O)=NC4=CC=C(C=C42)C)N	0.768	
CC1=CC=C(N=C(N2NC3=NN(C)C=C3)C4CC(NCC4)=O)C2=N	2.408	
CC1=CC(NC2=NC(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.576	
CC1=CC(NC2=NC(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.325	
CC1=CC(NC2=NC(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	1.63	
CC1=CC(NC2=NC(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.782	
CC(C=C12)=CC=C1N=C(C(N3CCN(C)C)C3)=O)N=C2NC4=NN(C5CC5)	0.016	
CC1=CC(NC2=NC(N3CCN(C)C)C3)=O)=NC4=CC=C(C)C=C42	0.218	
O=C1N(C)N(C)C@H1(C)C2=NC3=CC=C(C)C=C3(NC4=NN(C5O	0.026	
CC1=CC(NC2=NC(N3CCN(C)C)C3)=O)=NC4=CC=C(C)C=C42	0.151	
O=C1N(C)N(C)C@H1(C)C2=NC3=CC=C(C)C=C3(NC4=NN(C5O	0.017	

So go back in your excel file and convert the Ki into pki as follows

Preparing the dataset

The screenshot shows an Excel spreadsheet with the following data:

SMILES	ki	pKi
O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC(H)(H)C(C)C@H		0.009 =LOG10(C2/1000000)
O=C(N)C@H(C1=CC=CC=C1)CN(C)C(N)(C2(C)C)CC3=C2NN=C3N		0.026
CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42		0.71
CC1=CC=C(N=C(N2CCC(N)CC2)=C3NC4=CN(C5CC5)C=N4)C3=C		0.652
CC1=CC=C(N=C(N2CCC(N)CC2)=C3NC4=CN(C)C(C)C=N4)C3=C1		0.884
CC1=CC=C(N=C(N2CCC(N)CC2)=C3NC4=CN(C5=CC=CC=C5)C=N		1.44
CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42)=NN1		0.099
CC1=CC=C(N=C(N2CCC(N)CC2)=C3NC4=NN(C5CC5)=C4)C3=C		0.016
CC1=CC=C(N=C(N2CCC(N)CC2)=C3NC4=NN(C(C)C)C=C4)C3=C1		0.077
CC1=CC=C(N=C(N2CCC(N)CC2)=C3NC4=NN(C5=CC=CC=C5)=O		0.256
CC1=CC(NC2=NC(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)=NN		0.354
CC1=CC(NC2=NC(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)C)=NN		3.817
CC1=CC(NC2=NC(N3CCC(N)C)CC3)=O)=NC4=CC=C(C=C42)C)=N		0.674
CC1=CC(NC2=NC(N3CCC(O)C)CC3)=O)=NC4=CC=C(C=C42)C)=N		3.89
CC1=CC(NC2=NC(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)=N		0.672
CC1=CC(NC2=NC(N3CCN(C)C3)=O)=NC4=CC=C(C=C42)C)=NN1		0.045
CC1=CC(NC2=NC(N3CCOCC3)=O)=NC4=CC=C(C=C42)C)=NN1		2.172
CC1=CC(NC2=NC(N3CCN(C)C3)=O)=NC4=CC=C(C=C42)C)=NN		0.768
CC1=CC=C(N=C(N2NC3=NN(C)C=C3)C(N4C(NCC4)=O)=O)C2=N		2.408
CC1=CC(NC2=NC(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)=N		0.576
CC1=CC(NC2=NC(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1		0.325
CC1=CC(NC2=NC(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)=N		1.63
CC1=CC(NC2=NC(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1		0.782
CC(C=C12)=CC=C11=C(C(N3CCN(C)C3)=O)=N=C(N4=NN(C5CC5)		0.016
CC1=CC(NC2=NC(N3CCN(C)C@H)C)C=C(N4=CC=C(C)C)C=C42		0.218
O=C(N1CN(C)C@H)1)C2=NC3=CC=C(C)C=C3(NC4=NN(C5O		0.026
CC1=CC(NC2=NC(N3CCN(C)C@H)3)C=C(N4=CC=C(C)C)C=C42		0.151
O=C(N1CCN(C)C@H)1)C2=NC3=CC=C(C)C=C3(NC4=NN(C5CC		0.017

Click on the cell aside the first Ki value end insert”=log10(C2/1000000). This means you are going to transform the ki from micromolar to molar and make the log10.

Preparing the dataset

The screenshot shows a spreadsheet with two columns: 'SMILES' and 'pKi'. The 'pKi' column header is highlighted in blue, and a blue arrow points to it from the right. The spreadsheet contains 28 rows of data, each with a SMILES string and a numerical pKi value.

SMILES	pKi
O=C1NCC2=C1C3=C4C5=C2C6=C(C=C=C6)N5C7(C)OC(H)(H)C(C)C@	0.009
O=C(N)C@H(C1=CC=CC=C1)N(C)C(N(C2(C)C)CC3=C2NN=C3N	0.026
CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42	0.71
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5C5)C=N4)C3=C	0.652
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C)C(C)C=N4)C3=C1	0.884
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N	1.44
CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42)NN1	0.099
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NN(C5CC5)=C4)C3=C	0.016
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NN(C)C(C)C=C4)C3=C1	0.077
CC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NN(C5=CC=CC=C5)=O	0.256
CC1=CC(NC2=NC(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)N	0.354
CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)C)N	3.817
CC1=CC(NC2=NC(N3CCC(N)C)C(C3)=O)=NC4=CC=C(C=C42)C)N	0.674
CC1=CC(NC2=NC(N3CCC(O)C)C(C3)=O)=NC4=CC=C(C=C42)C)N	3.89
CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)C)N	0.672
CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.045
CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)C)N	2.172
CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)C)N	0.768
CC1=CC=C(N=C(N2NC3=NN(C)C=C3)C4CC(NCC4)=O)C2=N	2.408
CC1=CC(NC2=NC(C(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.576
CC1=CC(NC2=NC(C(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.325
CC1=CC(NC2=NC(C(N)C3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	1.63
CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)C)N	0.782
CC(C=C12)=CC=C1N=C(C(N3CCNCC3)=O)N=C2NC4=NN(C5CC5)	0.016
CC1=CC(NC2=NC(C(N3CCN(C)C@H)3C)C)N=C4=CC=C(C)C=C42	0.218
CC1=CC(NC2=NC(C(N)C3CCN(C)C@H)3C)C)N=C4=CC=C(C)C=C42	0.026
CC1=CC(NC2=NC(C(N3CCN(C)C@H)3C)C)N=C4=CC=C(C)C=C42	0.151
CC1=CC(NC2=NC(C(N)C3CCN(C)C@H)3C)C)N=C4=CC=C(C)C=C42	0.017

As soon as you click on enter the pKi value is calculated

Preparing the dataset

Row	Chemical Structure	pKi Value
17	CC1=CC(NC2=N(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1	0.045
18	CC1=CC(NC2=N(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)C)=NN1	2.172
19	CC1=CC(NC2=N(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)C)=NN	0.768
20	CC1=CC=C(N=C(N=C2NC3=NN(C)=C3)C(N4C(NC4)=O)=O)C2=	2.408
21	CC1=CC(NC2=N(C(N(C)3CCNCC3)=O)=NC4=CC=C(C=C42)C)=N	0.576
22	CC1=CC(NC2=N(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1	0.325
23	CC1=CC(NC2=N(C(N(C)3CCNCC3)=O)=NC4=CC=C(C=C42)C)=N	1.63
24	CC1=CC(NC2=N(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)C)=NN1	0.782
25	CC(C=C12)=CC=C1N=C(C(N3CCNCC3)=O)N=C2NC4=NNC(C5CC5)	0.016
26	CC1=CC(NC2=N(C(N3CCN(C)C@H)3C)=O)=NC4=CC=C(C)C=C42	0.218
27	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5O)	0.026
28	CC1=CC(NC2=N(C(N3CCN(C)C@H)3C)=O)=NC4=CC=C(C)C=C42	0.151
29	O=C(N1CCN(C)C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5CC)	0.017
30	CC1=CC(NC2=N(C(N3CCN(C)C@H)1C)3)=O)=NC4=CC=C(C)C=C42	0.051
31	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5C)	0.009
32	CC1=CC(NC2=N(C(N3CCN(C)C@H)1C)3)=O)=NC4=CC=C(C)C=C4	0.306
33	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5)	0.028
34	CC1=CC(NC2=N(C(N3CCN(C)C)3)=O)=NC4=CC=C(C)C=C42	0.119
35	O=C(N1CCN(C)C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5CC)	0.114
36	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5)	0.006
37	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5CC)	0.017
38	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5)	0.007
39	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5)	0.016
40	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C5)	0.011
41	O=C(N1CCN(C@H)1C)C2=NC3=CC=C(C)C=C3(NC4=NNC(C)	0.036

Now select the cell containing the first pKi value and by clicking on the little square on the right-bottom and slide down to apply the calculation of the pKi to all the other cells.

Select the pKi column and copy into the memory ...

Preparing the dataset

The screenshot shows a spreadsheet with two columns: SMILES strings and numerical values. A 'Paste Special' dialog box is open, with 'Numbers' selected under the 'Paste as' options. Blue arrows point to the 'Paste Special' button and the 'Numbers' option.

SMILES	Value 1	Value 2
O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC(H)C(C)C@H	0.009	8.04575749
O=C(N)C@H(C1=CC=CC=C1)N(C)C(N)C(C)C(C)C=C2N=C3N	0.026	7.58502665
C(C=N)C=C1NC2=N(C)N3CCC(N)CC3)=NC4=CC=C(C)C=C42	0.71	6.14874165
CC1=CC=C(N=C(N)C2CCC(N)CC2)N=C3NC4=CN(C)C=CC=C4	0.052	6.1857524
CC1=CC=C(N=C(N)C2CCC(N)CC2)N=C3NC4=CN(C)C=CC=C4	0.884	6.05354773
CC1=CC=C(N=C(N)C2CCC(N)CC2)N=C3NC4=CN(C)C=CC=C4	1.44	5.84163751
CC1=CC(NC2=N(C)N3CCC(N)CC3)=NC4=CC=C(C)C=C4	0.099	7.00436481
CC1=CC=C(N=C(N)C2CCC(N)CC2)N=C3NC4=NN(C)C=CC=C4	0.016	7.79589002
CC1=CC=C(N=C(N)C2CCC(N)CC2)N=C3NC4=NN(C)C=CC=C4	0.077	7.11350927
CC1=CC=C(N=C(N)C2CCC(N)CC2)N=C3NC4=NN(C)C=CC=C4	0.256	6.59176003
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.354	6.45099674
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	3.817	5.41827784
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.674	6.1713401
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	3.89	5.4100504
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.672	6.17263073
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.045	7.34678749
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	2.172	5.66314018
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.768	6.11463878
CC1=CC=C(N=C(N)C2N3=NN(C)C=C3)C3C(N)C4	2.408	5.61834352
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.576	6.23957752
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.325	6.48811664
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	1.63	5.7878124
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.782	6.10679325
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.016	7.79589002
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.218	6.66154351
O=C1N(C)C(N)C@H(C)C2=NC3=CC=C(C)C=C3C(N)C4=NN(C)C5O	0.026	7.58502665
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.151	6.82102305
CC1=CC(NC2=N(C)N3CCC(N)CC3)=O)N=C4=CC=C(C)C=C4	0.017	7.76051106

... then paste it on the Ki column by choose “paste special” and indicating just the numbers.

Preparing the dataset

The screenshot shows an Excel spreadsheet with the following data:

SMILES	pKi
<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=C=C6)N5C7(C)OC(H)C(C)@H</chem>	8.04575749056068
<chem>O=C(N)C@H(C1=CC=CC=C1)N(C)C(N)C2(C)C(C)CC3=C2NN=C3N</chem>	7.58502665202918
<chem>CN(C=N1)C=C1NC2=N(C)N3CCC(N)CC3=NC4=CC=C(C)C=C42</chem>	6.14874165128092
<chem>CC1=CC=C(N=C(N)N2CCC(N)CC2)N=C3NC4=CN(C)C5C5C=C=N4C3=C</chem>	6.18575240426908
<chem>CC1=CC=C(N=C(N)N2CCC(N)CC2)N=C3NC4=CN(C)C(C)C=C=N4C3=C1</chem>	6.05354773496993
<chem>CC1=CC=C(N=C(N)N2CCC(N)CC2)N=C3NC4=CN(C)C5=CC=CC=C5C=C4</chem>	5.84163750790475
<chem>CC1=CC(NC2=N(C)N3CCC(N)CC3)=NC4=CC=C(C)C=C42=NN1</chem>	7.00436480540245
<chem>CC1=CC=C(N=C(N)N2CCC(N)CC2)N=C3NC4=NN(C)C(C)C=C4C3=C</chem>	7.79588001734408
<chem>CC1=CC=C(N=C(N)N2CCC(N)CC2)N=C3NC4=NN(C)C(C)C=C4C3=C1</chem>	7.11350927482752
<chem>CC1=CC=C(N=C(N)N2CCC(N)CC2)N=C3NC4=NN(C)C5=CC=CC=C5=O</chem>	6.59176003468815
<chem>CC1=CC(NC2=N(C)C(N)N3CCC(N)CC3=O)=NC4=CC=C(C=C42)C)=NN</chem>	6.45099673797421
<chem>CC1=CC(NC2=N(C)C(N)N3CCC(O)CC3=O)=NC4=CC=C(C=C42)C)=NN</chem>	5.4182778400509
<chem>CC1=CC(NC2=N(C)C(N)N3CCC(N)C(C)C3=O)=NC4=CC=C(C=C42)C)=N</chem>	6.17134010346468
<chem>CC1=CC(NC2=N(C)C(N)N3CCC(O)C(C)C3=O)=NC4=CC=C(C=C42)C)=N</chem>	5.41005039867429
<chem>CC1=CC(NC2=N(C)C(N)N3CCC(N)C(C)C3=O)=NC4=CC=C(C=C42)C)=N</chem>	6.17263072694618
<chem>CC1=CC(NC2=N(C)C(N)N3CCN(C)C3=O)=NC4=CC=C(C=C42)C)=NN1</chem>	7.34678748622466
<chem>CC1=CC(NC2=N(C)C(N)N3CCOCC3=O)=NC4=CC=C(C=C42)C)=NN1</chem>	5.66314017908319
<chem>CC1=CC(NC2=N(C)C(N)N3CCN(C)C3=O)=NC4=CC=C(C=C42)C)=NN</chem>	6.11463877996849
<chem>CC1=CC=C(N=C(N)N2C3=NN(C)C=C3)C4CC(N)CC4=O=O)C2=O</chem>	5.61834351741421
<chem>CC1=CC(NC2=N(C)C(N)C3CCN(C)C3=O)=NC4=CC=C(C=C42)C)=N</chem>	6.23957751657679
<chem>CC1=CC(NC2=N(C)C(N)C3CCN(C)C3=O)=NC4=CC=C(C=C42)C)=NN1</chem>	6.48811663902113
<chem>CC1=CC(NC2=N(C)C(N)C3CCN(C)C3=O)=NC4=CC=C(C=C42)C)=N</chem>	5.78781239559604
<chem>CC1=CC(NC2=N(C)C(N)N3CCN(C)C3=O)=NC4=CC=C(C=C42)C)=NN1</chem>	6.10679324694015
<chem>CC(C=C12)=CC=C1N=C(N)N3CCC(N)CC3=O)=NC2NC4=NN(C)C5C5C=C</chem>	7.79588001734408
<chem>CC1=CC(NC2=N(C)C(N)N3CCN(C)C@H)C3=O)=NC4=CC=C(C)C=C42</chem>	6.6615435063954
<chem>O=C(N)CCN(C)C@H)C1C2=C3=CC=C(C)C=C3(N)C4=NN(C)C5O</chem>	7.58502665202918
<chem>CC1=CC(NC2=N(C)C(N)N3CCN(C)C@H)C3=O)=NC4=CC=C(C)C=C42</chem>	6.82102305270683
<chem>O=C(N)CCN(C)C@H)C1C2=C3=CC=C(C)C=C3(N)C4=NN(C)C5C5C=C</chem>	7.76055107852172

Pagina 34

Now you have the excel file containing RECORDID, SMILES and pKi columns

Preparing the dataset

Batch Upload 3.0 - File preview and column remapping

jm7b01342_sl_002

RECORDID	SMILES	pK
stauroporine	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C...</chem>	8.045757490560675
PF3758309	<chem>O=C([N]([C@@H])(C1=CC=CC=C1)CN(C)N)(C2(C...</chem>	7.585026652029182
10a	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=C...</chem>	6.1487416512809245
10b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=CN(...</chem>	6.185752404268079
10c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=CN(...</chem>	6.053547734986927
10d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=CN(...</chem>	5.841637507904751
11a	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C...</chem>	7.00436480540245
11b	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=NNC...</chem>	7.795880017344075
11c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=NNC...</chem>	7.113509274827518
11d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2))=C3NC4=NNC...</chem>	6.59176003468815

The ARTICLE column is missing, the stub unpublished article will be assigned by default

Green titles indicate recognized columns, red titles indicate errors. Please click on the red columns and select whether the column indicates a property, condition or another column type like name, value or molecule, then select the matching entity and confirm your selection by clicking on the green button on the left. If you have irrelevant columns in your sheet, you can leave them red and they will be ignored in the further process. If you

Pagina 35

Go to OCHEM.eu and repeat the uploading of the excel file as above described

Preparing the dataset

Batch Upload 3.0 - Entity remapping

Review and remap the properties, conditions, units, articles and baskets involved in the data upload

Database entities remapping

Property: pki

Values
Unit: $-\log(\text{mol/L})$, min value: 5.410050398674292,
max value: 8.221848749616356

Article: unpublished

Molecule set: #default

submit

Cancel Batch Upload Download Excel file

Now the unity of the activity potency is correctly reported, click on “Download Excel file”

Preparing the dataset

Batch upload 3.0 - records preview

Preview the records you are about to upload, select the desired actions

Batch upload preview browser

Summary:
All rows in the sheet Count: **40**
Status: valid, Count: **40**

Filter by row number: and row type: all Batch operations

1 - 10 of 40

Row 1
+Save
⊖Skip

C1=CC=C2C(=C1)C(=O)N(C2)C3=CC=CC=C3

● pKi = 8.045757490560675 (in -log(mol/L))
Ragno, R
jm7b01342_w_002.xls...
N: AUTO_1
MoleculeID: M4402773
RecordID: R-1
rino.ragno Only visible to rino.ra

Row 2
+Save
⊖Skip

C1=CC=C2C(=C1)C(=O)N(C2)C3=CC=CC=C3

● pKi = 7.585026652029182 (in -log(mol/L))
Ragno, R
jm7b01342_w_002.xls...
N: AUTO_2
MoleculeID: M9541999
RecordID: R-2
rino.ragno Only visible to rino.ra

jm7b01342_w_0_csw Show all

Pagina 37

And everything is OK, slide down to check all the rows ...

In this tutorial the OCHEM.eu portal will complain about data duplication with the user rino.ragno@uniroma1.it.

Ignore it and go on.

Preparing the dataset

The screenshot displays the 'Online chemical database' interface. It shows a list of chemical records with the following details:

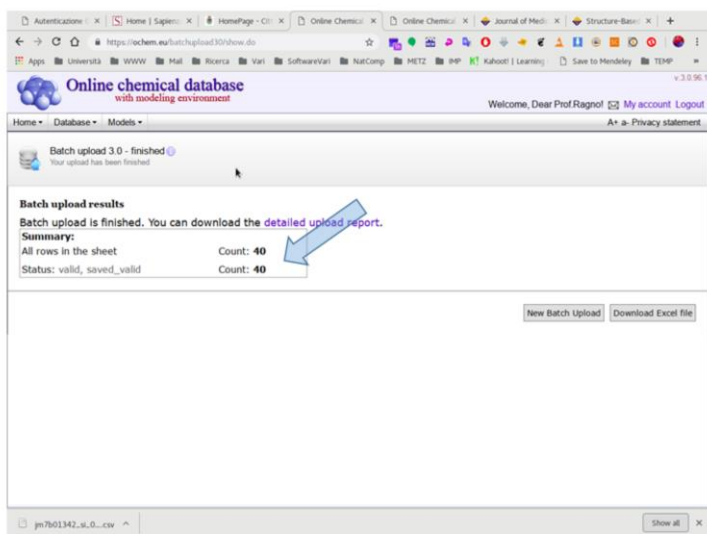
- Row 8:** pK1 = 7.79300027344073 (in -log(mol/L))
Ragno, R
jm7b01342_sl_002.xls...
N: AUTO_8
MoleculeID: M9713352
RecordID: R-8
rino.ragno Only visible to rino.ra
- Row 9:** pK1 = 7.113509274827518 (in -log(mol/L))
Ragno, R
jm7b01342_sl_002.xls...
N: AUTO_9
MoleculeID: M9713353
RecordID: R-9
rino.ragno Only visible to rino.ra
- Row 10:** pK1 = 6.59176003468815 (in -log(mol/L))
Ragno, R
jm7b01342_sl_002.xls...
N: AUTO_10
MoleculeID: M9713354
RecordID: R-10
rino.ragno Only visible to rino.ra

At the bottom of the list, there is a 'Proceed with upload' button and a 'Download Excel file' button. Two blue arrows point to these buttons. The interface also shows a 'Cancel Batch Upload' button and a 'Show all' button.

Pagina 38

... and if everything is OK click on "Proceed with upload" again

Preparing the dataset



The screenshot displays the 'Online chemical database' web interface. At the top, there is a navigation bar with 'Home', 'Database', and 'Models' menus. A notification banner at the top left states 'Batch upload 3.0 - finished' with a sub-message 'Your upload has been finished'. Below this, the 'Batch upload results' section is visible, containing a 'Summary' table. A blue arrow points to the 'detailed upload report' link in the text above the table. At the bottom right of the results section, there are two buttons: 'New Batch Upload' and 'Download Excel file'. The browser's address bar shows the URL 'https://chem.eu/submit/upload30/show.do'.

Summary:	
All rows in the sheet	Count: 40
Status: valid, saved_valid	Count: 40

Pagina 39

The portal will report on the data you have uploaded.

Preparing the dataset

The screenshot displays the 'Online chemical database' website interface. The browser's address bar shows the URL 'https://chem.eu/online-chemical-database'. The page header includes the site logo and the text 'v.3.0.96.1'. A navigation menu at the top contains 'Home', 'Database', and 'Models'. The 'Models' menu is open, showing a list of options: 'Create a model', 'Apply a model', 'Create multiple models', 'Create multiple models with conditions (experimental)', 'Open predictor', 'Upload a linear model', 'Upload a stub model', 'View pending tasks', 'View published tasks', 'SetCompare utility', 'MolOptimiser', 'Calculate descriptors', and 'Descriptors storage'. A blue arrow points to the 'Create a model' option. Below the menu, there is a 'Batch upload' section with a 'Summary' box and buttons for 'New Batch Upload' and 'Download Excel file'.

Pagina 40

Now let try to make some models. Click on the “Models” tab and select “Create a model”

Preparing the dataset

The screenshot shows the 'Online chemical database' interface. The main window is titled 'Create a model' and contains the following sections:

- Select the training and validation sets:** Includes a 'Training set (required): []' and an 'Add a validation set' button.
- Choose the learning method:** A list of 'Suggested modeling methods' with radio buttons next to each. The selected method is 'PLS - Partial Least Squares Support Vector Machine (GPU)'. Other methods include ASNN, CHEMCHAMBER, CNF, Consensus model, DEEPCHEM, DNN, EAGCNG, FSMLR, KNN, LASSO, LS-SVM, LSSVM, Multiple Linear Regression, PLS, RFR, WEKA-J48, WEKA-RF, and XGBoost.

Pagina 41

A window will give you the opportunity to select the statistical engine, for this purpose select PLS (although in the image is show selected MLR)

Preparing the dataset

The screenshot shows the 'Online chemical database' website interface. The page title is 'Online chemical database with modeling environment'. The user is logged in as 'Prof Ragnol'. The main content area is titled 'Models' and lists various machine learning models such as CHEMFASTER, CHEM, CHF, Consensus model, DEEPCHEM, DNN, EAGONS, FSMLR, KNN, LIBSVM, LSSVM, MLR, PLS, RFR, WEKA J48, WEKA RF, and XGBoost. Below the list, there are sections for 'Methods under development' and 'Model validation'. The 'Model validation' section includes options for 'Validation method' (N-Fold cross validation), 'Number of folds' (5), 'Stratified cross-validation' (checked), and 'Consider each record as a molecule' (checked). At the bottom of the configuration area, there is a 'Next>>' button. A blue arrow points to this button.

Pagina 42

... slide down and click on "Next>>"

Preparing the dataset

Online chemical database
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models

Create a model

Select the training and validation sets, the machine learning model and the validation protocol

Select the training and validation sets:

Training set (required):
Add a validation set
Click to change

Choose the learning method:

Suggested modeling methods:

- ASNN: Associative Neural Networks
- CHEMANNER: Chainer Chemistry models (GPU)
- CNF - Convolutional Neural Network Fingerprint (GPU)
- Consensus model (based on models developed for the same set)
- DEEPCHEM: several methods from DeepChem (GPU)
- DNN: Deep Neural Network (GPU)
- EAGCNG - Edge Attention based Multi-relational Graph Convolutional Networks
- FSMLR: Fast Stagewise Multiple Linear Regression
- KNN: k - Nearest Neighbors
- Library model (A local bias correction model based on another ASNN model)
- LIBSVM: grid-search parameter optimisation
- LSSVM: Least Squares Support Vector Machine (GPU)
- MLR: Multiple Linear Regression
- PLS: Partial Least Squares
- RFR: Random Forest regression and classification
- WEKA-HS: Weka C4.5 decision trees, only classification - use with bagging
- WEKA-RF: Random Forest, only classification

Show all

Pagina 43

... but it will ask you to select a training set.
Then click in the “[...]” link ...

Preparing the dataset

The screenshot shows the 'Online chemical database' interface. The main content area is titled 'Basket browser' and contains a table of records. The table has two columns: 'Selected records' and 'records'. The first row is highlighted in green and contains the text 'm7b01342_002.xls' and '40 records'. A blue arrow points to this row. Below the table, there is a button that says 'Click to select this basket'. Another blue arrow points to this button. The interface also includes a search bar, a 'Filter by name' field, and a 'Show public sets' link. The browser's address bar shows the URL 'https://occhem.eu/model/configurator/choose.do'.

Selected records	records
m7b01342_002.xls	40 records

Pagina 44

... and select the excel file containing the dataset as prepared above

Preparing the dataset

The screenshot shows the 'Online chemical database with modeling environment' interface. The main content area is titled 'Create a model' and 'Select the training and validation sets:'. It displays a training set 'jm7b01342_sl_002.xls' and a property 'pKi using unit log(mol/L)'. Under 'Choose the learning method:', a list of suggested modeling methods is shown, with 'PLS - Partial Least Squares' selected and highlighted in blue. Two blue arrows point to the 'PLS' option and the 'Show all' button at the bottom right of the list.

Online chemical database
with modeling environment

Welcome, Dear Prof Ragnol My account Logout

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): [jm7b01342_sl_002.xls](#) [details]

Add a validation set

The model will predict this property:
pKi using unit [log\(mol/L\)](#)

Choose the learning method:

Suggested modeling methods:

- ASN1: Associative Neural Networks
- CHEMCHAIER: Chainer Chemistry models (GPU)
- CNF - Convolutional Neural Network Fingerprint (GPU)
- Consensus model (based on models developed for the same set)
- DEEPCHEM: several methods from DeepChem (GPU)
- DNN: Deep Neural Network (GPU)
- EAGCNG - Edge Attention based Multi-relational Graph Convolutional Networks
- FSMLR: Fast Stagewise Multiple Linear Regression
- KNN: k-Nearest Neighbors
- Library model (lib) - local bias correction model based on another ASN1 model)
- LIBSVM: grid-search parameter optimisation
- LSVM3: Least Squares Support Vector Machine (GPU)
- PLS: Multiple Linear Regression
- PLS: Partial Least Squares
- RFR: Random Forest regression and classification
- WEKA-J48: Weka C4.5 decision trees, only classification - use with bagging

[jm7b01342_sl_002_cw](#) Show all

Pagina 45

Now the correct statistical method is selected (PLS) and there is also the training set.

Slide down ...

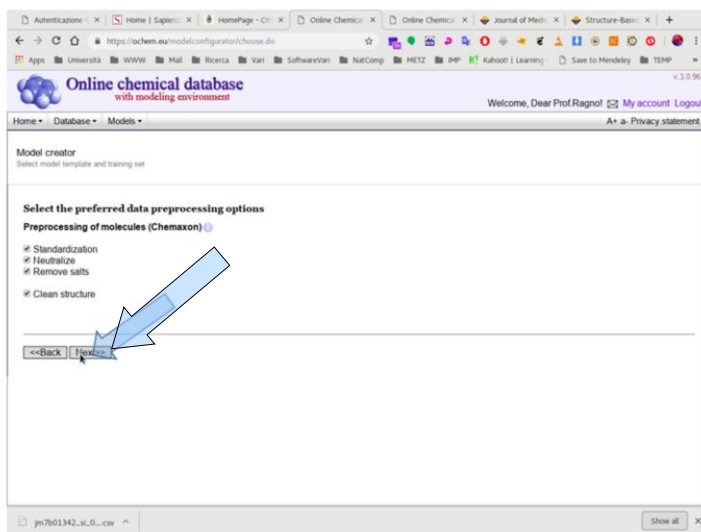
Preparing the dataset

The screenshot displays the 'Online chemical database' interface. The page title is 'Online chemical database with modeling environment'. The user is logged in as 'Prof Ragno!'. The main content area shows a list of models under the 'Models' tab, including CHENCHAIER, CHN, Consensus model, DEEPCHEM, DNN, EAGONG, FSMLR, KNN, LIBSVM, LSSVM, MLR, PLS, RFR, WEKA-J48, WEKA-RF, and XGBoost. Below the model list, there is a 'Model validation' section with a dropdown menu set to 'N-Fold cross-validation'. Below that, there are checkboxes for 'Number of folds: 5', 'Stratified cross-validation (classification only)', and 'Consider each record as a molecule'. At the bottom left, there is a 'Next >>' button, which is highlighted by a blue arrow. The browser's address bar shows 'https://occhem.eurmodel.configurator/choose.do'.

Pagina 46

... and click on "Next>>"

Preparing the dataset



Pagina 47

A window will ask you how to treat the molecules, leave everything as suggested and click on “Next>>”.

Preparing the dataset

The screenshot displays the 'Model creator' interface of the Online Chemical Database. The main section is titled 'Select the following descriptors' and contains a list of 'Recommended descriptor types'. A blue arrow points to the 'E-State' category, which is expanded to show sub-options: 'Atom counts', 'Bonds counts', and 'Aromaticity structures'. The 'Aromaticity structures' dropdown is currently set to 'Chemaxon Basic'. Other descriptor types listed include 'ALogPS (2)', 'CSF fragment (1138)', 'Dragon v 7 (52703D)', 'alvaDesc v1 0 14 (53053D)', 'ESDA fragments', 'Inductive descriptors (543D)', 'MERA descriptors (5293D)', 'MERSY descriptors (423D)', 'Chemaxon descriptors (4993D)', 'GNPR', and 'Spectrophores (1443D)'. On the right side, there are sections for 'Predictions by OCHEM's featured models' and 'Obsolete/Additional descriptor types'. The 'Obsolete/Additional descriptor types' section includes 'CDK 1.4.11 descriptors (2563D)', 'OESState', 'Dragon v 5.4 (16443D)', and 'Dragon v 5.5 (32243D)'. A 'Show all' button is located at the bottom right of the interface.

Pagina 48

In this windows the portal let you select which descriptors you want to use. Uncheck everything ...

Preparing the dataset

Model creator
Select model template and training set

Select the molecular descriptors

Recommended descriptor types

- E-state
- ALogPS (2)
- GSFfragment (1138)
- CDK 2.0 descriptors (256/3D)
- Dragon v. 7 (5270/3D)
- alogchec v.1.0.14 (5305/3D)
- ISIDA fragments
- 'Inductive' descriptors (54/3D)
- MERA descriptors (529/3D)
- MERSY descriptors (42/3D)
- Chemaxon descriptors (499/3D)
- GHPR
- Spectrophores (144/3D)
- Structural alerts (ToxAlerts)

Predictions by OCHEM's featured models

- Ames Iverenberg
- Toxicity against T. Pyriformis
- ALogPS 3.0
- CYP1A2 Estate+ALogPS
- CYP2C9 Estate+ALogPS
- CYP2C19 Estate+ALogPS
- CYP2D6 Estate+ALogPS
- CYP3A4 Estate+ALogPS
- Pyridyls point prediction (best Estate)
- Melting Point prediction (best Estate)
- Water solubility model based on logP and Melting Point
- ALOGPS 2.1 logP
- ALOGPS 2.1 logS
- Outputs of other OCHEM models

Obsolete/Additional descriptor types

- CDK 1.4.11 descriptors (256/3D)
- OEState
- Dragon v. 5.4 (1644/3D)
- Dragon v. 5.5 (3224/3D)

jn7h01342_sl_0_csv Show all

Pagina 49

Once all is unchecked ...

Preparing the dataset

The screenshot shows the 'Online chemical database' interface. The main heading is 'Select the molecular descriptors'. Under 'Recommended descriptor types', the following are listed:

- E-state
- AS-indices (1138)
- CDK descriptors (2563D)
- Dragon v 7 (52703D)

The 'Dragon v 7 (52703D)' item is highlighted in yellow. A blue arrow points to it. Below this list, there are two columns of other descriptor types, each with a checked checkbox:

- Constitutional descriptors (47)
- Topological indices (75)
- Connectivity indices (37)
- 2D matrix-based descriptors (607)
- Burden eigenvalues (96)
- ETA indices (23)
- Geometrical descriptors (3D, 38)
- 3D autocorrelations (3D, 86)
- 3D-MoRSE descriptors (3D, 224)
- GETAWAY descriptors (3D, 273)
- Functional group counts (3D, 154)
- Atom-type E-state indices (172)
- 2D Atom Pairs (1596)
- Charge descriptors (3D, 15)
- Drug-like indices (28)
- Ring descriptors (32)
- Walk and path counts (46)
- Information indices (50)
- 2D autocorrelations (213)
- P_VSA-like descriptors (55)
- Edge adjacency indices (324)
- 3D matrix-based descriptors (3D, 99)
- RDF descriptors (3D, 210)
- WHIM descriptors (3D, 114)
- Random molecular profiles (3D, 41)
- Atom-centred fragments (115)
- CATS-2D (150)
- 3D Atom Pairs (3D, 36)
- Molecular properties (20)
- CATS-3D (3D, 300)

On the right side, there are sections for 'Predictions by OCHEM's featured models' and 'Obsolete/Additional descriptor types'. The 'Obsolete/Additional descriptor types' section includes:

- CDK 1.4.11 descriptors (2563D)
- OESState
- Dragon v 5.4 (18443D)
- Dragon v 5.5 (32243D)
- Dragon v 6 (48853D)
- MOPAC 7.1 descriptors (253D)

Pagina 50

Select the Dragon v. 7 descriptors. Make sure all are selected.

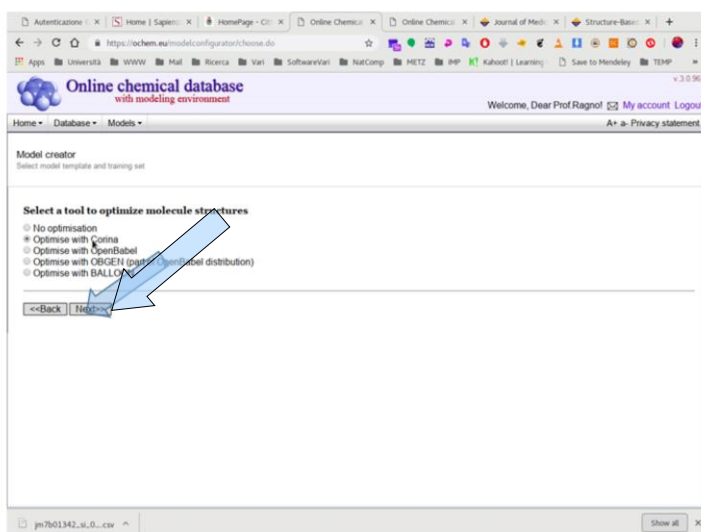
Preparing the dataset

The screenshot shows the 'Online chemical database' configuration interface. The page title is 'Preparing the dataset'. The browser address bar shows 'https://ochem.ejbmolconfigurator/choose.do'. The page content includes a navigation menu with 'Home', 'Database', and 'Models'. Under 'Database', there are options for 'QIPR', 'Spectrophores (144/3D)', and 'Structural alerts (ToxAlerts)'. A section titled 'Special descriptors (scaffolds, fingerprints):' lists 'Chemaxon Scaffolds', 'Silicos-It Scaffolds', 'E-SP: Fingerprints (not supported by your installation)', and 'MolPrint Fingerprints'. Below this is a warning: 'Under development: can change anytime and backward compatibility is not guaranteed. Use at your own risk!'. A list of descriptors follows: 'JPligP', 'CDK 2.2 descriptors (256/3D)', 'RDKit descriptors (3D)', 'RDKit additional descriptors (3D)', 'MORDRED descriptors (1826/3D)', 'CDOD', 'MORPC2016 descriptors (35/3D)', 'SIRMS', 'PyDescriptor descriptors (16251/3D)', 'External descriptors', and 'Allow Merging Descriptors (experimental)'. At the bottom, there are '<<Back' and 'Next>>' buttons. A blue arrow points to the 'Next>>' button, and another blue arrow points to the scrollbar on the right side of the page.

Pagina 51

Slide down and click on “Next>>”

Preparing the dataset



Pagina 52

Here you can choose how to convert into 3D the smiles, accept “optimise with Corina” and click on “Next>”

Preparing the dataset

Model creator
Select model template and training set

Select filters of descriptors

- Eliminate descriptors with less than 2 unique values
- Delete descriptors that have absolute values larger than 999996
- Delete descriptors that have variance smaller than 0.01
- Group descriptors, that have pair-wise correlations Pearson's correlation coefficient R larger than 0.95
- Use Unsupervised Forward Selection to delete variables using the above value of multiple correlation coefficient R
- After filtering, I want to select necessary descriptors myself (advanced)

Normalisation parameters

Descriptors normalization Do not normalize

Values normalization Do not normalize

<<Back Next>>

Pagina 53

In this next window you can transform and filter the descriptor data

Deselect "Group descriptors" and ...

Preparing the dataset

The screenshot shows the 'Model creator' interface of the Online Chemical Database. It features a 'Select filters of descriptors' section with several checkboxes and input fields for filtering descriptors based on unique values, absolute values, variance, and correlations. Below this is the 'Normalisation parameters' section, where the 'Descriptors normalization' dropdown is set to 'Standardize (zero mean and unit variance)'. The 'Values normalization' dropdown is set to 'Do not normalize'. Blue arrows point to the 'Standardize' option and the 'Next>>' button. The interface also includes a navigation bar at the top with 'Home', 'Database', and 'Models' menus, and a footer with 'Pagina 54'.

Model creator
Select model template and training set

Select filters of descriptors

- Eliminate descriptors with less than 2 unique values
- Delete descriptors that have absolute values larger than 999996
- Delete descriptors that have variance smaller than 0.01
- Group descriptors, that have pair-wise correlations Pearson's correlation coefficient R larger than 0.95
- Use Unsupervised Forward Selection to delete variables using the above value of multiple correlation coefficient R
- After filtering, I want to select necessary descriptors myself (advanced)

Normalisation parameters

Descriptors normalization: Standardize (zero mean and unit variance)

Values normalization: Do not normalize

<<Back Next>>

... set the normalization as shown in the image.

Then click on "Next>>"

Preparing the dataset

The screenshot displays the 'Online chemical database with modeling environment' interface. The main content area is titled 'Model creator' and contains a 'Configure PLS method' section. This section includes a 'Number of latent variables' input field set to 0, a checkbox for 'Optimize the number of latent variables automatically', and a checkbox for 'Limit predicted values to training set range'. A blue arrow points to the 'Next>>' button at the bottom of the configuration area. The browser's address bar shows the URL 'https://chem.su.se/modelling/prepare1/buona.do'. The page footer contains the text 'Pagina 55'.

... everything is ready, just click on “Next>>” to launch the calculation.

Preparing the dataset

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: pk6_PLS_Dragon7 (blocks: 1-30)

Save models

Task priority:
 High priority (please, use for fast tasks)
 Normal priority
 Low priority (for long tasks)

<<Back | Start calculation | Discard

Pagina 56

Before launching the portal will ask to set the name of the under creating model. Modify it as you like and click on “Start calculation”

Preparing the dataset

The screenshot displays a web browser window with the URL <https://ochem.eurymodel.configurator/choose.do>. The page header includes the logo for "Online chemical database with modeling environment" and a user greeting: "Welcome, Dear Prof Ragnoli". The main content area is titled "Run model builder" and shows a progress indicator with the text "Finished posting ... - Processing task: Corina - Waiting for a free server -- 09:16". Below this, there are two buttons: "<<Back" and "Next>>". The browser's address bar and several open tabs are visible at the top.

Pagina 57

A windows will open and showing the process ...

Preparing the dataset

The screenshot shows the 'Model creator' interface for the 'Online chemical database with modeling environment'. The model name is 'pKi_PLS_Dragon7 (bloc)'. The predicted property is 'pKi modeled in -log(mol/L)'. The training method is 'PLS'. The data set table shows 40 records with an R2 of 0.4 ± 0.2. A scatter plot shows a positive correlation between predicted and actual pKi values. The standardized latent variables are listed on the right.

Data Set	#	R ²	RMSE	MAE	
Training set: jm7b01342_sl_002_als	40 records	0.4 ± 0.2	0.3 ± 0.3	0.7 ± 0.2	0.46 ± 0.08

Model name: pKi_PLS_Dragon7 (blocks: 1-30) | 336922 [wname]
Temporal Public ID: 37774824 - use this link to share the model

Predicted property: pKi modeled in -log(mol/L)
Training method: PLS

scale X: STANDARDIZED latent variables 3D by Cornea
5-fold cross-validation
3008 pre-filtered descriptors

scale X: STANDARDIZED latent variables
Y = -1.03 + 7.3E-6*W + 0.001437*MMW +
9.56E-5*Sw + 7.81E-5*Sp + 9.37E-5*Sp + 7.9E-
5*Pc - 0.06647*W + 0.00879*W + 0.0737*W +
0.0959*W + 0.0383*GD + 8.21E-5*W +
7.64E-5*W - 0.00167*TA + 1.01E-4*W +
1.66E-4*W + 4.82E-4*W + 3.19E-5*W +
0.00108*W + 0.0679*W + 1.51E-4*W +
4.65E-4*W + 1.92E-4*W + 9.4E-5*W +
0.00143*W - 0.00104*W - 0.00236*W +
0.00246*W - 3.41E-4*W - 8.88E-4*W +
3.52E-4*W - 1.95E-4*W + 1.81E-4*W +
6.79E-4*W + 4.88E-4*W + 0.00108*W + ...

After some minutes the results will be shown.

Here you can see the R2 is not so high (2 Principal components).

But if you have payed attention in the training set were also included the staurosporine and PF3758309 molecules that were used as reference drug in the article.

Therefore it could be assumed that those two molecules being quite different from the new molecules they should be removed from the training set.

Preparing the dataset

The screenshot displays the 'Online chemical database' interface. The 'Database' tab is active, and the 'Baskets' menu item is highlighted with a blue arrow. The interface shows a list of baskets, with 'pk1_PLS_Dragon7 (bloc)' selected. A table below the list provides statistical data for the training set:

Data Set	#	R2	q2	RMSE	MAE
Training set: jm7b01342_sl_002_als	40 records	0.4 ± 0.2	0.3 ± 0.3	0.7 ± 0.2	0.46 ± 0.08

Below the table is a scatter plot showing the relationship between two variables, with a regression line. The plot has a y-axis ranging from 7.0 to 8.0 and an x-axis ranging from 0.0 to 1.0. The data points are red dots, and a blue regression line is drawn through them.

Pagina 59

To remove molecules from the training set it is possible to create a selection of the original dataset and work on it.

To do that click on “Database” tab and select baskets

Preparing the dataset

The screenshot shows the 'Basket browser' interface of the Online Chemical Database. The page title is 'Online chemical database with modeling environment'. The user is logged in as 'Dear Prof Ragno!'. The interface includes a search bar, a 'Filter by name' field, and a 'Create new' button. A table lists baskets with columns for 'Selected records' and 'pending models'. The basket 'jm7b01342_sl_002' is highlighted in green, showing 40 records and 1 pending model. A blue arrow points to the basket name, and another blue arrow points to the '40 records' value. A tooltip 'Open basket profile' is visible over the basket name.

Basket Name	Selected records	pending models
Selected records	0 records	
jm7b01342_sl_002	40 records	1 pending models

Pagina 60

The Baskets (Datasets) will be listed. Let's click on the one we are working on ...

Preparing the dataset

The screenshot displays the 'Basket editor' interface of the Online chemical database. The page title is 'Online chemical database with modeling environment'. The user is logged in as 'Prof Ragno!'. The interface includes a navigation menu with 'Home', 'Database', and 'Models'. The main content area is titled 'Basket editor' and contains a form for creating a new basket. The form has three input fields: 'Name' (containing 'jm7b01342_sl_002.xls'), 'Description (optional)', and 'Excluded implicit records (under development)'. Below the form is an 'Actions' section with several options: 'Create a copy of this basket', 'Create a primary records basket', 'Add or delete particular records', 'Discretize the numerical values', 'Models summary for 1 models', 'Split the basket into two sets', 'Transform the basket using OScript', and 'Export this basket into Excel, CSV or PDF'. The 'Statistics of the basket' section shows 'Properties' as 'pkf' and 'Records' as '40 records' and '40 compounds'. A blue arrow points to the 'Records' link, which is highlighted in blue. The URL in the browser address bar is 'https://chem.eu/ligandscout/show.do?basket/select=189582&property=481947'.

Pagina 61

... and after the windows have changed to the one similar to the image click on the "records" link

Preparing the dataset

The screenshot displays the 'Online chemical database' interface. The main content area shows a list of molecules with their pKa values and chemical structures. The list is filtered to show records 1-5 of 40. The molecules listed are:

- RecordID: R38465835, pKa = 7.443697499232712 (in -log(mol/L)), Ragnio, R, jm7001342_sl_002.xls, R: AUTO_40, MoleculeID: M57133554, Private record.
- RecordID: R38465834, pKa = 7.958607314841775 (in -log(mol/L)), Ragnio, R, jm7001342_sl_002.xls, R: AUTO_39, MoleculeID: M57133553, Private record.
- RecordID: R38465834, pKa = 7.795880017344075 (in -log(mol/L)), Ragnio, R, jm7001342_sl_002.xls, Private record.

Two blue arrows point to the right side of the interface, indicating the scroll bar area.

Pagina 62

... the window will now show all the records (molecules) in groups of 5.
Slide down ...

Preparing the dataset

The screenshot displays the 'Online chemical database' interface. On the left, there are search filters including 'Molecular mass' (set to 'between' and 'and'), 'MISCELLANEOUS' (Current test: jm701342_sl_002.xls), 'Data origin and quality' (Data origin: All users, Data visibility: All data), and 'Discover issues with the data' (Error records, Empty molecules, etc.). The main area shows a list of records with chemical structures and pKa values. A blue arrow points to the 'Items on page' dropdown menu, which is currently set to 5. The bottom of the page shows a dark red bar with the text 'Pagina 63'.

... and change it from 5 to 50 (a number greater of your molecule number)

Preparing the dataset

The screenshot displays the 'Online chemical database' interface. On the left, there are filter sections for 'SOURCE' (Article/Source), 'PROPERTY' (Activity/Property), and 'MOLECULE FILTERS'. The 'PROPERTY' section has 'pKi' selected. The main area shows a list of molecules with their chemical structures, names, and pKi values. A blue arrow points to a green double-check icon in the top right of the molecule list, indicating the selection of all records.

RecordID	pKi (in -log(mol/L))	Record Name
R39465635	7.44369/499252/112	Ragno, R jm7001342_sl_002.xls N: AUTO_40
R39465634	7.958607314841775	Ragno, R jm7001342_sl_002.xls N: AUTO_39
R39465633	7.795880017344075	Ragno, R jm7001342_sl_002.xls N: AUTO_38

Pagina 64

Now you have all the 40 molecules displayed (sorted in reverse mode from the last to the first), look for the staurosporine (AUTO_1) and PF3758309 (AUTO_2) molecules (they should be in the bottom. But first select all the record by clickn on the double check green icon

Preparing the dataset

The screenshot displays the 'Online chemical database' interface. The main content area shows a list of molecules with their pKa values and associated records. The molecules are:

- RecordID: R38465798, pKa = 6.1487416512809245 (in -log(mol/L)), MoleculeID: M57133547
- RecordID: R38465797, pKa = 7.585026652029182 (in -log(mol/L)), MoleculeID: M55119959
- RecordID: R38465796, pKa = 8.045757490560675 (in -log(mol/L)), MoleculeID: M4430773

Each molecule entry includes a chemical structure, a 'molecule profile' link, and a 'Private record' status. Two blue arrows point to the 'Private record' status boxes for the first and second molecules, indicating the action of clicking to deselect them.

Then slide down and deselect the two molecules by clicking in the little boxes as shown

Preparing the dataset

The screenshot displays the Online Chemical Database (OCDB) interface. The browser window shows the URL <https://ocdb.eurfx.it/ocdb/show.do>. The page title is "Online chemical database" with the tagline "an open modeling environment". The user is logged in as "Prof. Ragnoli".

The main content area is titled "Compounds properties browser" and shows a list of 40 records. The first three records are visible:

- Record 1: **pKa = 7.443697499232712 (in -log(mol/L))**
Ragno, R
jm7001342_sl_002.xls
N: AUTO_40
RecordID: R39465635
MoleculeID: M97133554
Private record
- Record 2: **pKa = 7.958607314841775 (in -log(mol/L))**
Ragno, R
jm7001342_sl_002.xls
N: AUTO_39
RecordID: R39465634
MoleculeID: M97133553
Private record
- Record 3: **pKa = 7.795880017344075 (in -log(mol/L))**
Ragno, R
jm7001342_sl_002.xls
N: AUTO_38

Each record includes a chemical structure diagram and a "molecule profile" link. The interface also features a sidebar with filters for SOURCE, PROPERTY, and MOLECULE FILTERS.

Pagina 66

Now click on the “Molecule sets” tab and ...

Preparing the dataset

The screenshot displays the 'Online chemical database with modeling environment' interface. The 'Basket browser' section shows a table of records. Two blue arrows point to the 'Selected records' column and the '38 records' value. The table lists two records: 'Selected records' (38 records) and 'jm7b01342_sl_602.xls' (40 records, 1 pending models). The interface includes a search filter, navigation links, and a taskbar at the bottom.

Filter by name:	Selected records	38 records
1-2 of 2		
	jm7b01342_sl_602.xls	40 records 1 pending models
1-2 of 2		

Pagina 67

... the "Selected records" will indicate 38 records.
Click on it ...

Preparing the dataset

The screenshot displays the 'Basket editor' interface of the Online Chemical Database. The page title is 'Preparing the dataset'. The browser address bar shows 'https://ochem.pdsc.it/show.do'. The navigation bar includes 'Home', 'Database', and 'Models'. The 'Molecule sets' tab is highlighted. The 'Basket editor' section contains the following elements:

- Name:** A text input field containing '38 Mols' with a note '(min. 2 characters)'. A blue arrow points to this field.
- Description (optional):** A text input field.
- Excluded implicit records (under development):** A text input field.
- Actions:** A list of actions including 'Create a copy of this basket', 'Split the basket into two sets', 'Create a primary records basket', 'Transform the basket using OSCript', 'Add or delete particular records', 'Export this basket into Excel, CSV or SDF', and 'Discretize the numerical values'.
- Statistics of the basket:** A summary showing '38 records' and '38 compounds'. A 'Show MMP's' button is visible.

The bottom of the browser window shows several open files: 'LigandScout_4...dmg', 'LigandScout_4...exe', 'LigandScout...tar.gz', and 'jm7b01342_sl.0...csr'. A 'Show all' button is located at the bottom right of the file list.

Pagina 68

And rename it as you like and click again on the “Molecule sets” tab

Preparing the dataset

The screenshot shows the Online Chemical Database interface. The browser address bar displays <https://ochem.eu/basket/>. The page header includes the logo for 'Online chemical database' and the text 'Welcome, Dear Prof Ragno! My account Logout'. Below the header, there is a 'Basket browser' section with the subtext 'Browse, Compare or Join molecule sets'. A 'Filter by name:' input field is present, with a dropdown menu showing 'Create new' and a 'Show public sets' button. Below the filter, a table lists two molecule sets:

Set Name	Records	Pending Models
38 Mols	38 records	0 pending models
jm7b01342_v_602.xls	40 records	1 pending models

The '38 Mols' set is highlighted in green. A blue arrow points to the '38 Mols' set, and another blue arrow points to the 'jm7b01342_v_602.xls' set. The bottom of the browser window shows several open files, including 'LigandScout_4...dmg', 'LigandScout_4...exe', 'LigandScout...tar.gz', and 'jm7b01342_v_602.csv'. A 'Show all' button is visible in the bottom right corner of the browser window.

Pagina 69

Now create a model using the “38 Mols” set

Preparing the dataset

Online chemical database
with modeling environment

Welcome, Dear Prof Ragnol [My account](#) [Logout](#)

Home Database Models

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): 38 Mols [\[details\]](#)
Add a validation set

The model will predict this property:
pK_a using unit [_log\(mol/L\)](#)

Skip model configuration and use the predefined settings

Choose the learning method:

Suggested modeling methods:

- ASNN: Associative Neural Networks
- CHEMCHAINER: Chainer Chemistry models (GPU)
- CHF - Convolutional Neural Network Fingerprint (GPU)
- Consensus model (based on models developed for the same set)
- DEEPCHEM: several methods from DeepChem (GPU)
- DNN: Deep Neural Network (GPU)
- EAGCHG - Edge Attention based Multi-relational Graph Convolutional Networks
- FSMLR: Fast Stagewise Multiple Linear Regression
- KNN: k - Nearest Neighbors
- Library model (A local bias correction model based on another ASNN model)
- LSSVM: gnd-search parameter optimisation
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: Multiple Linear Regression

LigandScout_A...dmg LigandScout_A...exe LigandScout...tar.gz jm7b01342_sl.0...csv [Show all](#)

Pagina 70

Everything should be already set, but CHECK IT!

Preparing the dataset

The screenshot displays the 'Online chemical database' web application. The page title is 'Online chemical database with modeling environment'. The user is logged in as 'Dear Prof Ragnol'. The main content area is titled 'Models' and lists various machine learning models for selection. A large blue arrow points to the 'Next>>' button at the bottom of the model selection area. Another blue arrow points to the 'N-Fold cross-validation' dropdown menu under the 'Model validation' section.

Models available for selection:

- CHEMCHARIC: Chainer Chemistry models (GPU)
- CHN - Convolutional Neural Network Fingerprint (GPU)
- Consensus model (based on models developed for the same set)
- DEEPCHEM: several methods from DeepChem (GPU)
- DBN: Deep Neural Network (GPU)
- EAGCND - Edge Attention based Multi-relational Graph Convolutional Networks
- FSMLR: Fast Stagewise Multiple Linear Regression
- KNN: k - Nearest Neighbors
- Library model (A local bias correction model based on another ASH/N model)
- LSSVM: grid search parameter optimisation
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: Multiple Linear Regression
- PLS: Partial Least Squares
- RF: Random Forest regression and classification
- WEKA J48: Weka C4.5 decision trees, only classification - use with bagging
- WEKA RF: Random Forest, only classification
- XGBoost: Scalable and Flexible Gradient Boosting

Methods under development:

Model validation

Validation method: **N-Fold cross-validation**

Number of folds: 5

Stratified cross-validation (classification only)

Consider each record as a molecule

You can create a model from template: import an XML model template or use another model as a template

Next>>

Pagina 71

Click on "Next>>"

Preparing the dataset

The screenshot displays the 'Model creator' section of the 'Online chemical database' website. The page title is 'Select the preferred data preprocessing options'. Under the heading 'Preprocessing of molecules (Chemaxon)', there are four checked options: 'Standardization', 'Neutralize', 'Remove salts', and 'Clean structure'. At the bottom of the form, there are two buttons: '<<Back' and 'Next>>'. Two blue arrows are overlaid on the image: one points to the 'Next>>' button, and another points to the 'Clean structure' option.

Pagina 72

Click on "Next>>"

Preparing the dataset

Model creator
Select model template and training set

Select the molecular descriptors

Recommended descriptor types

- E-state
- ALogPS (2)
- GSFfragment (1138)
- CDK 2.0 descriptors (256/3D)
- Dragon v 7 (5270/3D)

(select all) (select none)

- Constitutional descriptors (47)
- Topological indices (75)
- Connectivity indices (37)
- 2D matrix based descriptors (607)
- Burden eigenvalues (96)
- ETA indices (23)
- Geometrical descriptors (3D, 38)
- 3D autocorrelations (3D, 80)
- 3D-MoRSE descriptors (3D, 224)
- OE TAWAY descriptors (3D, 273)
- Functional group counts (3D, 154)
- Atom-type E-state indices (172)
- 2D Atom Pairs (1596)
- Charge descriptors (3D, 15)
- Drug-like indices (28)
- Ring descriptors (32)
- Walk and path counts (46)
- Information indices (50)
- 2D autocorrelations (213)
- P_VSA-like descriptors (55)
- Edge adjacency indices (324)
- 3D matrix-based descriptors (3D, 99)
- RDF descriptors (3D, 210)
- WHIM descriptors (3D, 114)
- Randic molecular profiles (3D, 41)
- Atom-centred fragments (115)
- CATS 2D (150)
- 3D Atom Pairs (3D, 36)
- Molecular properties (20)
- CATS 3D (3D, 300)

Predictions by OCHEM's featured models

- Ames levenberg
- Toxicity against *Pythium*
- ALogPS 3
- CYP1A2 *Estate+ALogPS*
- CYP2C9 *Estate+ALogPS*
- CYP2D6 *Estate+ALogPS*
- CYP3A4 *Estate+ALogPS*
- Pyrolysis point prediction (best Estate)
- Water solubility prediction (best Estate)
- Melting Point prediction (best Estate)
- Water solubility model based on logP and Melting Point
- ALogPS 2.1 logP
- ALogPS 2.1 logS

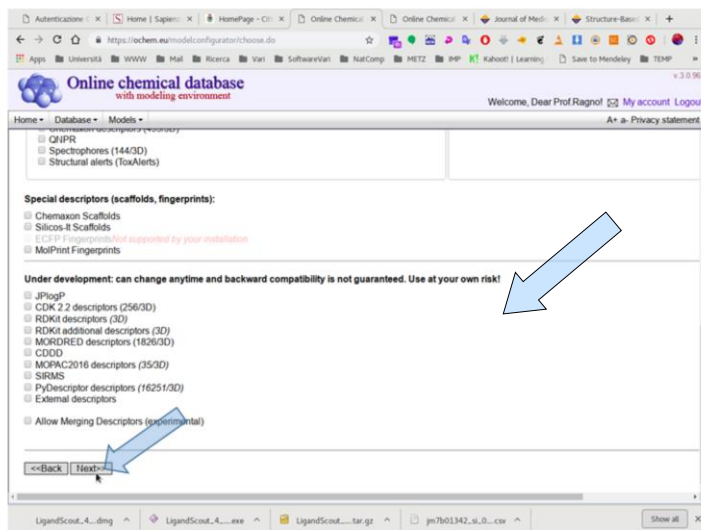
Obsolete/Additional descriptor types

- Outputs of other OCHEM models
- CDK 1.4.11 descriptors (256/3D)
- OEState
- Dragon v 5.4 (1644/3D)
- Dragon v 5.5 (3224/3D)

LigandScout_4..._dmg LigandScout_4..._exe LigandScout...tar.gz jm7b01342_sl_0..._csw Show all

Click on "Next>>"

Preparing the dataset



Pagina 74

Click on "Next>>"

Preparing the dataset

Model creator
Select model template and training set

Select a tool to optimize molecule structures

- No optimisation
- Optimise with Corina
- Optimise with OpenBabel
- Optimise with OBGEN (part of OpenBabel distribution)
- Optimise with BALLOON

<<Back Next>>

Pagina 75

Click on “Next>>”

Preparing the dataset

The screenshot shows the 'Model creator' interface of the 'Online chemical database with modeling environment'. The page is titled 'Model creator' and includes a sub-header 'Select model template and training set'. Under 'Select filters of descriptors', there are several options with checkboxes and input fields: 'Eliminate descriptors with less than 2 unique values', 'Delete descriptors that have absolute values larger than 999996', 'Delete descriptors that have variance smaller than 0.01', 'Group descriptors, that have pair-wise correlations Pearson's correlation coefficient R larger than 0.95', and 'Use Unsupervised Forward Selection to delete variables using the above value of multiple correlation coefficient R '. Below this is the 'Normalisation parameters' section, which includes 'Descriptors normalization: Standardize (zero mean and unit variance)' and 'Values normalization: Do not normalize'. At the bottom of the form are '<<Back' and 'Next>>' buttons. A blue arrow points to the 'Next>>' button.

Pagina 76

Click on "Next>>"

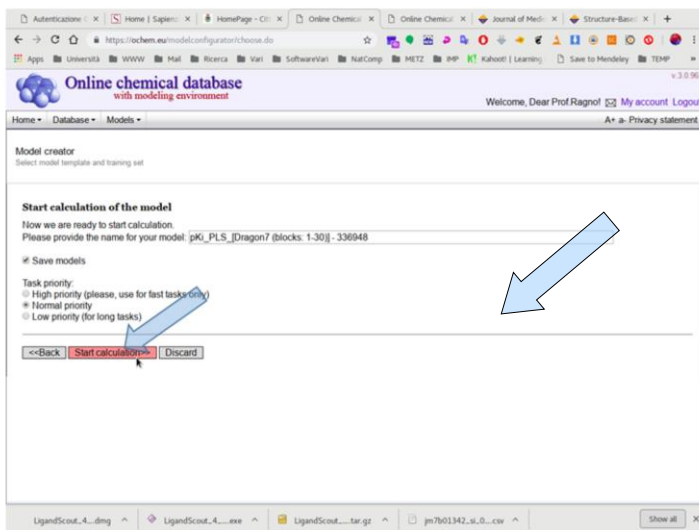
Preparing the dataset

The screenshot shows a web browser window with the URL <https://chem.esi-model.configurator.chimie.de>. The page title is "Online chemical database with modeling environment". The user is logged in as "Dear Prof Ragnol". The main content area is titled "Model creator" and contains the "Configure PLS method" section. This section includes a "Number of latent variables" input field set to "0", a checked option for "Optimize the number of latent variables automatically", and an unchecked option for "Limit predicted values to the training set range". At the bottom of the form are two buttons: "<<Back" and "Next>>". A blue arrow points to the "Next>>" button. The browser's taskbar at the bottom shows several open files, including "LigandScout_4...dmg", "LigandScout_4...exe", "LigandScout...tar.gz", and "jm7b01342_sl.0...csw".

Pagina 77

Click on "Next>>"

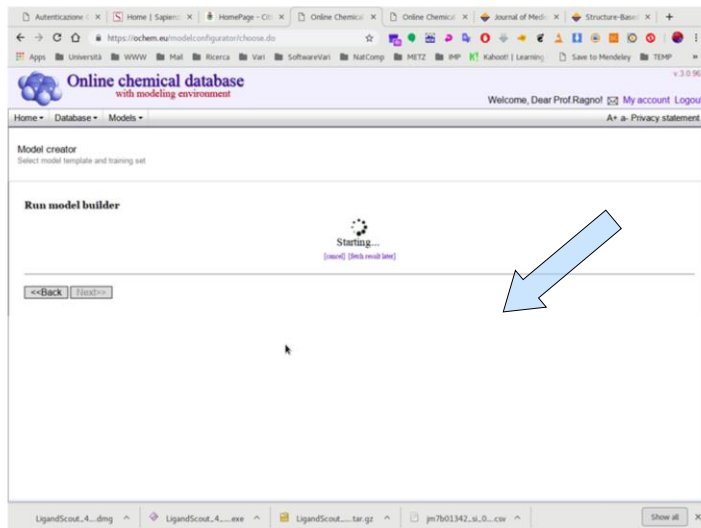
Preparing the dataset



Pagina 78

Click on "Next>>"

Preparing the dataset



Wait for the model

Preparing the dataset

The screenshot shows a web browser window with the URL <https://chem.esi-model.configurator.chimie.it>. The page header includes the logo for 'Online chemical database with modeling environment' and a user greeting: 'Welcome, Dear Prof Ragnoli'. Below the header, there are navigation tabs for 'Home', 'Database', and 'Models'. The main content area is titled 'Run model builder' and displays the status: 'Finished posting ... - Processing task, Corina - Waiting for a free server -- 09:37'. A blue arrow points to this status message. At the bottom of the main content area, there are two buttons: '<<Back' and 'Next>>'. The browser's taskbar at the bottom shows several open files related to 'LigandScout'.

Pagina 80

Wait for the model

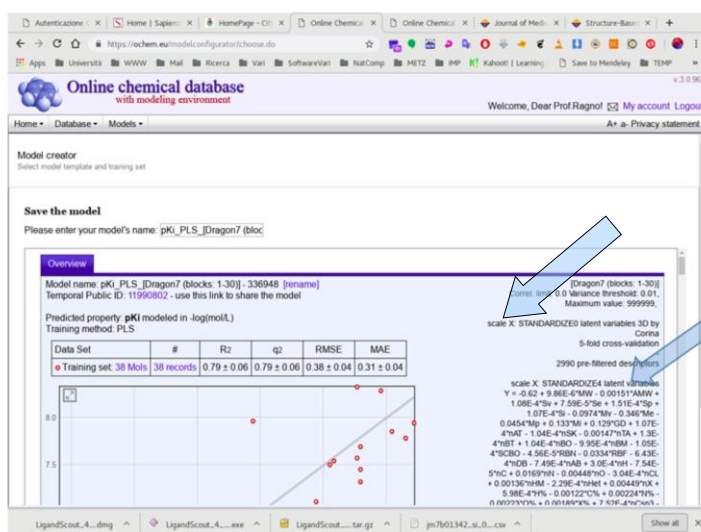
Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.es/modelconfigurator/home.do>. The page title is "Online chemical database with modeling environment" and the version is "v.3.0.96.1". The user is logged in as "Prof Ragnoli". The main content area is titled "Run model builder" and displays the message: "Finished posting ... - Processing task Descriptors - Tasks are sent for calculations -- 09:37". Below this message are two buttons: "<<Back" and "Next>>". A blue arrow points to the "Next>>" button. The browser's taskbar at the bottom shows several files related to "LigandScout" and "jm7b01342_sl.0...".

Pagina 81

Wait for the model

Preparing the dataset



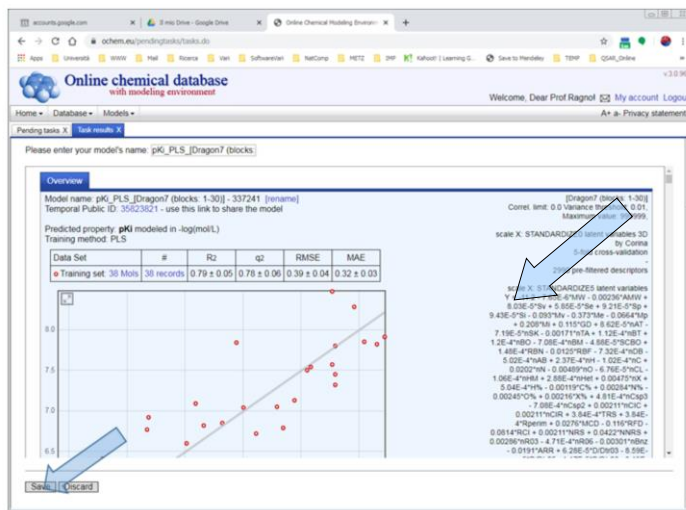
Pagina 82

Once the model is ready you can have the results.

In the right you have the QSAR equation and on the bottom (not shown) you can download everything in an excel file and inspect it as youlike.

Slide down and save the model!

Preparing the dataset



Pagina 83

Once the model is ready you can have the results.

In the right you have the QSAR equation and on the bottom (not shown) you can download everything in an excel file and inspect it as youlike.

Slide down and save the model!

Preparing the dataset

The screenshot displays a web browser window with the URL <https://chem.esi.model.configurator.chimie.de>. The page title is "Online chemical database with modeling environment" and the version is "v.3.0.96.1". The user is logged in as "Prof Ragnol". The main content area is titled "Model creator" and contains the following text:

Your model has been saved
Thank you for your cooperation.
Your next possible actions are:
[Apply your model](#)
[View your model's properties](#)

Two blue arrows are overlaid on the page: one points to the "View your model's properties" link, and the other points to the right side of the page.

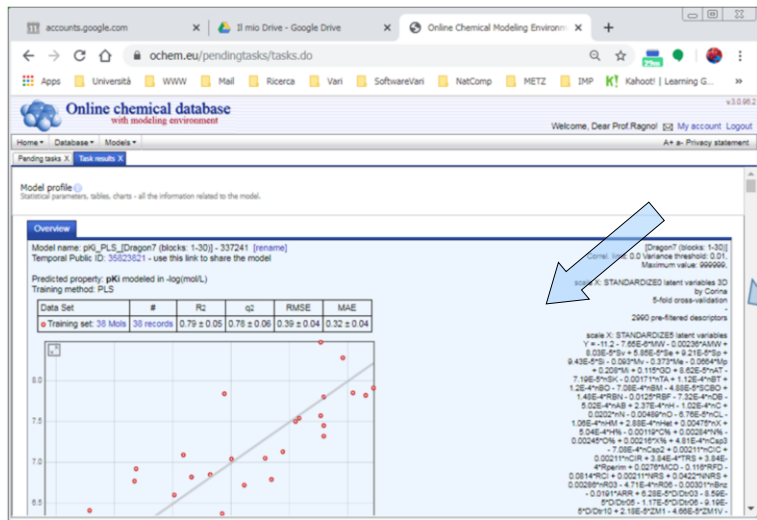
Pagina 84

You have created a QSAR model by means of PLS.

You can go back to your model later on to check the details.

Click on "View your model's properties"

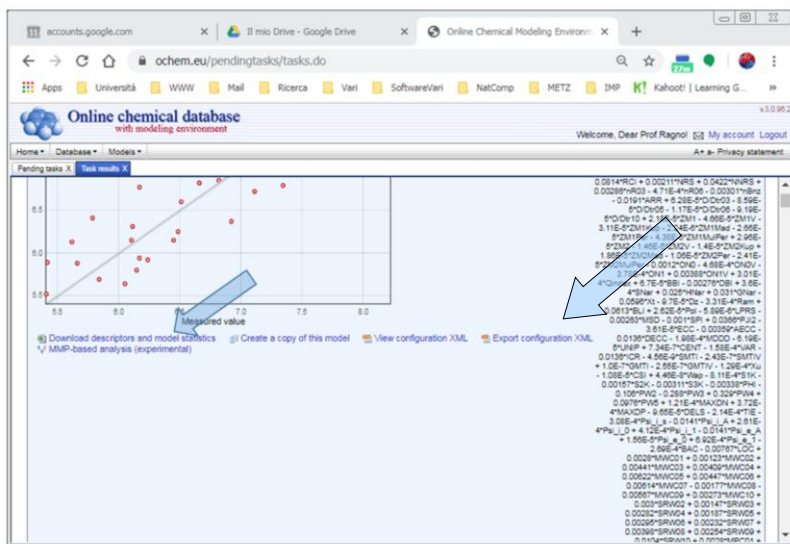
Preparing the dataset



Pagina 85

The model again will be shown and slide down to look for a excel file to be downloaded

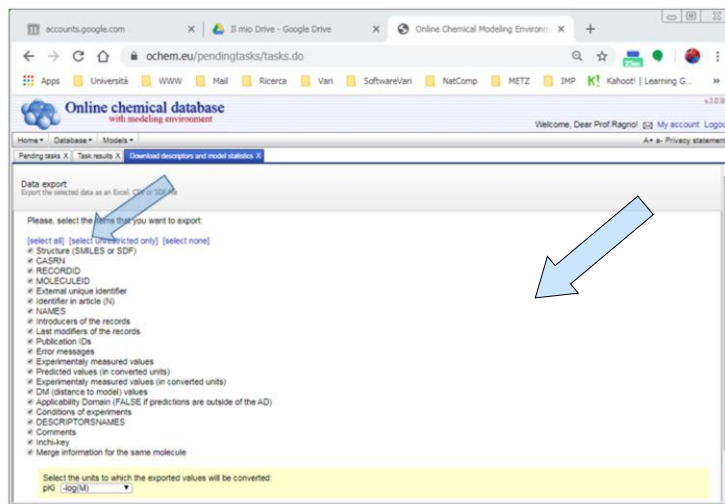
Preparing the dataset



Pagina 86

The model again will be shown and slide down to look for a excel file to be downloaded

Preparing the dataset



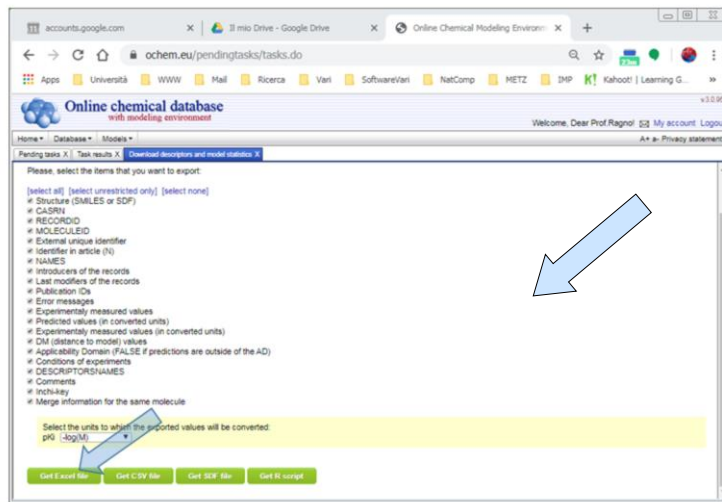
Pagina 87

The model again will be shown.

The portal will ask what you want to save, just select everything

Slide down to download the file

Preparing the dataset



The screenshot shows the 'Online chemical database' interface. The main content area is titled 'Please, select the items that you want to export:' and contains a list of checkboxes for various data fields. A blue arrow points to the 'Get Excel file' button at the bottom left. Another blue arrow points to the 'Select the units to which the reported values will be converted' dropdown menu, which is currently set to 'pH _log10_'. The interface also includes a navigation bar with 'Home', 'Database', and 'Models' menus, and a user profile section with 'Welcome, Dear Prof Ragnoli', 'My account', and 'Logout' links.

Online chemical database
with modeling environment

Please, select the items that you want to export:

- [select all] [select unrestricted only] [select none]
- Structure (SMILES or SDF)
- CASRN
- RECORDID
- MOLECULED
- External unique identifier
- Identifier in article (N)
- NAMES
- Introducers of the records
- Last modifiers of the records
- Publication Cts
- Error messages
- Experimentally measured values
- Predicted values (in converted units)
- Experimentally measured values (in converted units)
- DM (distance to model) values
- Applicability Domain (FALSE if predictions are outside of the AD)
- Conditions of experiments
- DESCRIPTORSNAMES
- Comments
- InChIkey
- Merge information for the same molecule

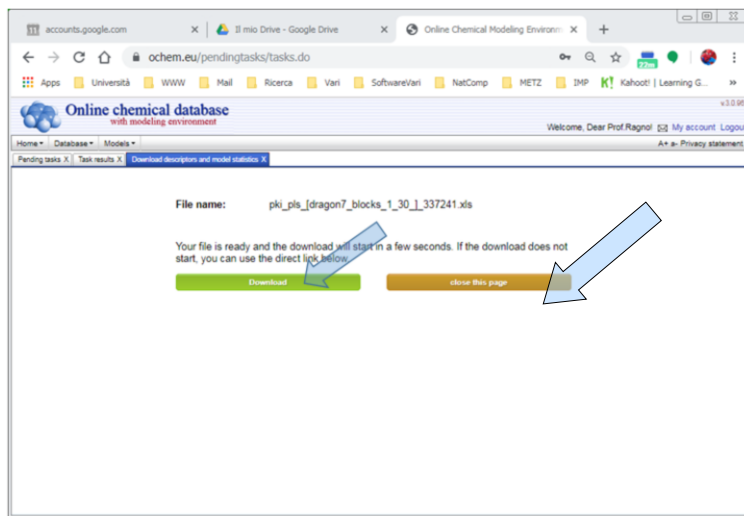
Select the units to which the reported values will be converted
pH _log10_

Get Excel file Get CSV file Get XML file Get R script

Pagina 88

Click on "Get Excel File"

Preparing the dataset



Pagina 89

Then click on “Download” and you can open it with excel or libreoffice to inspect the results