

# 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>

# Preparing the dataset

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

Autenticazione | Home | Sapienz... | HomePage - Citt... | Online Chemical... | Online Chemical... | Journal of Medic... | Structure-Based... | +

https://pubs.acs.org/doi/10.1021/acs.jmedchem.7b01342

ACS Publications C&EN CAS Access provided by UNIVERSITA DI ROMA LA SAPIENZA Log In

ACS Publications Most Trusted. Most Cited. Most Read.

Search text, DOI, authors, etc.

My Activity Publications

RETURN TO ISSUE | < PREV ARTICLE NEXT >

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

Chenzhou Hao<sup>1</sup>, Fan Zhao<sup>1</sup>, Hongyan Song<sup>1</sup>, Jing Guo<sup>1</sup>, Xiaodong Li<sup>1</sup>, Xiaolin Jiang<sup>1</sup>, Ran Huan<sup>1</sup>, Shuai Song<sup>1</sup>, Qiaoling Zhang<sup>1</sup>, Ruifeng Wang<sup>1</sup>, Kai Wang<sup>1</sup>, Yu Pang<sup>1</sup>, Tongchao Liu<sup>1</sup>, Tianqi Lu<sup>2</sup>, Wanxu Huang<sup>1</sup>, Jian Wang<sup>1</sup>, Bin Lin<sup>1</sup>, Zhonggui He<sup>1</sup>, Haitao Li<sup>1</sup>, Feng Li<sup>2</sup>, Dongmei Zhao<sup>1\*</sup> and Maosheng Cheng<sup>1\*</sup>

**View Author Information**

Cite this: *J. Med. Chem.* 2018, 61, 1, 265-285  
Publication Date: November 30, 2017  
https://doi.org/10.1021/acs.jmedchem.7b01342  
Copyright © 2017 © 2017 American Chemical Society

Article Views 6322 Altmetric 4 Citations -

LEARN ABOUT THESE METRICS

Share Add to Export

Journal of Medicinal Chemistry

PDF (8 MB) Supporting Info (2)

## 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 PAKs, we chose a 2,4-diaminoquinazoline 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-aminoquinazoline-2-carboxamide PAK4 inhibitors were designed and synthesized. The inhibitors' selectivity, therapeutic potency, and pharmaceutical properties were optimized. One of the best compounds, **31** (CZh226), showed remarkable PAK4 selectivity (346-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 investigation of group II PAKs.



This website uses cookies to improve your user experience. By continuing to use the site, you are accepting our use of cookies. Read the ACS privacy policy. CONTINUE

# Preparing the dataset

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

Autenticazione | Home | Sapientz | HomePage - Citt | Online Chemical | Online Chemical | Journal of Medic | Structure-Based

https://pubs.acs.org/doi/10.1021/acs.jmedchem.7b01342

Journal of Medicinal Chemistry

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

### Introduction

ARTICLE SECTIONS

The p21-activated kinases (PAKs) are serine/threonine (Ser/Thr) protein kinases that have been identified as downstream signaling effectors of Rho-family GTPases.(1,2) The six mammalian PAK isoforms are categorized into two groups: PAK1–3 (group I) and PAK4–6 (group II), based on their structural homologies and biochemical features.(3) As key components of the Ras-Rac/Cdc42-PAK pathway, PAKs have pivotal roles in many fundamental cellular processes, including cytoskeletal reorganization, focal adhesion, cell motility, morphological changes, cell-cycle progression, etc.(2) Moreover, the overexpression, amplification, and mutational activation of PAK isoforms, in particular, PAK1 and PAK4, have been linked to many human diseases, including breast cancer, lung cancer, prostate cancer, colon cancer, and human head and neck squamous cell carcinoma.(4) Consequently, PAKs have emerged as attractive targets for new anticancer therapies and have been the subject of extensive drug discovery efforts.(5)

Although the two groups of PAK proteins are similar in overall sequence and structure, they are differentiated by their tissue expression profiles,(6) subcellular localization,(7) GTPase specificity,(7) activation mechanism,(8) and downstream substrate specificity.(2,7) Studies using knockout mice lacking one or more specific PAK isoforms revealed the role of each isoform in normal tissue development, with phenotypes that range from no apparent effect to early embryonic death.(2,9–10) Among all of the PAKs, PAK4 is the most studied group II PAK member, and it has a place at critical nodal points in multiple signaling pathways that are associated with cell growth, cytoskeletal dynamics, cell polarity, survival, and development.(11) PAK4 is particularly highly expressed in prostate, testis, lung, heart, brain, and liver.(12) It has attracted considerable interest because of its role in cancer invasion, metastasis, and proliferation of BRAF- or KRAS-driven cancers.(13) In addition to PAK4, there is emerging evidence for the roles for PAK5 and PAK6 in cancer progression.(14) Moreover, a recent study revealed that PAK2 inhibition correlates with increased acute cardiovascular toxicity, which may be enhanced by PAK1 inhibition.(15) Thus, the development of specific and potent PAK4 inhibitors is highly desirable for minimizing the risk of the potential side effects associated with the inhibition of normal function of group I PAKs and will also shed further light on its role in cancer progression.

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.7b01342.

Sequence alignment of the active site residues of PAK1–6; crystallographic parameters of all X-ray structures (cocrystal structures of 10a, 30, and 31 bound to PAK4); comparison of the orientation differences of the DFG Asp (Asp458<sup>PAK4</sup>/Asp407<sup>PAK1</sup>); binding mode analysis of compounds 12, 17, 18, and 20; detailed kinase selectivity data of compound 31; <sup>1</sup>H NMR, <sup>13</sup>C NMR, and HRMS spectra of compounds 10a–d, 11a–d, and 12–41; liver microsomal stability, rat plasma stability, physicochemical properties determination, CYP450, and hERG inhibition reports of compound 31 (PDF) Molecular formula strings and associated biological data (CSV)

+ jm7b01342\_sl002.csv (4.13 kb)

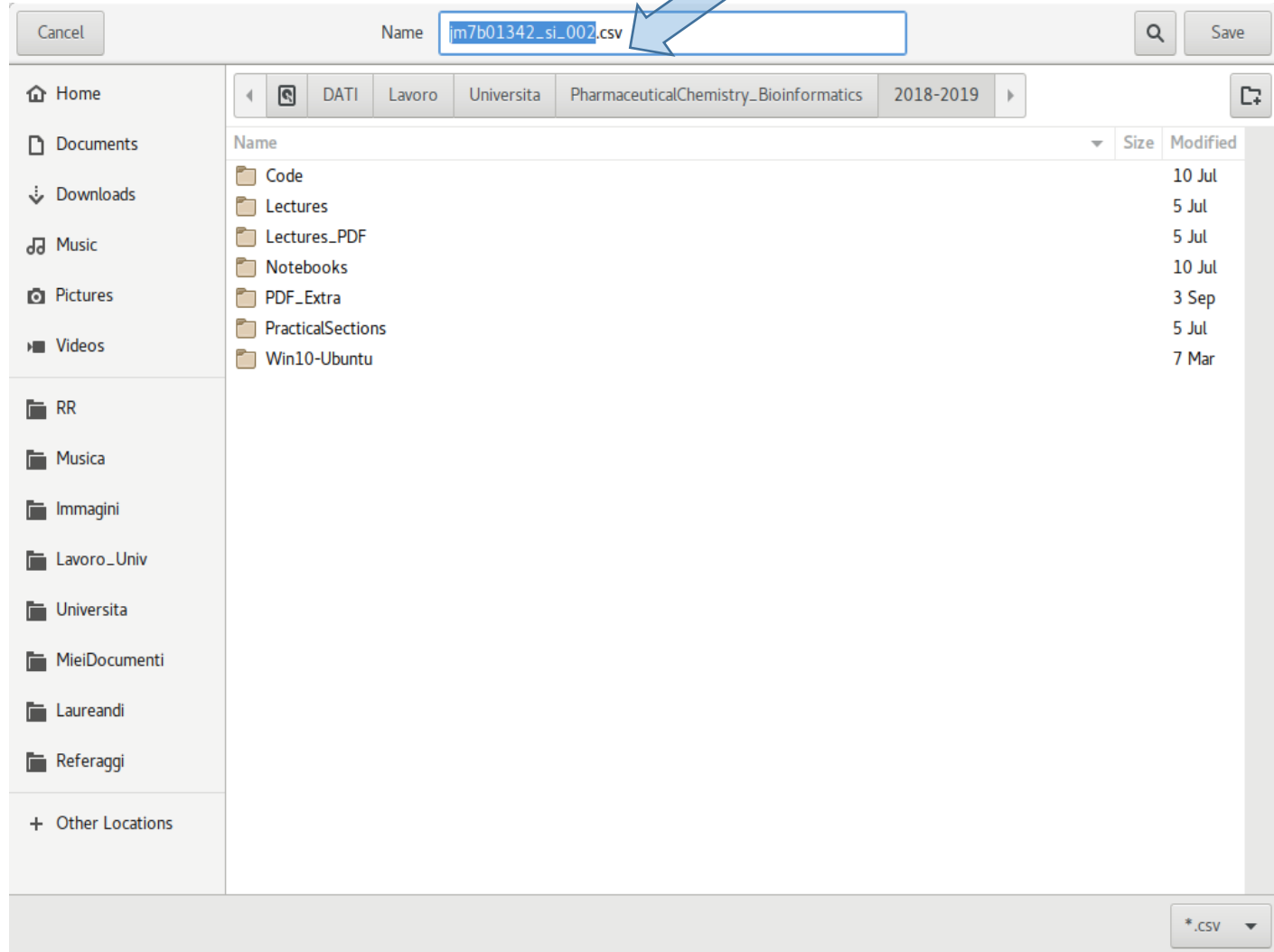
pdf

+ jm7b01342\_sl001.pdf (8.69 MB)

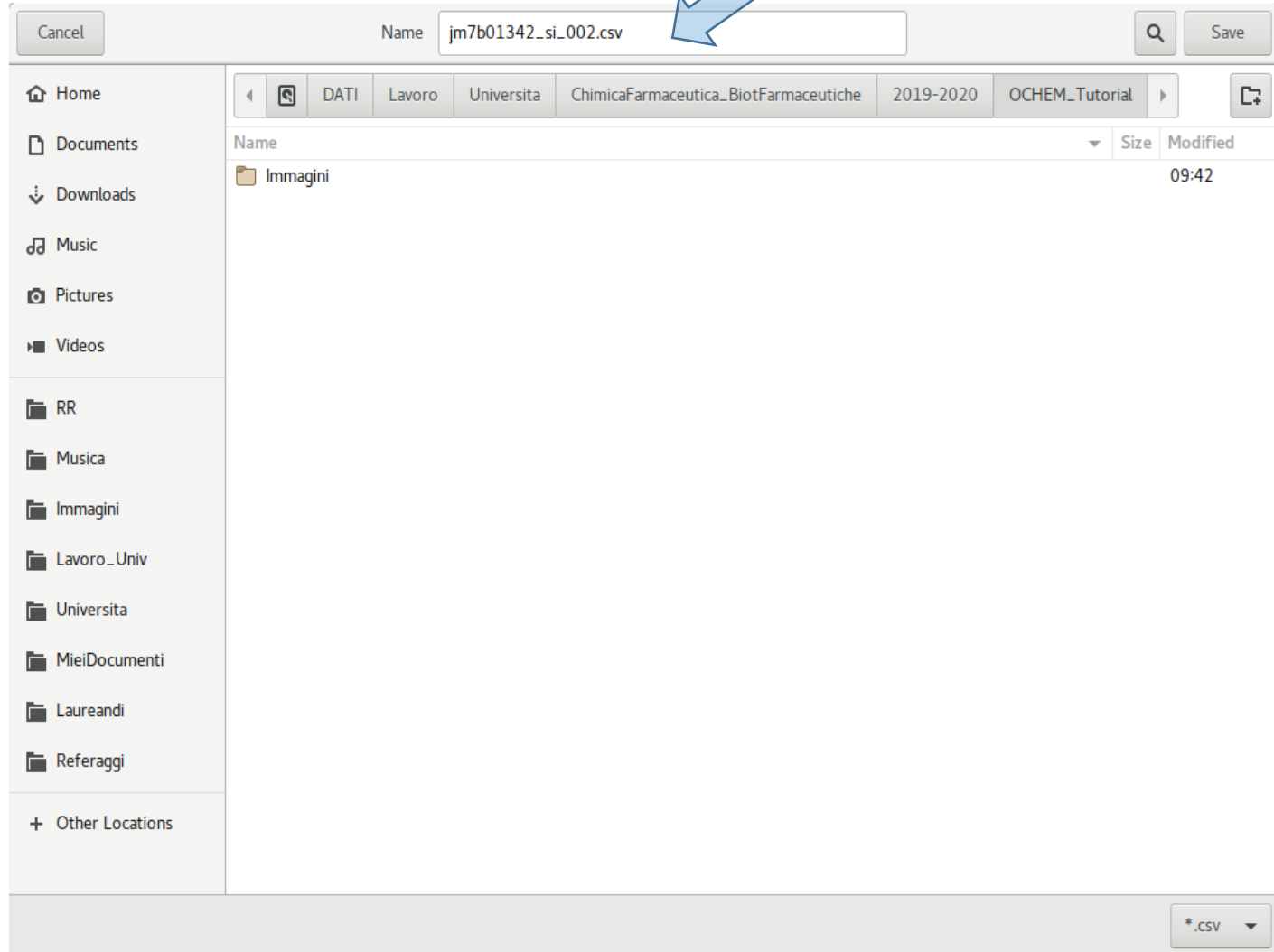
This website uses cookies to improve your user experience. By continuing to use the site, you are accepting our use of cookies. Read the ACS privacy policy. **CONTINUE**



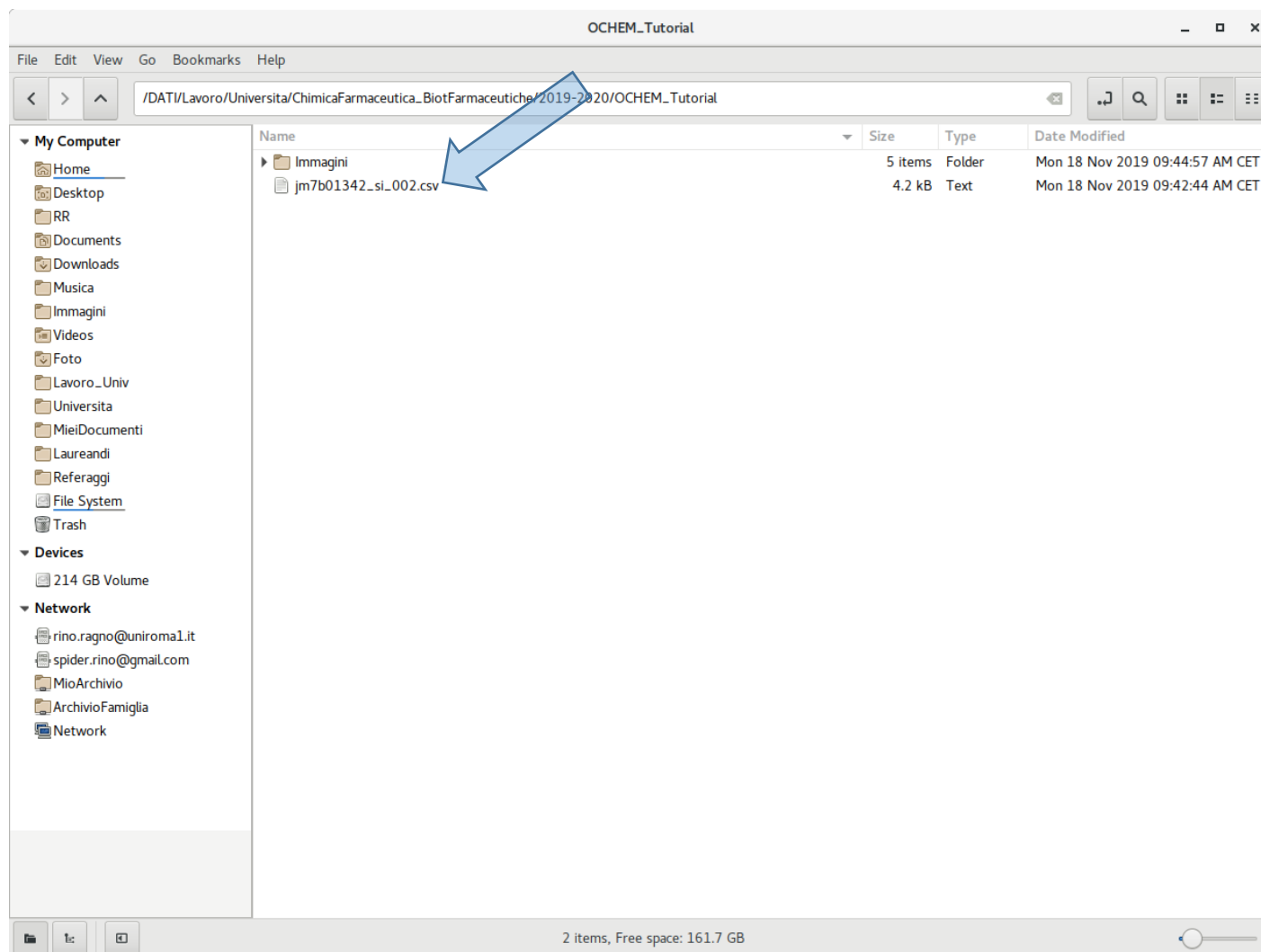
# Preparing the dataset



# Preparing the dataset



# Preparing the dataset



# Preparing the dataset

The screenshot shows a file explorer window titled "OCHEM\_Tutorial" with a menu bar (File, Edit, View, Go, Bookmarks, Help) and a search bar. The address bar shows the path: "/DATI/Lavoro/Universita/ChimicaFarmaceutica..BiotFarmaceutiche/2019-2020/OCHEM\_Tutorial". The file list shows a folder "Immagini" and a file "jm7b01342\_si\_002.csv" (4.2 kB, Text). A "Text Import" dialog box is open over the file, titled "Text Import - [jm7b01342\_si\_002.csv] (on rcmd-ProBook.rcmd.it)". The dialog has the following settings:

- Character set: Western Europe (ISO-8859-14)
- Language: English (USA)
- From row: 1
- Separator Options:  Fixed width,  Separated by,  Tab,  Comma,  Semicolon,  Space,  Other,  Merge delimiters,  Trim spaces, String delimiter: "
- Other Options:  Format quoted field as text,  Detect special numbers
- Fields: Column type: (empty)

The "Fields" section contains a table with two columns: "Standard" and "SMILES".

Standard	SMILES
1 Compound	SMILES
2 Staurosporine	O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)NSC7(C)OC([H]
3 PF3758309	O=C(N[C@@H](C1=CC=CC=C1)CN(C)N(C2(C)C)CC3=C2N
4 10a	CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C1)C=C
5 10b	C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4
6 10c	C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C)C)C=N4
7 10d	C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C
8 11a	CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C1)C=C42)=N

The status bar at the bottom indicates: "jm7b01342\_si\_002.csv" selected (4.2 kB), Free space: 161.7 GB.

# Preparing the dataset

The image shows a screenshot of a LibreOffice Calc spreadsheet titled "jm7b01342\_si\_002.csv - LibreOffice Calc (on rcmd-ProBook.rcmd.it)". The spreadsheet has two columns: "Compound" (Column A) and "SMILES" (Column B). The data is as follows:

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)CN(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(CI)C=C42</chem>
10b	<chem>CIC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C1</chem>
10c	<chem>CIC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C(C)C)C=N4)C3=C1</chem>
10d	<chem>CIC1=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(CI)C=C42)=NN1</chem>
11b	<chem>CIC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5CC5)=C4)C3=C1</chem>
11c	<chem>CIC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C(C)C)=C4)C3=C1</chem>
11d	<chem>CIC1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5=CC=CC=C5)=C4)C3=C1</chem>
12	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
13	<chem>CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
14	<chem>CC1=CC(NC2=NC(C(N3CCC(N)(C)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
15	<chem>CC1=CC(NC2=NC(C(N3CCC(O)(C)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
16	<chem>CC1=CC(NC2=NC(C(N3CCC(NC)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
17	<chem>CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
18	<chem>CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
19	<chem>CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
20	<chem>CIC1=CC=C(N=C(N=C2NC3=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)CI)=NN1</chem>
22	<chem>CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>
23	<chem>CC1=CC(NC2=NC(C(N(C)C3CNCC3)=O)=NC4=CC=C(C=C42)CI)=NN1</chem>

A blue arrow points to the SMILES string for compound 11c.

# Preparing the dataset

jm7b01342\_si\_002.csv - LibreOffice Calc (on rcmd-ProBook.rcmd.it)

File Edit View Insert Format Styles Sheet Data Tools Window Help

LibreOffice Calc toolbar

fx = PAK1 Inhibitory Activity (Ki, ?M)

	C	D	E	F	G	H	I	J
1	MW	PAK1 Inhibitory Activity (Ki, ?M)	PAK4 Inhibitory Activity (Ki, ?M)	Tmax-T0 (PAK4, FC)	clogP	PSA		
2	466	0.003	0.009	ND	3.7	69.8		
3	490	0.052	0.026	ND	4.3	98.3		
4	357.5	>10	0.71	0.5	0.0	2.6	71.3	
5	383.5	ND	0.65	2.0	5	0.0	3	70.9
6	385.5	ND	0.88	4.0	8	0.3	3.2	70.4
7	419.5	ND	1.44	0.9	0.3		3.8	71
8	357.5	1.891	0.099	2.1	0.0		2.3	86.4
9	383.5	0.288	0.016	4.7	0.3		2.9	85
10	385.5	ND	0.077	2.3	0.3		3.1	84.8
11	419.5	ND	0.256	0.7	0.3		3.5	85.1
12	385.5	ND	0.354	0.5	0.0		1.7	108.9
13	386.5	ND	3.81	7.0	1	0.3	2.1	105.3
14	399.5	ND	0.674	0.4	0.3		2.1	104.3
15	400.5	ND	3.89	-0.1	0.0		2.9	101.5
16	399.5	ND	0.672	0.3	0.3		2.3	97
17	371.5	>10	0.045	2.4	0.5		1.6	98.7
18	372.5	ND	2.172	0.2	0.3		2.2	92.5
19	385.5	ND	0.768	ND			2	88.9
20	385.5	ND	2.408	ND			0.8	132.2
21	399.5	ND	0.576	0.2	0.3		2.7	95.7
22	385.5	ND	0.325	0.2	0.3		2.4	108.8
23	399.5	ND	1.63	0.0	0.3		2.7	95.6
24	357.5	ND	0.782	0.7	0.3		2.8	111.4
25	397.5	2.75	0.016	4.1	0.3		2.2	97.4
26	385.5	>4.52	0.218	1.5	0.3		2.4	91.4
27	411.5	0.383	0.026	3.6	0.3		3	90.1
28	385.5	>4.52	0.151	1.8	0.3		1.9	92.5
29	411.5	2.08	0.017	3.6	0.3		2.5	91.1

Sheet 1 of 1 | Default | English (USA) | Average: ; Sum: 0 | 140%

# Preparing the dataset

jm7b01342\_si\_002.csv - LibreOffice Calc (on rcmd-ProBook.rcmd.it)

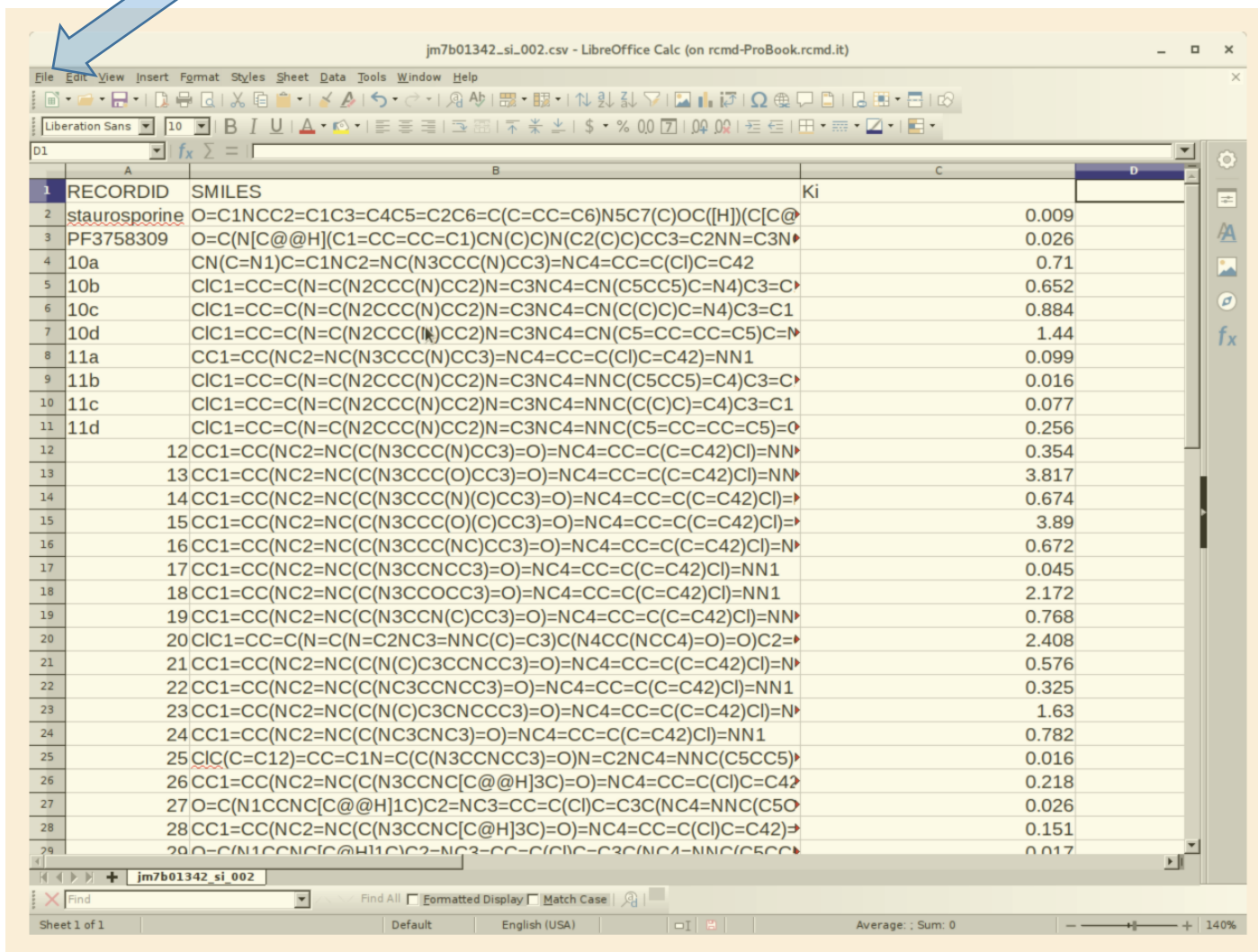
File Edit View Insert Format Styles Sheet Data Tools Window Help

LibreOffice Calc interface showing a spreadsheet with columns C through J. A blue arrow points to the header of column E, which is labeled 'Ki'. The spreadsheet contains data for various compounds, including MW, PAK1 Inhibitory Activity (Ki, ?M), Tmax-T0 (PAK4, FC), clogP, and PSA.

	C	D	E	F	G	H	I	J
1	MW	PAK1 Inhibitory Activity (Ki, ?M)	Ki	Tmax-T0 (PAK4, FC)	clogP	PSA		
2	466		0.003	0.009ND		3.7	69.8	
3	490		0.052	0.026ND		4.3	98.3	
4	357.5	>10		0.710.5f0.0		2.6	71.3	
5	383.5	ND		0.6520.5f0.0		3	70.9	
6	385.5	ND		0.8840.8f0.3		3.2	70.4	
7	419.5	ND		1.440.9f0.3		3.8	71	
8	357.5		1.891	0.0992.1f0.0		2.3	86.4	
9	383.5		0.288	0.0164.7f0.3		2.9	85	
10	385.5	ND		0.0772.3f0.3		3.1	84.8	
11	419.5	ND		0.2560.7f0.3		3.5	85.1	
12	385.5	ND		0.3540.5f0.0		1.7	108.9	
13	386.5	ND		3.8170.1f0.3		2.1	105.3	
14	399.5	ND		0.6740.4f0.3		2.1	104.3	
15	400.5	ND		3.89-0.1f0.0		2.9	101.5	
16	399.5	ND		0.6720.3f0.3		2.3	97	
17	371.5	>10		0.0452.4f0.5		1.6	98.7	
18	372.5	ND		2.1720.2f0.3		2.2	92.5	
19	385.5	ND		0.768ND		2	88.9	
20	385.5	ND		2.408ND		0.8	132.2	
21	399.5	ND		0.5760.2f0.3		2.7	95.7	
22	385.5	ND		0.3250.2f0.3		2.4	108.8	
23	399.5	ND		1.6300.0f0.3		2.7	95.6	
24	357.5	ND		0.7820.7f0.3		2.8	111.4	
25	397.5		2.75	0.0164.1f0.3		2.2	97.4	
26	385.5	>4.52		0.2181.5f0.3		2.4	91.4	
27	411.5		0.383	0.0263.6f0.3		3	90.1	
28	385.5	>4.52		0.1511.8f0.3		1.9	92.5	
29	411.5		2.08	0.0173.6f0.3		2.5	91.1	

Sheet 1 of 1 | Default | English (USA) | Average: 0.884; Sum: 0.884 | 140%

# Preparing the dataset

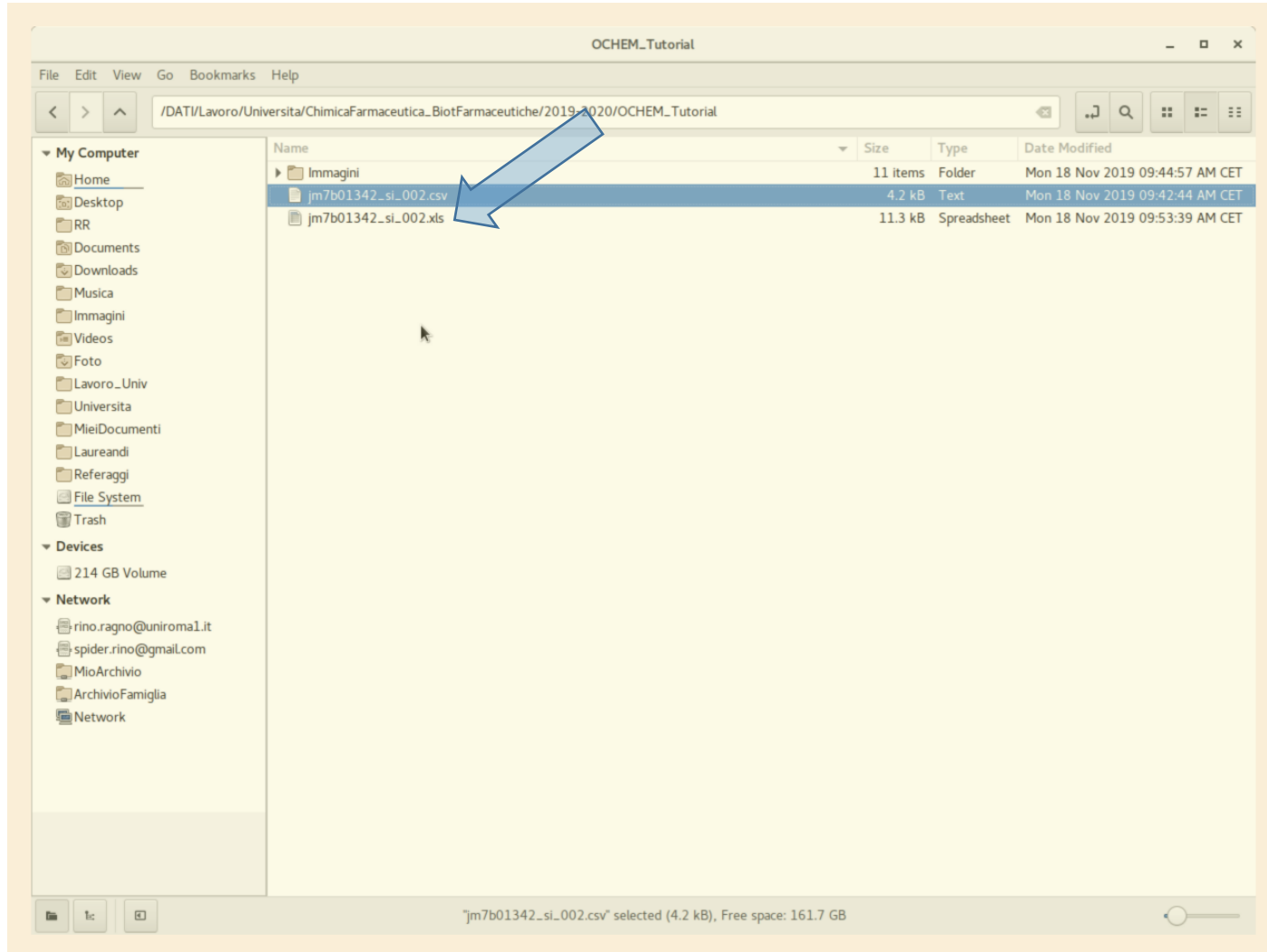


The screenshot shows a LibreOffice Calc spreadsheet with the following data:

RECORDID	SMILES	Ki
1	stauosporine	0.009
3	PF3758309	0.026
4	10a	0.71
5	10b	0.652
6	10c	0.884
7	10d	1.44
8	11a	0.099
9	11b	0.016
10	11c	0.077
11	11d	0.256
12	CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.354
13	CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	3.817
14	CC1=CC(NC2=NC(C(N3CCC(N)(C)CC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.674
15	CC1=CC(NC2=NC(C(N3CCC(O)(C)CC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	3.89
16	CC1=CC(NC2=NC(C(N3CCC(NC)CC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.672
17	CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.045
18	CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	2.172
19	CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.768
20	C1C1=CC=C(N=C(N=C2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=NN1	2.408
21	CC1=CC(NC2=NC(C(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.576
22	CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.325
23	CC1=CC(NC2=NC(C(N(C)C3CNCCC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	1.63
24	CC1=CC(NC2=NC(C(NC3CNC3)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.782
25	C1C1=CC=C(C=C1)N=C(C(N3CCNCC3)=O)N=C2NC4=NNC(C5CC5)=O	0.016
26	CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.218
27	O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C=C3)C(NC4=NNC(C5CC5)=O)C2=O	0.026
28	CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C=C4)Cl)=NN1	0.151
29	O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C=C3)C(NC4=NNC(C5CC5)=O)C2=O	0.017



# Preparing the dataset



# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | Home Page - C | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

v.3.0.96.1

log in create account

Home Database Models A+ a- Privacy statement

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**  
<https://cbianchi.eu/epidemiology/show/tutorials> to know more about the

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  
LogL(water) LogD logPI(+)

**Water solubility**  
LogL(blood) LogL(oil) ER

Cbrain/Cplasma IC50

Papp(Caco-2) Papp(MDCK)

**Oral absorption** LIC 50  
Papp ratio(Caco-2)

**Plasma protein binding**  
Papp ratio(MDCK-mdr1) pIC50

%Human FA Human IA  
Human FA

**fraction unbound (fu)**

Latest active users

- rino.ragno: Prof. Rino Ragno seconds ago
- Xingguomeng: Miss. guomeng xing seconds ago
- vkovalishyn: Dr. Vasyi Kovalishyn seconds ago
- echmstry: Mr. Ely Setiawan seconds ago
- corde: Mr. Jose Andres Cordero Solano seconds ago
- zaira1: Mrs. Zaira Rehman about an hour ago

Latest published models

- IC50 model published by carpvovv 1 hours ago
- Ld50 mouse oral model published by Tinkov\_Oleg 9 months ago
- Drug-Induced Rhabdomyolysis model published by qingshuang0501 9 months ago
- o ooo orl LD model published by

jm7b01342\_sl\_0...csv Show all


# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Citi | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +


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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

 **Online chemical database**  
with modeling environment v.3.0.96.1

[log in](#) [create account](#)

Home ▾ Database ▾ Models ▾ A+ 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

jm7b01342\_sl\_0...csv

# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Cit | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

v.3.0.96.1

log in create account

Home Database Models A+ a- Privacy statement

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 be bound by the terms of this Agreement. If you do not agree to the terms of this

**By clicking on 'I 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](#).**

I ACCEPT, CREATE MY ACCOUNT. I REJECT.

jm7b01342\_sl\_0...csv Show all


# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Citi | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

 **Online chemical database**  
with modeling environment v.3.0.96.1

[log in](#) [create account](#)

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

## 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 enter you 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](#)

jm7b01342\_sl\_0\_0...csv Show all

# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Cit | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Version: v.3.0.96.1

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**  
<https://fbtlankeu/epibioasect/dhewaticto-know-more-about-the>

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**  
LogL(water) logBB LogD logPI(+)
- Water solubility**  
LogL(blood) LogL(oil) ER  
Cbrain/Cplasma IC50
- Papp(Caco-2) Papp(MDCK)**
- Oral absorption** LIC 50  
Papp ratio(Caco-2)
- Plasma protein binding**  
Papp ratio(MDCK-mdr1) pIC50  
%Human FA Human IA  
Human FA
- fraction unbound (fu)**

Latest active users

- rino.ragno: Prof. Rino Ragno seconds ago
- vkovalishyn: Dr. Vasyil Kovalishyn seconds ago
- corde: Mr. Jose Andres Cordero Solano seconds ago
- Xingguomeng: Miss. guomeng xing 2 minutes ago
- echmstry: Mr. Ely Setiawan 8 minutes ago
- zaira1: Mrs. Zaira Rehman about 1 hours ago

Latest published models

- IC50 model published by carpovpv 1 hours ago
- Ld50 mouse oral model published by Tinkov\_Oleg 9 months ago
- Drug-Induced Rhabdomyolysis model published by qingshuang0501 9 months ago
- o oaa orl LD model published by

jm7b01342\_sl\_0...csv

Show all

# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Cit | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

Online chemical database  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

**Compound properties**

- Properties
- Conditions
- Units
- Articles/Books
- Journals
- ToxAlerts
- MatchedPairs
- Baskets
- Tags
- Set area of interest...
- User-related changes
- Batch data upload**
- Trash

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  
LogL(water) LogD logPI(+)

**Water solubility**  
LogL(blood) LogL(oil) ER

Cbrain/Cplasma IC50  
Papp(Caco-2) Papp(MDCK)  
Oral absorption LIC 50  
Papp ratio(Caco-2)

**Plasma protein binding**  
Papp ratio(MDCK-mdr1) pIC50  
%Human FA Human IA  
Human FA  
fraction unbound (fu)

**Latest active users**

- rino.ragno: Prof. Rino Ragno seconds ago
- vkovalishyn: Dr. Vasyl Kovalishyn seconds ago
- corde: Mr. Jose Andres Cordero Solano seconds ago
- Xingguomeng: Miss. guomeng xing 2 minutes ago
- echmstry: Mr. Ely Setiawan 8 minutes ago
- zaira1: Mrs. Zaira Rehman about 1 hours ago

**Latest published models**

- IC50 model published by carpovpv 1 hours ago
- Ld50 mouse oral model published by Tinkov\_Oleg 9 months ago
- Drug-Induced Rhabdomyolysis model published by qingshuang0501 9 months ago
- o oaa orl LD model published by

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

https://ochem.eu/batchupload30/show.do

jm7b01342\_sl\_0\_...csv

Show all

# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Cit | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

https://ochem.eu/batchupload30/show.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

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

Home | Database | Models | A+ a- 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 "\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 difficulties uploading your data, feel free to drop us an e-mail at [info@ochem.eu](mailto:info@ochem.eu).

Select a file to upload

Upload file

Choose File No file chosen

#### Settings

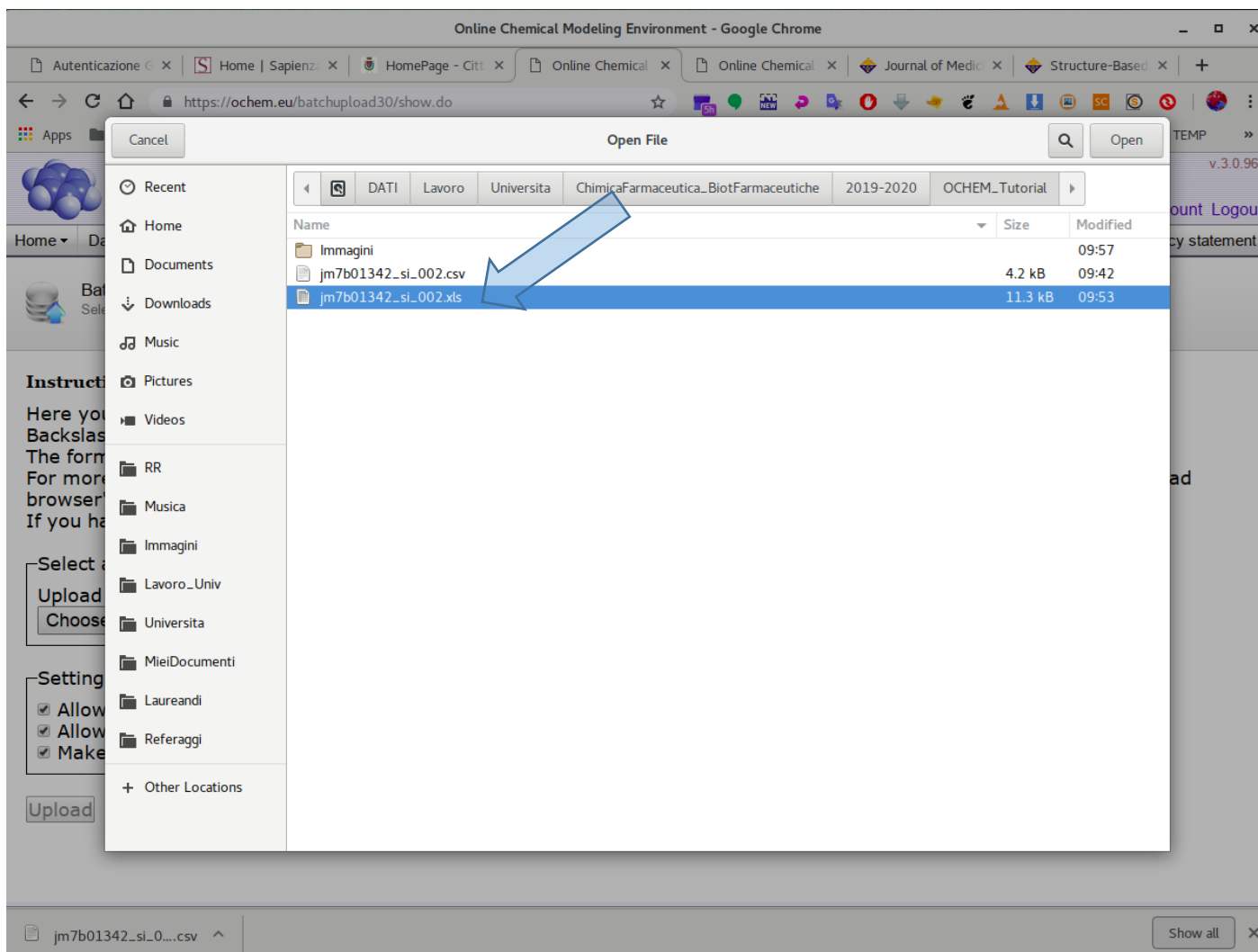
- Allow molecule lookup by name on PubChem
- Allow article lookup by PMID on PubMed
- Make the uploaded records hidden

Upload

jm7b01342\_sl\_0....csv Show all



# Preparing the dataset



# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapienz | HomePage - Citi | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

https://ochem.eu/batchupload30/show.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

Online chemical database  
with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- 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 "\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 difficulties uploading your data, feel free to drop us an e-mail at [info@ochem.eu](mailto:info@ochem.eu).

Select a file to upload

Upload file  
Choose File jm7b01342\_si\_002.xls

Settings

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

Upload

jm7b01342\_si\_0...csv Show all

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citi x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

https://ochem.eu/batchupload30/show.do

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

**Online chemical database**  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Batch Upload 3.0 - File preview and column remapping  
Preview your data, select the sheet and the columns you would like to upload

jm7b01342\_si\_002

<input checked="" type="checkbox"/> RECORDID	<input checked="" type="checkbox"/> SMILES	<input checked="" type="checkbox"/> 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 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\_si\_0...csv Show all x

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citi x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

https://ochem.eu/batchupload30/show.do

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP

**Online chemical database**  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

<input checked="" type="checkbox"/> RECORDID	<input checked="" type="checkbox"/> SMILES	<input checked="" type="checkbox"/> 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 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 need help, feel free to drop us an e-mail at [info@ochem.eu](mailto:info@ochem.eu).

Upload this sheet

Cancel Batch Upload Download Excel file


jm7b01342\_si\_0....csv Show all x

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citi x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +


https://ochem.eu/batchupload30/show.do

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP K! Kahoot! Learning Save to Mendeley TEMP >>

 **Online chemical database**  
with modeling environment v.3.0.96.1

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

Home Database Models A+ a- Privacy statement

 **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: **Ki**

Values  
Unit: **-log(M)**, min value: 0.006, max value: 3.89

Article: **unpublished**

Molecule set: **default**


jm7b01342\_si\_0....csv  x

# Preparing the dataset

Autenticazione | Home | Sapienz | HomePage - Citi | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

https://ochem.eu/batchupload30/show.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

 **Online chemical database**  
with modeling environment

v.3.0.96.1

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

Home | Database | Models | A+ a- Privacy statement

Batch upload (reloaded) X | Select unit (Ki) X

				<b>pM</b>	(Concentration)	<a href="#">Show records</a>	1 pM( picomolar ) corresponds ...	Pankaj_OCHEM
				<b>ppb</b>	(Concentration)	<a href="#">Show records</a>	parts per billion Denotes one ...	bhataba / itetko
				<b>ppb food</b>	(Concentration)	<a href="#">Show records</a>	in food - parts per billion De ...	charochkina
				<b>pph</b>	(Concentration)	<a href="#">Show records</a>	parts per hundred, one molecu ...	bhataba / admin
				<b>ppm</b>	(Concentration)	<a href="#">Show records</a>	Parts per million ('ppm') deno ...	itetko
				<b>ppm food</b>	(Concentration)	<a href="#">Show records</a>	in food - Parts per million (' ...	charochkina
				<b>ppt</b>	(Concentration)	<a href="#">Show records</a>	Parts per trillion ('ppt') den ...	indykpol
				<b>ug</b>	(Concentration)	<a href="#">Show records</a>	ug/ml -- ug stands for micro g ...	Pankaj_OCHEM / enamine
				<b>ug/bee</b>	(Concentration)	<a href="#">Show records</a>	The quantity of substance in o ...	mojca / itetko
				<b>ug/g</b>	(Concentration)	<a href="#">Show records</a>	The unit to measure concentrat ...	charochkina
				<b>ug/kg</b>	(Concentration)	<a href="#">Show records</a>	The unit to measure concentrat ...	charochkina
				<b>ug/kg food</b>	(Concentration)	<a href="#">Show records</a>	The unit to measure concentrat ...	charochkina
				<b>ug/L</b>	(Concentration)	<a href="#">Show records</a>	ug/l -- micrograms per liter, ...	charochkina
				<b>ug/L/d</b>	(Concentration)	<a href="#">Show records</a>	Concentration in micrograms (u ...	charochkina
				<b>ug/m3</b>	(Concentration)	<a href="#">Show records</a>	Microgram per cubic meter unit ...	bhataba / itetko
				<b>ug/ml</b>	(Concentration)	<a href="#">Show records</a>	Micrograms / milliliter = (10- ...	ucb
				<b>ul/L</b>	(Concentration)	<a href="#">Show records</a>	We assume that 1ul = 1 micro l ...	kamel / itetko
				<b>uM</b>	(Concentration)	<a href="#">Show records</a>	Micro mol corresponds to 10	ucb
				<b>umol/L</b>	(Concentration)	<a href="#">Show records</a>	Micromol per liter. The mole i ...	amaziz / itetko
				<b>µg/cm^3</b>	(Concentration)	<a href="#">Show records</a>	µg/cm^3 -- micrograms per cubi ...	tanzeem
				<b>µg/dcl</b>	(Concentration)	<a href="#">Show records</a>	µg/dcl -- micrograms per deci ...	kamel / tanzeem
				<b>µg/L</b>	(Concentration)	<a href="#">Show records</a>	µg/l -- micrograms per liter, ...	itetko

jm7b01342\_si\_0....csv | Show all X

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/batchupload30/show.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. Ragno. The main content area is titled "Batch Upload 3.0 - Entity remapping" and contains the following information:

- Property:** Ki
- Values:** Unit:  $\mu\text{M}$ , min value: 0.006, max value: 3.89
- Article:** unpublished
- Molecule set:** default

A blue arrow points to the "submit" button. At the bottom right, there are buttons for "Cancel Batch Upload" and "Download Excel file". The browser's taskbar at the bottom shows a file named "jm7b01342\_si\_0....csv" and a "Show all" button.

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citi x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

https://ochem.eu/batchupload30/show.do

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP

**Online chemical database** with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Batch upload 3.0 - records preview Preview the records you are about to upload, select the desired actions

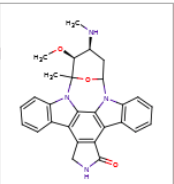
### Batch upload preview browser

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

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

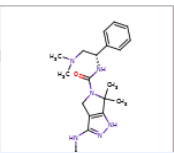
1 - 10 of 40 10 items on page 1 of 4

**Row 1**  
 Save as error  
 Skip

  
● **Ki = 0.009000000000000001 (in uM) = 8.05 (in -log(M))**  
Ragno, R  
jm7b01342\_si\_002.xls...  
N: AUTO\_1  
MoleculeID: M4402773  
Some obligatory conditions for property Ki have not been specified: [Target] RecordID: R-1  
rino.ragno Only visible to rino.ra

**error: Some obligatory conditions for property Ki have not been specified: [Target]**

**Row 2**  
 Save as error  
 Skip

  
● **Ki = 0.026000000000000002 (in uM) = 7.59 (in -log(M))**  
Ragno, R  
jm7b01342\_si\_002.xls...  
N: AUTO\_2  
MoleculeID: M95419909  
Some obligatory conditions for property Ki have not been specified: [Target] RecordID: R-2

jm7b01342\_si\_0...csv Show all x



# Preparing the dataset

	B	C	D	E
1	SMILES	ki	pKi	
2	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC([H])(C)C@H</chem>	0.009	=log10(C2)	
3	<chem>O=C(N[C@@H](C1=CC=CC=C1)CN(C)C)N(C2(C)C)CC3=C2NN=C3N</chem>	0.026		
4	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42</chem>	0.71		
5	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C</chem>	0.652		
6	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C)C(C)C=N4)C3=C1</chem>	0.884		
7	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N</chem>	1.44		
8	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42)=NN1</chem>	0.099		
9	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5CC5)=C4)C3=C</chem>	0.016		
10	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C)C=C4)C3=C1</chem>	0.077		
11	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5=CC=CC=C5)=</chem>	0.256		
12	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>	0.354		
13	<chem>CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>	3.817		
14	<chem>CC1=CC(NC2=NC(C(N3CCC(N)C)CC3)=O)=NC4=CC=C(C=C42)Cl)=</chem>	0.674		
15	<chem>CC1=CC(NC2=NC(C(N3CCC(O)C)CC3)=O)=NC4=CC=C(C=C42)Cl)=</chem>	3.89		
16	<chem>CC1=CC(NC2=NC(C(N3CCC(NC)CC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>	0.672		
17	<chem>CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	0.045		
18	<chem>CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	2.172		
19	<chem>CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>	0.768		
20	<chem>C1C1=CC=C(N=C(N=C2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=</chem>	2.408		
21	<chem>CC1=CC(NC2=NC(C(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>	0.576		
22	<chem>CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	0.325		
23	<chem>CC1=CC(NC2=NC(C(N(C)C3CNCCC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>	1.63		
24	<chem>CC1=CC(NC2=NC(C(NC3CNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	0.782		
25	<chem>C1C(C=C12)=CC=C1N=C(C(N3CCNCC3)=O)N=C2NC4=NNC(C5CC5)</chem>	0.016		
26	<chem>CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C)C=C42</chem>	0.218		
27	<chem>O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5O</chem>	0.026		
28	<chem>CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C)C=C42)</chem>	0.151		
29	<chem>O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5O</chem>	0.017		

# Preparing the dataset

The screenshot shows a spreadsheet with the following data:

SMILES	ki	pKi
O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC([H])(C[C@H]1C=CC=CC=C1)CN(C)N(C2(C)C)CC3=C2NN=C3N		0.009
CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(CI)C=C42)		0.026
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C		0.71
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C(C)C)C=N4)C3=C1		0.652
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C(C)C)C=N4)C3=C1		0.884
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N4)C3=C1		1.44
CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(CI)C=C42)=NN1		0.099
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5CC5)=C4)C3=C		0.016
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C(C)C)=C4)C3=C1		0.077
C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5=CC=CC=C5)=C4)C3=C1		0.256
CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.354
CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1		3.817
CC1=CC(NC2=NC(C(N3CCC(N)(C)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.674
CC1=CC(NC2=NC(C(N3CCC(O)(C)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1		3.89
CC1=CC(NC2=NC(C(N3CCC(NC)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.672
CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.045
CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)CI)=NN1		2.172
CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.768
C1C1=CC=C(N=C(N=C2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=NC3=CC=C(C1)C=C3C(NC4=NNC(C5CC5)=C4)C3=C1		2.408
CC1=CC(NC2=NC(C(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.576
CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.325
CC1=CC(NC2=NC(C(N(C)C3CNCCC3)=O)=NC4=CC=C(C=C42)CI)=NN1		1.63
CC1=CC(NC2=NC(C(NC3CNC3)=O)=NC4=CC=C(C=C42)CI)=NN1		0.782
C1C(C=C12)=CC=C1N=C(C(N3CCNCC3)=O)N=C2NC4=NNC(C5CC5)=C4)C3=C1		0.016
CC1=CC(NC2=NC(C(N3CCNC[C@H]3C)=O)=NC4=CC=C(CI)C=C42)		0.218
O=C(N1CCNC[C@H]1C)C2=NC3=CC=C(CI)C=C3C(NC4=NNC(C5CC5)=C4)C3=C1		0.026
CC1=CC(NC2=NC(C(N3CCNC[C@H]3C)=O)=NC4=CC=C(CI)C=C42)		0.151
O=C(N1CCNC[C@H]1C)C2=NC3=CC=C(CI)C=C3C(NC4=NNC(C5CC5)=C4)C3=C1		0.017

# Preparing the dataset

The screenshot shows an Excel spreadsheet with the following data:

	B	C	D	E
1	SMILES	ki	pKi	
2	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC([H])(C[C@</chem>		0.009	8.04575749
3	<chem>O=C(N[C@@H](C1=CC=CC=C1)CN(C)C)N(C2(C)C)CC3=C2NN=C3N</chem>		0.026	
4	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C1)C=C42</chem>		0.71	
5	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C</chem>		0.652	
6	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C(C)C)C=N4)C3=C1</chem>		0.884	
7	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N</chem>		1.44	
8	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C1)C=C42)=NN1</chem>		0.099	
9	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5CC5)=C4)C3=C</chem>		0.016	
10	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C(C)C)=C4)C3=C1</chem>		0.077	
11	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5=CC=CC=C5)=</chem>		0.256	
12	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>		0.354	
13	<chem>CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>		3.817	
14	<chem>CC1=CC(NC2=NC(C(N3CCC(N)(C)CC3)=O)=NC4=CC=C(C=C42)Cl)=</chem>		0.674	
15	<chem>CC1=CC(NC2=NC(C(N3CCC(O)(C)CC3)=O)=NC4=CC=C(C=C42)Cl)=</chem>		3.89	
16	<chem>CC1=CC(NC2=NC(C(N3CCC(NC)CC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>		0.672	
17	<chem>CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>		0.045	
18	<chem>CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>		2.172	
19	<chem>CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>		0.768	
20	<chem>C1C1=CC=C(N=C(N=C2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=</chem>		2.408	
21	<chem>CC1=CC(NC2=NC(C(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>		0.576	
22	<chem>CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>		0.325	
23	<chem>CC1=CC(NC2=NC(C(N(C)C3CNCCC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>		1.63	
24	<chem>CC1=CC(NC2=NC(C(NC3CNC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>		0.782	
25	<chem>C1C(C=C12)=CC=C1N=C(C(N3CCNCC3)=O)N=C2NC4=NNC(C5CC5)</chem>		0.016	
26	<chem>CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C1)C=C42</chem>		0.218	
27	<chem>O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C1)C=C3C(NC4=NNC(C5O</chem>		0.026	
28	<chem>CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C1)C=C42)</chem>		0.151	
29	<chem>O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C1)C=C3C(NC4=NNC(C5O</chem>		0.017	

# Preparing the dataset

	B	C	D	E
17	CC1=CC(NC2=NC(C(N3CCNCCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1	0.045	7.34678749	
18	CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1	2.172	5.66314018	
19	CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN	0.768	6.11463878	
20	ClC1=CC=C(N=C(N=C2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=	2.408	5.61834352	
21	CC1=CC(NC2=NC(C(N(C)C3CCNCCC3)=O)=NC4=CC=C(C=C42)Cl)=N	0.576	6.23957752	
22	CC1=CC(NC2=NC(C(N3CCNCCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1	0.325	6.48811664	
23	CC1=CC(NC2=NC(C(N(C)C3CNCCC3)=O)=NC4=CC=C(C=C42)Cl)=N	1.63	5.7878124	
24	CC1=CC(NC2=NC(C(N3CNCCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1	0.782	6.10679325	
25	ClC(C=C12)=CC=C1N=C(C(N3CCNCCC3)=O)N=C2NC4=NNC(C5CC5)	0.016	7.79588002	
26	CC1=CC(NC2=NC(C(N3CCN(C)C@@H]3C)=O)=NC4=CC=C(C)C=C4	0.218	6.66154351	
27	O=C(N1CCN(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5O	0.026	7.58502665	
28	CC1=CC(NC2=NC(C(N3CCN(C)C@@H]3C)=O)=NC4=CC=C(C)C=C4	0.151	6.82102305	
29	O=C(N1CCN(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5CC	0.017	7.76955108	
30	CC1=CC(NC2=NC(C(N3CCN(C)C@@H]3C)=O)=NC4=CC=C(C)C=C4	0.051	7.29242982	
31	O=C(N1CCN(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5C	0.009	8.04575749	
32	CC1=CC(NC2=NC(C(N3CCN(C)C@@H]3C)=O)=NC4=CC=C(C)C=C4	0.306	6.51427857	
33	O=C(N1CCN(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5	0.028	7.55284197	
34	CC1=CC(NC2=NC(C(N3CC(C)NC(C)C3)=O)=NC4=CC=C(C)C=C4	0.119	6.92445304	
35	O=C(N1CC(C)NC(C)C1)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5CC	0.114	6.94309515	
36	O=C(N1CCN(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5	0.006	8.22184875	
37	O=C(N1C(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5CC	0.017	7.76955108	
38	O=C(N1C(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5	0.007	8.15490196	
39	O=C(N1C(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5	0.016	7.79588002	
40	O=C(N1C(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5	0.011	7.95860731	
41	O=C(N1C(C)C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C	0.036	7.4436975	
42				
43				
44				
45				



# Preparing the dataset

The screenshot shows the LibreOffice Calc interface with a spreadsheet containing SMILES strings in column B and numerical values in columns C and D. A 'Paste Special' dialog box is open, showing options for Selection, Operations, and Shift Cells. A blue arrow points to the 'Paste all' option in the Selection section.

SMILES	C	D
O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC([H])(C[C@H]1C=CC=CC=C1)CN(C)N(C2(C)C)CC3=C2NN=C3N	0.009	8.04575749
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.026	7.58502665
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.71	6.14874165
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.652	6.1857524
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.884	6.05354773
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	1.44	5.84163751
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.099	7.00436481
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.016	7.79588002
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.077	7.11350927
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.256	6.59176003
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.354	6.45099674
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	3.817	5.41827784
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.674	6.1713401
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	3.89	5.4100504
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.672	6.17263073
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.045	7.34678749
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	2.172	5.66314018
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.768	6.11463878
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	2.408	5.61834352
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.576	6.23957752
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.325	6.48811664
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	1.63	5.7878124
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.782	6.10679325
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.016	7.79588002
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.218	6.66154351
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.026	7.58502665
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.151	6.82102305
CC1=CC=CC=C1C=CC=C(C=C1)N(C)N(C2(C)C)CC3=C2NN=C3N	0.017	7.76955108

# Preparing the dataset

The screenshot shows a spreadsheet with the following data:

	B	C	D	E
1	SMILES	pKi		
2	<chem>O=C1NCC2=C1C3=C4C5=C2C6=C(C=CC=C6)N5C7(C)OC([H])(C[C@H]1C=CC=CC=C1)CN(C)C(C)CC3=C2NN=C3N</chem>	8.04575749056068		
3	<chem>O=C(N[C@@H](C1=CC=CC=C1)CN(C)C(C)CC3=C2NN=C3N</chem>	7.58502665202918		
4	<chem>CN(C=N1)C=C1NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42</chem>	6.14874165128092		
5	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5CC5)C=N4)C3=C</chem>	6.18575240426808		
6	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C)C(C)C=N4)C3=C1</chem>	6.05354773498693		
7	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(C5=CC=CC=C5)C=N</chem>	5.84163750790475		
8	<chem>CC1=CC(NC2=NC(N3CCC(N)CC3)=NC4=CC=C(C)C=C42)=NN1</chem>	7.00436480540245		
9	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5CC5)=C4)C3=C</chem>	7.79588001734408		
10	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C)C(C)=C4)C3=C1</chem>	7.11350927482752		
11	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC(C5=CC=CC=C5)=</chem>	6.59176003468815		
12	<chem>CC1=CC(NC2=NC(C(N3CCC(N)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>	6.45099673797421		
13	<chem>CC1=CC(NC2=NC(C(N3CCC(O)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>	5.4182778400509		
14	<chem>CC1=CC(NC2=NC(C(N3CCC(N)C)CC3)=O)=NC4=CC=C(C=C42)Cl)=</chem>	6.17134010346468		
15	<chem>CC1=CC(NC2=NC(C(N3CCC(O)C)CC3)=O)=NC4=CC=C(C=C42)Cl)=</chem>	5.41005039867429		
16	<chem>CC1=CC(NC2=NC(C(N3CCC(N)C)CC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>	6.17263072694618		
17	<chem>CC1=CC(NC2=NC(C(N3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	7.34678748622466		
18	<chem>CC1=CC(NC2=NC(C(N3CCOCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	5.66314017908319		
19	<chem>CC1=CC(NC2=NC(C(N3CCN(C)CC3)=O)=NC4=CC=C(C=C42)Cl)=NN</chem>	6.11463877996849		
20	<chem>C1C1=CC=C(N=C(N=C2NC3=NNC(C)=C3)C(N4CC(NCC4)=O)=O)C2=</chem>	5.61834351741421		
21	<chem>CC1=CC(NC2=NC(C(N(C)C3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>	6.23957751657679		
22	<chem>CC1=CC(NC2=NC(C(NC3CCNCC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	6.48811663902113		
23	<chem>CC1=CC(NC2=NC(C(N(C)C3CNCCC3)=O)=NC4=CC=C(C=C42)Cl)=N</chem>	5.78781239559604		
24	<chem>CC1=CC(NC2=NC(C(NC3CNC3)=O)=NC4=CC=C(C=C42)Cl)=NN1</chem>	6.10679324694015		
25	<chem>C1C(C=C12)=CC=C1N=C(C(N3CCNCC3)=O)N=C2NC4=NNC(C5CC5)</chem>	7.79588001734408		
26	<chem>CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C)C=C42</chem>	6.6615435063954		
27	<chem>O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5O</chem>	7.58502665202918		
28	<chem>CC1=CC(NC2=NC(C(N3CCNC[C@@H]3C)=O)=NC4=CC=C(C)C=C42)</chem>	6.82102305270683		
29	<chem>O=C(N1CCNC[C@@H]1C)C2=NC3=CC=C(C)C=C3C(NC4=NNC(C5O</chem>	7.76055107862173		

# Preparing the dataset

Autenticazione | Home | Sapienz. | HomePage - Citi. | Online Chemical | Online Chemical | Journal of Medic. | Structure-Based | +

https://ochem.eu/batchupload30/show.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

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

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

**Batch Upload 3.0 - File preview and column remapping**  
Preview your data, select the sheet and the columns you would like to upload

jm7b01342\_si\_002

<input checked="" type="checkbox"/> RECORDID	<input checked="" type="checkbox"/> SMILES	<input checked="" type="checkbox"/> pKi
staurosporine	<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)N=C3NC4=CN(...</chem>	6.185752404268079
10c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=CN(...</chem>	6.053547734986927
10d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=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)N=C3NC4=NNC...</chem>	7.795880017344075
11c	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=C3NC4=NNC...</chem>	7.113509274827518
11d	<chem>C1C1=CC=C(N=C(N2CCC(N)CC2)N=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

jm7b01342\_si\_0...csv Show all

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/batchupload30/show.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. Ragno. The main content area is titled "Batch Upload 3.0 - Entity remapping" and includes a sub-section "Database entities remapping".

Under "Database entities remapping", there are three input fields:

- Property: **pKi**
- Values: Unit: **-log(mol/L)**, min value: 5.410050398674292, max value: 8.221848749616356
- Article: **unpublished**
- Molecule set: **default**

A blue arrow points to the "submit" button. At the bottom right, there are buttons for "Cancel Batch Upload" and "Download Excel file". The browser's taskbar at the bottom shows a file named "jm7b01342\_si\_0...csv" and a "Show all" button.



# Preparing the dataset

Autenticazione | Home | Sapienz. | HomePage - Citi. | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +


https://ochem.eu/batchupload30/show.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database** with modeling environment v.3.0.96.1

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

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

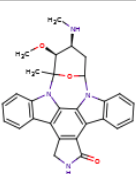
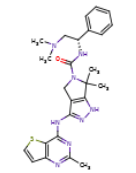
 **Batch upload 3.0 - records preview** ?  
Preview the records you are about to upload, select the desired actions

**Batch upload preview browser**

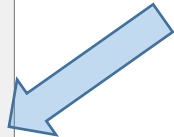
**Summary:**  
All rows in the sheet Count: **40**  
Status: valid, Count: **40**

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

1 - 10 of 40 10 items on page 1 of 4 > >>

<b>Row 1</b> <input checked="" type="radio"/> Save <input type="radio"/> Skip		<p><b>pK<sub>i</sub> = 8.045757490560675</b> (in -log(mol/L))</p> <p>Ragno, R <a href="#">jm7b01342_si_002.xls...</a> N: AUTO_1 MoleculeID: <i>M4402773</i></p> <p style="text-align: right;">RecordID: <i>R-1</i> rino.ragno <input checked="" type="checkbox"/> Only visible to rino.ra</p>
<b>Row 2</b> <input checked="" type="radio"/> Save <input type="radio"/> Skip		<p><b>pK<sub>i</sub> = 7.585026652029182</b> (in -log(mol/L))</p> <p>Ragno, R <a href="#">jm7b01342_si_002.xls...</a> N: AUTO_2 MoleculeID: <i>M95419909</i></p> <p style="text-align: right;">RecordID: <i>R-2</i> rino.ragno <input checked="" type="checkbox"/> Only visible to rino.ra</p>

jm7b01342\_si\_0...csv Show all ×



# Preparing the dataset

Autenticazione x Home | Sapienz. x HomePage - Citi. x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

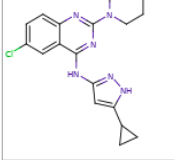
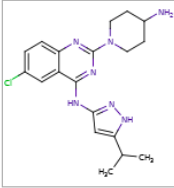
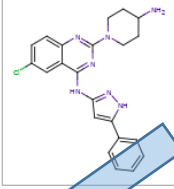
https://ochem.eu/batchupload30/show.do

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP

**Online chemical database** with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

<b>Row 8</b> <input type="radio"/> Save <input type="radio"/> Skip		<b>pKi = 7.793880017344073 (in -log(mol/L))</b> Ragno, R jm7b01342_si_002.xls... N: AUTO_8 MoleculeID: M97153552 RecordID: R-8 rino.ragno Only visible to rino.ra
<b>Row 9</b> <input type="radio"/> Save <input type="radio"/> Skip		<b>pKi = 7.113509274827518 (in -log(mol/L))</b> Ragno, R jm7b01342_si_002.xls... N: AUTO_9 MoleculeID: M97153553 RecordID: R-9 rino.ragno Only visible to rino.ra
<b>Row 10</b> <input type="radio"/> Save <input type="radio"/> Skip		<b>pKi = 6.59176003468815 (in -log(mol/L))</b> Ragno, R jm7b01342_si_002.xls... N: AUTO_10 MoleculeID: M97153554 RecordID: R-10 rino.ragno Only visible to rino.ra

1 - 10 of 40  
Proceed with upload

10 items on page 1 of 4

Cancel Batch Upload Download Excel file

jm7b01342\_si\_0...csv Show all x

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/batchupload30/show.do>. The page header includes the logo for the Online Chemical Database (v.3.0.96.1) and a welcome message for Prof. Ragno. The main content area displays a notification for a completed batch upload and a summary of the results. A blue arrow points to the 'detailed upload report' link in the summary text.

Batch upload 3.0 - finished Your upload has been finished

**Batch upload results**

Batch upload is finished. You can download the [detailed upload report](#).

**Summary:**

All rows in the sheet	Count: <b>40</b>
Status: valid, saved_valid	Count: <b>40</b>

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

jm7b01342\_si\_0...csv

# Preparing the dataset

The screenshot shows the Online Chemical Database (ochem.eu) interface. The browser address bar displays <https://ochem.eu/batchupload30/show.do>. The page header includes the logo and text "Online chemical database with modeling environment" and a version number "v.3.0.96.1". A navigation bar contains "Home", "Database", and "Models" menus. The "Models" menu is open, showing options: "Create 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 model" option. The main content area shows a "Batch upload" section with a "New Batch Upload" button and a "Download Excel file" button. The footer shows a file upload area with a file named "jm7b01342\_si\_0...csv" and a "Show all" button.

Autenticazione x Home | Sapienz. x HomePage - Citi x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

https://ochem.eu/batchupload30/show.do

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

Online chemical database with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Batch upload Your upload has

Batch upload res

Batch upload is

Summary:  
All rows in the sh

Status: valid, sav

Create 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  
Descriptors storage

oad report.

New Batch Upload Download Excel file

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

jm7b01342\_si\_0...csv Show all x

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.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. Ragno. The main content area is titled "Create a model" and includes instructions to "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](#)

**Choose the learning method:**

*Suggested modeling methods:*

- ASNN: ASsociative Neural Networks
- CHEMCHAINER: 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
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: 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
- WEKA-RF: Random Forest, only classification
- XGBoost: Scalable and Flexible Gradient Boosting

At the bottom of the page, there is a file upload area with the filename "jm7b01342\_si\_0...csv" and a "Show all" button.

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.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.Ragno!". The main content area is titled "Model validation" and includes the following options:

- CHEMCHAINER: 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
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: 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
- WEKA-RF: Random Forest, only classification
- XGBoost: Scalable and Flexible Gradient Boosting

Below the list, there is a section "Methods under development:" and a "Model validation" section with the following settings:

Validation method:

Number of folds:

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](#)

At the bottom left, there is a "Next>>" button. A blue arrow points from the text "You can create a model from template" to this button. At the bottom right, there is a "Show all" button.

# Preparing the dataset

Autenticazione | Home | Sapienz. | HomePage - Citi. | Online Chemical | Online Chemical | Journal of Medic. | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

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

Home | Database | Models | A+ a- Privacy statement

**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](#)

Click to change

**Choose the learning method:**

*Suggested modeling methods:*

- ASNN: ASsociative Neural Networks
- CHEMCHAINER: 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
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: 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
- WEKA-RF: Random Forest, only classification

javascript:void(0)

jm7b01342\_si\_0...csv ^ Show all x



# Preparing the dataset

The screenshot shows the Online Chemical Database interface. The browser address bar displays <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo and text "Online chemical database with modeling environment" and a version number "v.3.0.96.1". A navigation bar contains "Home", "Database", and "Models" menus. A user greeting "Welcome, Dear Prof.Ragno!" and links for "My account" and "Logout" are visible. Below the navigation bar, there are tabs for "Model Builder X" and "Select compound set X".

The main content area is titled "Basket browser" with a sub-header "Browse, Compare or Join molecule sets". It features a "Filter by name:" input field, a "[Create new]" button, and a "Show public sets" checkbox. Below this, a table displays the contents of the basket:

Selected records	0 records
<a href="#">jm7b01342_si_002.xls</a>	40 records

A blue arrow points to the "jm7b01342\_si\_002.xls" record, and a tooltip box below it says "Click to select this basket". At the bottom of the page, a file download bar shows "jm7b01342\_si\_0...csv" and a "Show all" button.

# Preparing the dataset

Autenticazione | Home | Sapienz... | HomePage - Citt... | Online Chemical | Online Chemical | Journal of Medic... | Structure-Based... | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

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

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

**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\\_si\\_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:*

- ASNN: ASsociative Neural Networks
- CHEMCHAINER: 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
- LSVMG: Least Squares Support Vector Machine (GPU)
- MLR: 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\_si\_0...csv ^ Show all ×

# Preparing the dataset

The screenshot shows the 'Online chemical database with modeling environment' interface. The browser address bar is at <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the site logo, the text 'Online chemical database with modeling environment', and a user greeting: 'Welcome, Dear Prof.Ragno! My account Logout'. Below the header is a navigation menu with 'Home', 'Database', and 'Models' tabs. The main content area lists various machine learning models, each with a radio button for selection:

- CHEMCHAINER: 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
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: 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
- WEKA-RF: Random Forest, only classification
- XGBoost: Scalable and Flexible Gradient Boosting

Below the list, there is a section 'Methods under development:' followed by a 'Model validation' section. In 'Model validation', the 'Validation method' is set to 'N-Fold cross-validation'. The 'Number of folds' is set to 5. There are two checkboxes: 'Stratified cross-validation (classification only)' and 'Consider each record as a molecule.' Below this, a text line says: 'You can create a model from template: [import an XML model template](#) or [use another model as a template](#)'. At the bottom left of the main content area, there is a 'Next>' button, which is highlighted by a blue arrow pointing from the left. The footer of the page shows a file upload area with 'jm7b01342\_si\_0....csv' and a 'Show all' button.

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for 'Online chemical database with modeling environment' and a version number 'v.3.0.96.1'. A navigation bar contains 'Home', 'Database', and 'Models' menus. The main content area is titled 'Model creator' and contains the instruction 'Select model template and training set'. Below this, a section titled 'Select the preferred data preprocessing options' is shown. Under the sub-section 'Preprocessing of molecules (Chemaxon)', four options are listed, all of which are checked: 'Standardization', 'Neutralize', 'Remove salts', and 'Clean structure'. At the bottom of this section are two buttons: '<<Back' and 'Next>>'. A blue arrow points to the 'Next>>' button. The browser's address bar and tabs are visible at the top, and a file upload notification for 'jm7b01342\_si\_0...csv' is visible at the bottom.

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP »

**Online chemical database**  
with modeling environment

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

Home Database Models A+ a- Privacy statement

**Model creator**  
Select model template and training set

**Select the preferred data preprocessing options**

**Preprocessing of molecules (Chemaxon)**

- Standardization
- Neutralize
- Remove salts
- Clean structure

<<Back Next>>

jm7b01342\_si\_0...csv Show all x

# Preparing the dataset

Autenticazione x Home | Sapienz. x HomePage - Citt. x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP

**Online chemical database**  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

**Model creator**  
Select model template and training set

**Select the molecular descriptors**

**Recommended descriptor types**

- E-state
  - E-State types:
    - Atom indices
    - Bonds indices
    - Atom counts
    - Bonds counts
- Aromatize structures: Chemaxon Basic
- ALogPS (2)
- GSFragment (1138)
- CDK 2.0 descriptors (256/3D)
- Dragon v. 7 (5270/3D)
- alvaDesc v.1.0.14 (5305/3D)
- ISIDA fragments
- 'Inductive' descriptors (54/3D)
- MERA descriptors (529/3D)
- MERSY descriptors (42/3D)
- Chemaxon descriptors (499/3D)
- QNPR
- Spectrophores (144/3D)
- Structural alerts (ToxAlerts)

**Predictions by OCHEM's featured models**

- Ames levenberg
- Toxicity against T. Pyriformis
- ALogPS 3.0
- CYP1A2 Estate+ALogPS
- CYP2C9 Estate+ALogPS
- CYP2C19 Estate+ALogPS
- CYP2D6 Estate+ALogPS
- CYP3A4 Estate+ALogPS
- Pyrolysis 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)
- OESState
- Dragon v. 5.4 (1644/3D)
- Dragon v. 5.5 (3224/3D)

jm7b01342\_sl\_0....csv Show all

# Preparing the dataset

Autenticazione x Home | Sapienz. x HomePage - Citt. x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

**Online chemical database**  
with modeling environment

v.3.0.96.1

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

Home Database Models A+ a- Privacy statement

**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)
- alvaDesc v.1.0.14 (5305/3D)
- ISIDA fragments
- 'Inductive' descriptors (54/3D)
- MERA descriptors (529/3D)
- MERSY descriptors (42/3D)
- Chemaxon descriptors (499/3D)
- QNPR
- Spectrophores (144/3D)
- Structural alerts (ToxAlerts)

**Predictions by OCHEM's featured models** ⓘ

- Ames levenberg
- Toxicity against T. Pyriformis
- ALogPS 3.0
- CYP1A2 Estate+ALogPS
- CYP2C9 Estate+ALogPS
- CYP2C19 Estate+ALogPS
- CYP2D6 Estate+ALogPS
- CYP3A4 Estate+ALogPS
- Pyrolysis 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)
- OESate
- Dragon v. 5.4 (1644/3D)
- Dragon v. 5.5 (3224/3D)

jm7b01342\_sl\_0....csv ^ Show all x

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP

**Online chemical database** with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Select model template and training set

### Select the molecular descriptors

**Recommended descriptor types**

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

[select all] [select none]

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

alvaDesc v.1.0.14 (5305/3D)

ISIDA fragments

**Predictions by OCHEM's featured models**

- Ames levenberg
- Toxicity against T. Pyriformis
- ALogPS 3.0
- CYP1A2 Estate+ALogPS
- CYP2C9 Estate+ALogPS
- CYP2C19 Estate+ALogPS
- CYP2D6 Estate+ALogPS
- CYP3A4 Estate+ALogPS
- Pyrolysis 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)
- OESState
- Dragon v. 5.4 (1644/3D)
- Dragon v. 5.5 (3224/3D)
- Dragon v. 6 (4885/3D)
- MOPAC 7.1 descriptors (25/3D)

jm7b01342\_sl\_0....csv Show all



# Preparing the dataset

The screenshot shows the 'Online chemical database with modeling environment' interface. The browser address bar displays 'https://ochem.eu/modelconfigurator/choose.do'. The page header includes the site logo, version 'v.3.0.96.1', and a welcome message for 'Prof.Ragno!' with links for 'My account' and 'Logout'. A navigation menu at the top contains 'Home', 'Database', and 'Models'. The main content area is a configuration form with several sections:

- Chemaxon descriptors (439/3D):** Includes checkboxes for QNPR, Spectrophores (144/3D), and Structural alerts (ToxAlerts).
- Special descriptors (scaffolds, fingerprints):** Includes checkboxes for Chemaxon Scaffolds, Silicos-It Scaffolds, ECFP Fingerprints (with a note 'Not supported by your installation'), and MolPrint Fingerprints.
- Under development: can change anytime and backward compatibility is not guaranteed. Use at your own risk!** This section contains a list of descriptors with checkboxes: JPligP, CDK 2.2 descriptors (256/3D), RDKit descriptors (3D), RDKit additional descriptors (3D), MORDRED descriptors (1826/3D), CDDD, MOPAC2016 descriptors (35/3D), SIRMS, PyDescriptor descriptors (16251/3D), External descriptors, and Allow Merging Descriptors (experimental).

At the bottom of the form, there are two buttons: '<<Back' and 'Next>>'. A blue arrow points to the 'Next>>' button. Another blue arrow points to the right side of the configuration area. The browser's taskbar at the bottom shows a file named 'jm7b01342\_si\_0....csv' and a 'Show all' button.


# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for the Online chemical database with modeling environment (v.3.0.96.1) and a welcome message for Prof. Ragno, with links for My account and Logout. The navigation menu contains Home, Database, and Models. The main content area is titled "Model creator" and "Select model template and training set". Below this, there is a section "Select a tool to optimize molecule structures" with five radio button options: "No optimisation", "Optimise with Corina", "Optimise with OpenBabel", "Optimise with OBGEN (part of OpenBabel distribution)", and "Optimise with BALLOON". A blue arrow points to the "Next" button in the navigation controls at the bottom of the section. The browser's address bar shows several tabs, including "Autenticazione", "Home | Sapienz", "HomePage - Citt", "Online Chemical", "Journal of Medic", and "Structure-Based". The browser's taskbar at the bottom shows a file named "jm7b01342\_sl\_0...csv" and a "Show all" button.

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

← → ↻ 🏠 <https://ochem.eu/modelconfigurator/choose.do> ☆

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP »

 **Online chemical database**  
with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! [✉ My account](#) [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

**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

---


jm7b01342\_sl\_0...csv ^ Show all x

# Preparing the dataset

Autenticazione x Home | Sapienz. x HomePage - Citt. x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP »

 **Online chemical database**  
with modeling environment v.3.0.96.1

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

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

**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  $R$
- After filtering, I want to select necessary descriptors myself (*advanced*)

**Normalisation parameters**

Descriptors normalization

Values normalization

jm7b01342\_si\_0...csv

# Preparing the dataset

Autenticazione x Home | Sapienz. x HomePage - Citt. x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

**Online chemical database** with modeling environment v.3.0.96.1

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

Home Database Models A+ a- Privacy statement

**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  $R$
- After filtering, I want to select necessary descriptors myself (advanced)

**Normalisation parameters**

Descriptors normalization

Values normalization

<<Back Next>>

jm7b01342\_si\_0...csv ^ Show all x

# Preparing the dataset

Online Chemical Modeling Environment - Google Chrome

Autenticazione | Home | Sapientz | HomePage - Citi | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

v.3.0.96.1

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

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

**Model creator**  
Select model template and training set

**Configure PLS method**

Number of latent variables:

Optimize the number of latent variables automatically

Limit predicted values to the training set range

jm7b01342\_sl\_0...csv Show all ×

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP »

**Online chemical database** with modeling environment v.3.0.96.1

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

Home Database Models A+ a- Privacy statement

**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 (for long tasks)

jm7b01342\_sl\_0...csv ^ Show all x

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for the Online chemical database with modeling environment (v.3.0.96.1) and a welcome message for Prof. Ragno, with links for My account and Logout. A navigation menu contains Home, Database, and Models. The main content area is titled 'Model creator' with the instruction 'Select model template and training set'. Below this is the 'Run model builder' section, which displays a progress indicator and the message: 'Finished posting ... - Processing task Corina - Waiting for a free server -- 09:16'. There are links for [cancel] and [fetch result later]. At the bottom of the main area are '<<Back' and 'Next>>' buttons. A file upload bar at the bottom shows a file named 'jm7b01342\_sl\_0...csv' and a 'Show all' button.



# Preparing the dataset

Autenticazione G x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

**Online chemical database**  
with modeling environment

v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

**Model creator**  
Select model template and training set

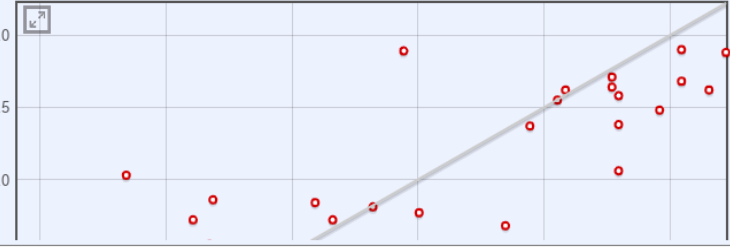
**Save the model**  
Please enter your model's name:

**Overview**

Model name: pKi\_PLS\_[Dragon7 (blocks: 1-30)] - 336922 [rename] [Dragon7 (blocks: 1-30)]  
Temporal Public ID: 37774824 - use this link to share the model  
Correl. limit: 0.0 Variance threshold: 0.01, Maximum value: 999999,

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

Data Set	#	R2	q2	RMSE	MAE
Training set: <a href="#">jm7b01342_si_002.xls</a>	40 records	0.4 ± 0.2	0.3 ± 0.3	0.7 ± 0.2	0.46 ± 0.08



scale X: STANDARDIZE0 latent variables 3D by Corina  
5-fold cross-validation  
3008 pre-filtered descriptors

scale X: STANDARDIZE2 latent variables  
Y = -1.03 + 7.3E-6\**MW* - 0.00143\**AMW* + 9.56E-5\**Sv* + 7.81E-5\**Se* + 9.37E-5\**Sp* + 7.9E-5\**Si* - 0.0664\**Mv* - 0.108\**Me* - 0.0773\**Mp* + 0.0959\**Mi* + 0.0383\**GD* + 8.21E-5\**nAT* + 7.64E-5\**nSK* - 0.00167\**nTA* + 1.01E-4\**nBT* + 1.66E-4\**nBO* - 4.82E-4\**nBM* + 3.18E-5\**SCBO* + 0.00108\**RBN* + 0.0675\**RBF* + 1.51E-4\**nDB* - 4.65E-4\**nAB* + 1.93E-4\**nH* + 9.4E-5\**nC* + 0.00145\**nN* - 0.00104\**nO* - 0.00236\**nCL* - 0.00246\**nHM* - 3.41E-4\**nHet* - 8.88E-4\**nX* + 3.52E-4\**H%* - 1.59E-4\**C%* + 1.81E-4\**N%* - 6.33E-4\**O%* - 4.98E-4\**X%* + 0.00108\**nCsp3* -

LigandScout\_4....dmg ^ LigandScout\_4....exe ^ LigandScout....tar.gz ^ jm7b01342\_si\_0....csv ^ Show all x

# Preparing the dataset

Autenticazione G x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

**Online chemical database** with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Model c  
Select m

Save

Pleas MatchedPairs pKi\_PLS\_[Dragon7 (bloc

**Compound properties**

- Properties
- Conditions
- Units
- Articles/Books
- Journals
- ToxAlerts
- Baskets
- Tags
- Set area of interest...
- User-related changes

**Batch data upload**

- Trash

Dragon7 (blocks: 1-30)) - 336922 [rename]  
24 - use this link to share the model

Correl. limit: 0.0 Variance threshold: 0.01,  
Maximum value: 999999,

scale X: STANDARDIZE0 latent variables 3D by  
Corina  
5-fold cross-validation  
-  
3008 pre-filtered descriptors

scale X: STANDARDIZE2 latent variables  
Y = -1.03 + 7.3E-6\**MW* - 0.00143\**AMW* +  
9.56E-5\**Sv* + 7.81E-5\**Se* + 9.37E-5\**Sp* + 7.9E-  
5\**Si* - 0.0664\**Mv* - 0.108\**Me* - 0.0773\**Mp* +  
0.0959\**Mi* + 0.0383\**GD* + 8.21E-5\**nAT* +  
7.64E-5\**nSK* - 0.00167\**nTA* + 1.01E-4\**nBT* +  
1.66E-4\**nBO* - 4.82E-4\**nBM* + 3.18E-5\**SCBO*  
+ 0.00108\**RBN* + 0.0675\**RBF* + 1.51E-4\**nDB* -  
4.65E-4\**nAB* + 1.93E-4\**nH* + 9.4E-5\**nC* +  
0.00145\**nN* - 0.00104\**nO* - 0.00236\**nCL* -  
0.00246\**nHM* - 3.41E-4\**nHet* - 8.88E-4\**nX* +  
3.52E-4\**H%* - 1.59E-4\**C%* + 1.81E-4\**N%* -  
6.33E-4\**O%* - 4.98E-4\**X%* + 0.00108\**nCsn3* -

Data Set	#	R2	q2	RMSE	MAE
Training set: jm7b01342_si_002.xls	40 records	0.4 ± 0.2	0.3 ± 0.3	0.7 ± 0.2	0.46 ± 0.08

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

LigandScout\_4....dmg ^ LigandScout\_4....exe ^ LigandScout....tar.gz ^ jm7b01342\_si\_0....csv ^ Show all x

# Preparing the dataset

The screenshot shows the Online Chemical Database interface. The browser address bar displays <https://ochem.eu/basket/show.do>. The page header includes the logo and text "Online chemical database with modeling environment" and a user greeting "Welcome, Dear Prof.Ragno!". Navigation tabs for "Home", "Database", and "Models" are visible. The main content area is titled "Basket browser" and contains a table of records. A blue arrow points to the "jm7b01342\_si\_002.xls" record, which has a tooltip that says "Open basket profile".

Filter by name:  [Create new] Show public sets

1 - 2 of 2

Selected records	0 records
jm7b01342_si_002.xls	40 records 1 pending models

1 - 2 of 2

Open basket profile

javascript:void(0)

LigandScout\_4....dmg ^ | LigandScout\_4.....exe ^ | LigandScout.....tar.gz ^ | jm7b01342\_si\_0....csv ^ | Show all x

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP »

**Online chemical database** with modeling environment v.3.0.96.1

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Molecule sets X Edit basket X

### Basket editor

Add new basket or edit existing basket

Name: jm7b01342\_si\_002.xls (min. 2 characters)

Description (optional):

Excluded implicit records (under development):

#### Actions

- 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
- Export this basket into Excel, CSV or SDF

#### Statistics of the basket

Properties	Records	Unique compounds	
pKi	40 records	40 compounds	Show MMPs

https://ochem.eu/epbrowser/show.do?basket-select=189582&property=48967

LigandScout\_4....dmg ^ LigandScout\_4\_....exe ^ LigandScout\_....tar.gz ^ jm7b01342\_si\_0....csv ^ Show all x

# Preparing the dataset

The screenshot displays the 'Online chemical database with modeling environment' interface. The browser address bar shows the URL <https://ochem.eu/basket/show.do>. The page title is 'Compounds properties browser' with a subtitle 'Search for numerical compounds properties linked to scientific articles'. The user is logged in as 'Prof. Ragno'.

The interface includes a navigation menu with 'Home', 'Database', and 'Models'. Below this, there are tabs for 'Molecule sets X', 'Edit basket X', and 'records X'. The main content area is divided into a 'FILTERS' sidebar on the left and a main list of records on the right.

**Filters:**

- SOURCE:** Article/Source [select]
- PROPERTY:** Activity/Property [select], with 'pKi' highlighted in yellow.
- MOLECULE FILTERS:** Name / OCHEM ID / Inchi-Key

**Records List:**

Chemical Structure	pKi Value	Record ID	Visibility
	7.443697499232712 (in -log(mol/L))	R38465835	Only visible to rino.ragno
	7.958607314841775 (in -log(mol/L))	R38465834	Only visible to rino.ragno
	7.795880017344075 (in -log(mol/L))		

The records list also includes the name 'Ragno, R', file names like 'jm7b01342\_si\_002.xls', and molecule IDs such as 'M97153584'. A blue arrow points to the right side of the records list, highlighting the visibility controls.

The bottom of the browser shows several open tabs: 'LigandScout\_4....dmg', 'LigandScout\_4....exe', 'LigandScout....tar.gz', and 'jm7b01342\_si\_0....csv'.

# Preparing the dataset

The screenshot displays the Online Chemical Database interface. The browser address bar shows the URL <https://ochem.eu/basket/show.do>. The page header includes the logo and name "Online chemical database with modeling environment" and a version number "v.3.0.96.1". A navigation menu contains "Home", "Database", and "Models". A user is logged in as "Prof. Ragno", with options for "My account" and "Logout".

The main content area is divided into a left sidebar with filters and a right main panel with a list of records. The sidebar includes a "Molecular mass" filter, "ADVANCED MOLECULE FILTERS", "MISCELLANEOUS" (with "Current set" set to "jm7b01342\_si\_002.xls"), "Data origin and quality" (with "Data introducers" set to "All users" and "Data visibility" set to "All data"), and "Discover issues with the data" (with checkboxes for "Error records", "Error inchies", "Mismatching names", "Include stereochem.", "Empty molecules", "Show only duplicates", and "No stereochemistry"). At the bottom of the sidebar are "REFRESH" and "RESET" buttons.

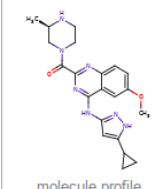
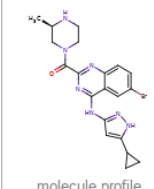
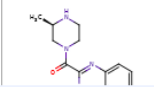
The main panel displays a list of records. Each record includes a chemical structure (molecule profile), a pKi value, the name "Ragno, R", a file name "jm7b01342\_si\_002.xls", a note "N: AUTO\_37", the molecule ID "M97153581", and the record ID "R38465832". The records are sorted by "Creation time" in "Ascending order". A pagination bar at the bottom of the list shows "1 - 5 of 40" items, with "5" items on page "1" of "8". A blue arrow points to the "5" in the pagination bar.

The taskbar at the bottom shows several open files: "LigandScout\_4....dmg", "LigandScout\_4....exe", "LigandScout....tar.gz", and "jm7b01342\_si\_0...csv". A "Show all" button is visible on the right side of the taskbar.

# Preparing the dataset

The screenshot displays the 'Online chemical database with modeling environment' interface. The browser address bar shows 'https://ochem.eu/basket/show.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. Ragno'.

The main content area is titled 'Compounds properties browser' and shows a list of records. The records are filtered by 'pKi' (highlighted in yellow in the filters section). The records are displayed in a table-like format with columns for 'molecule profile', 'pKi', 'RecordID', and 'visibility'. A blue arrow points to the 'Records' button in the toolbar, which has a tooltip that says 'Select all records matching current filters'.

molecule profile	pKi	RecordID	visibility
	$7.443697499232712$ (in -log(mol/L))	R38465835	Only visible to rino.ragno
	$7.958607314841775$ (in -log(mol/L))	R38465834	Only visible to rino.ragno
	$7.795880017344075$ (in -log(mol/L))	R38465833	Only visible to rino.ragno



# Preparing the dataset

The screenshot displays the 'Online chemical database with modeling environment' interface. The browser address bar shows 'https://ochem.eu/basket/show.do'. The page header includes the site logo, version 'v.3.0.96.1', and a welcome message for 'Prof. Ragno'. Navigation tabs include 'Home', 'Database', and 'Models'. A sub-menu at the top shows 'Molecule sets X', 'Edit basket X', and 'records X'. The main content area lists three molecule records, each with a chemical structure, pKi value, name, molecule ID, and record ID. Two blue arrows point to the 'Select/unselect this record' checkboxes for the second and third records.

Chemical Structure	pKi (in -log(mol/L))	Name	MoleculeID	RecordID	Timestamp	Visibility
	6.1487416512809245	Ragno, R jm7b01342_si_002.xls N: AUTO_3	M97153547	R38465798	09:09, 18 Nov 19	Only visible to rino.ragno
	7.585026652029182	Ragno, R jm7b01342_si_002.xls N: AUTO_2	M95419909	R38465797	09:09, 18 Nov 19	Only visible to rino.ragno
	8.045757490560675	Ragno, R jm7b01342_si_002.xls N: AUTO_1	M4402773	R38465796	09:09, 18 Nov 19	Only visible to rino.ragno

1 - 40 of 40

Taskbar: LigandScout\_4....dmg, LigandScout\_4....exe, LigandScout....tar.gz, jm7b01342\_si\_0....csv, Show all

# Preparing the dataset

The screenshot displays the OCHEM web interface. At the top, the browser address bar shows the URL <https://ochem.eu/basket/show.do>. The page header includes the OCHEM logo, the text "Online chemical database with modeling environment", and a welcome message for Prof. Ragno. A navigation menu contains "Home", "Database", and "Models". Below the menu, there are tabs for "Molecule sets", "Edit basket", and "records".

The main content area is titled "Compounds properties browser" and includes a search bar and a "Filters" sidebar. The sidebar has sections for "SOURCE", "PROPERTY" (with "pKi" selected), and "MOLECULE FILTERS". The main list shows three records, each with a chemical structure, a pKi value, and associated metadata. The records are:

Record ID	pKi	Source	Molecule ID	Visibility
R38465835	7.443697499232712	Ragno, R jm7b01342_si_002.xls	M97153584	Private record
R38465834	7.958607314841775	Ragno, R jm7b01342_si_002.xls	M97153583	Private record
R38465833	7.795880017344075	Ragno, R jm7b01342_si_002.xls	M97153582	Private record

The bottom of the screenshot shows a Windows taskbar with several open files, including "LigandScout\_4....dmg", "LigandScout\_4....exe", "LigandScout\_....tar.gz", and "jm7b01342\_si\_0....csv".

# Preparing the dataset

The screenshot shows the Online Chemical Database interface. The browser address bar displays <https://ochem.eu/basket/show.do>. The page header includes the logo for the Online chemical database with modeling environment, the version number v.3.0.96.1, and a welcome message for Prof. Ragno with links for My account and Logout. The navigation menu includes Home, Database, and Models. The main content area is titled 'Basket browser' and contains a filter by name field, a 'Create new' button, and a 'Show public sets' checkbox. Below this, there are two entries in the basket:

Entry Name	Records	Pending Models
Selected records	38 records	
jm7b01342_si_002.xls	40 records	1 pending models

A blue arrow points to the '38 records' value for the 'Selected records' entry. The taskbar at the bottom shows several files related to LigandScout, including .dmg, .exe, .tar.gz, and .csv files.

# Preparing the dataset

The screenshot shows 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 the following fields and actions:

- 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 table showing the following data:

Properties	Records	Unique compounds	
pKi	38 records	38 compounds	Show MMPs

The bottom of the browser window shows several open files: 'LigandScout\_4...dmg', 'LigandScout\_4...exe', 'LigandScout...tar.gz', and 'jm7b01342\_si\_0...csv'. A 'Show all' button is visible in the bottom right corner.

# Preparing the dataset

The screenshot shows the Online Chemical Database interface. The browser address bar displays <https://ochem.eu/basket/show.do>. The page header includes the logo and text "Online chemical database with modeling environment" and a version number "v.3.0.96.1". A welcome message reads "Welcome, Dear Prof.Ragno!" with links for "My account" and "Logout". Navigation tabs for "Home", "Database", and "Models" are visible. The main content area is titled "Basket browser" and contains a filter by name field, a "[Create new]" button, and a "Show public sets" checkbox. Below this, a table lists two items:

Item Name	Records	Models
38 Mols	38 records	
jm7b01342_si_002.xls	40 records	1 pending models

A blue arrow points to the "38 Mols" entry. The bottom of the browser window shows a taskbar with files like "LigandScout\_4....dmg", "LigandScout\_4.....exe", "LigandScout.....tar.gz", and "jm7b01342\_si\_0....csv".

# Preparing the dataset

The screenshot shows the 'Online chemical database with modeling environment' interface. The browser address bar displays 'https://ochem.eu/modelconfigurator/choose.do'. The page header includes the site logo, version 'v.3.0.96.1', and a welcome message for 'Prof. Ragno!' with links for 'My account' and 'Logout'. A navigation menu contains 'Home', 'Database', and 'Models'. The main content area is titled 'Create a model' and includes instructions to 'Select the training and validation sets, the machine learning method and the validation protocol'. A blue arrow points to the '38 Mols [details]' link under the 'Training set (required)' section. Below this, there is a dropdown for 'pKi using unit: -log(mol/L)', a checkbox for 'Skip model configuration and use the predefined settings', and a 'Choose the learning method' section with a list of suggested modeling methods. The taskbar at the bottom shows several files related to 'LigandScout' and a 'Show all' button.

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

**Online chemical database**  
with modeling environment

Welcome, Dear Prof. Ragno! My account Logout

Home Database Models At a- Privacy statement

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:  
pKi 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)
- 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
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: Multiple Linear Regression

LigandScout\_4....dmg ^ LigandScout\_4\_....exe ^ LigandScout\_....tar.gz ^ jm7b01342\_si\_0....csv ^ Show all x


# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.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.Ragno". The navigation menu includes "Home", "Database", and "Models". A list of models is displayed, with "PLS: Partial Least Squares" selected. Below the list, there is a section for "Model validation" with a dropdown for "Validation method" set to "N-Fold cross-validation", a text input for "Number of folds" set to "5", and two checkboxes: "Stratified cross-validation (classification only)" and "Consider each record as a molecule.". A blue arrow points to a "Next" button at the bottom left. The taskbar at the bottom shows several files related to "LigandScout\_4" and a "Show all" button.

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

← → ↻ 🏠 <https://ochem.eu/modelconfigurator/choose.do> ☆

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP >>

 **Online chemical database**  
with modeling environment v.3.0.96.1

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

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

- CHEMCHAINER: 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
- LSSVMG: Least Squares Support Vector Machine (GPU)
- MLR: 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
- WEKA-RF: Random Forest, only classification
- XGBoost: Scalable and Flexible Gradient Boosting

*Methods under development:*

**Model validation**

Validation method:

Number of folds:

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](#)

LigandScout\_4....dmg ^ | LigandScout\_4.....exe ^ | LigandScout.....tar.gz ^ | jm7b01342\_si\_0....csv ^ |  x

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for 'Online chemical database with modeling environment' and a welcome message for 'Prof. Ragno!'. The main content area is titled 'Model creator' and contains the following elements:

- Model creator**  
Select model template and training set
- Select the preferred data preprocessing options**
- Preprocessing of molecules (Chemaxon)**
  - Standardization
  - Neutralize
  - Remove salts
  - Clean structure
- Navigation buttons: <<Back and Next>>

A blue arrow points to the 'Next' button. The browser's taskbar at the bottom shows several files, including 'LigandScout\_4\_...dmg', 'LigandScout\_4\_...exe', 'LigandScout\_...tar.gz', and 'jm7b01342\_si\_0...csv'.



# Preparing the dataset

Autenticazione | Home | Sapienz | HomePage - Citt | Online Chemical | Online Chemical | Journal of Medic | Structure-Based | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendeley | TEMP

**Online chemical database**  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models At a- Privacy statement

**Model creator**  
Select model template and training set

**Select the molecular descriptors**

**Recommended descriptor types**

- E-state
- ALogPS (2)
- GSFragment (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)
- 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)
- 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 T. Pyriformis
- ALogPS 3.0
- CYP1A2 Estate+ALogPS
- CYP2C9 Estate+ALogPS
- CYP2C19 Estate+ALogPS
- CYP2D6 Estate+ALogPS
- CYP3A4 Estate+ALogPS
- Pyrolysis 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)
- OESate
- Dragon v. 5.4 (1644/3D)
- Dragon v. 5.5 (3224/3D)

LigandScout\_4....dmg | LigandScout\_4....exe | LigandScout\_....tar.gz | jm7b01342\_si\_0....csv | Show all

# Preparing the dataset

The screenshot shows the 'Online chemical database with modeling environment' configuration interface. The browser address bar is at <https://ochem.eu/modelconfigurator/choose.do>. The page title is 'Online chemical database with modeling environment' and the version is 'v.3.0.96.1'. A welcome message reads 'Welcome, Dear Prof.Ragno!' with links for 'My account' and 'Logout'. A navigation menu includes 'Home', 'Database', and 'Models'. The 'Database' section is expanded to show a list of descriptors with checkboxes:

- Chemaxon descriptors (400/3D)
- QNPR
- Spectrophores (144/3D)
- Structural alerts (ToxAlerts)

Below this is the 'Special descriptors (scaffolds, fingerprints):' section:

- Chemaxon Scaffolds
- Silicos-It Scaffolds
- ECFP Fingerprints *Not supported by your installation*
- MolPrint Fingerprints

A warning message states: 'Under development: can change anytime and backward compatibility is not guaranteed. Use at your own risk!'. This is followed by a list of descriptors:

- JPllogP
- CDK 2.2 descriptors (256/3D)
- RDKit descriptors (3D)
- RDKit additional descriptors (3D)
- MORDRED descriptors (1826/3D)
- CDDD
- MOPAC2016 descriptors (35/3D)
- SIRMS
- PyDescriptor descriptors (16251/3D)
- External descriptors

At the bottom of the list is the option:  Allow Merging Descriptors (experimental). Below the list are two buttons: '<<Back' and 'Next>>'. A blue arrow points to the 'Next>>' button. The Windows taskbar at the bottom shows several files, including 'LigandScout\_4\_...dmg', 'LigandScout\_4\_...exe', 'LigandScout\_...tar.gz', and 'jm7b01342\_si\_0...csv'.

# Preparing the dataset

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

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

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP

**Online chemical database**  
with modeling environment

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

**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>>

LigandScout\_4....dmg LigandScout\_4.....exe LigandScout.....tar.gz jm7b01342\_si\_0....csv Show all x

# Preparing the dataset

The screenshot shows the 'Model creator' 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 'Model creator' and contains the following sections:

- 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  $R$
  - After filtering, I want to select necessary descriptors myself (*advanced*)
- Normalisation parameters**
  - Descriptors normalization:
  - Values normalization:

At the bottom of the form, there are two buttons: '<<Back' and 'Next>>'. A blue arrow points from the 'Next>>' button towards the 'Standardize' dropdown menu. The taskbar at the bottom shows several files: 'LigandScout\_4....dmg', 'LigandScout\_4.....exe', 'LigandScout.....tar.gz', and 'jm7b01342\_si\_0....csv'.

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for the Online chemical database with modeling environment, the version number v.3.0.96.1, and a welcome message for Prof. Ragno with links for My account and Logout. The navigation menu contains Home, Database, and Models. The main content area is titled "Model creator" and "Select model template and training set". Under the heading "Configure PLS method", there is a text input field for "Number of latent variables" containing the value "0". Below this, there are two checkboxes: "Optimize the number of latent variables automatically" (checked) and "Limit predicted values to the training set range" (unchecked). At the bottom of the configuration section, there are two buttons: "<<Back" and "Next>>". A blue arrow points from the "Next>>" button towards the "Number of latent variables" input field. The browser's taskbar at the bottom shows several files: LigandScout\_4....dmg, LigandScout\_4\_....exe, LigandScout\_....tar.gz, and jm7b01342\_si\_0....csv, along with a "Show all" button.

# Preparing the dataset

The screenshot shows a web browser window with the URL `https://ochem.eu/modelconfigurator/choose.do`. The page header includes the logo for the Online chemical database (with modeling environment) and a welcome message for Prof. Ragno. The main content area is titled 'Model creator' and 'Start calculation of the model'. It contains a text input field with the value 'pKi\_PLS\_[Dragon7 (blocks: 1-30)] - 336948', a checked checkbox for 'Save models', and radio buttons for task priority: 'High priority (please, use for fast tasks only)', 'Normal priority' (selected), and 'Low priority (for long tasks)'. At the bottom, there are three buttons: '<<Back', 'Start calculation' (highlighted with a blue arrow), and 'Discard'. The browser's taskbar at the bottom shows several files related to 'LigandScout' and a CSV file.

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for "Online chemical database with modeling environment" and a welcome message: "Welcome, Dear Prof.Ragno! My account Logout". A navigation menu contains "Home", "Database", and "Models". The main content area is titled "Model creator" with the instruction "Select model template and training set". Below this, the "Run model builder" section is active, displaying a "Starting..." message with a circular loading icon and links for "[cancel]" and "[fetch result later]". At the bottom of the main content area, there are navigation buttons: "<<Back" and "Next>>". The browser's taskbar at the bottom shows several files: "LigandScout\_4....dmg", "LigandScout\_4.....exe", "LigandScout.....tar.gz", and "jm7b01342\_si\_0....csv".


# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for the Online chemical database with modeling environment, the version number v.3.0.96.1, and a welcome message for Prof. Ragno with links for My account and Logout. A navigation menu contains Home, Database, and Models. The main content area is titled 'Model creator' and 'Run model builder'. A status message indicates 'Finished posting ... - Processing task Corina - Waiting for a free server -- 09:37' with links for [cancel] and [fetch result later]. Navigation buttons for '<<Back' and 'Next>>' are visible. The taskbar at the bottom shows several files related to LigandScout.

Autenticazione x Home | Sapienz x HomePage - Citt x Online Chemical x Online Chemical x Journal of Medic x Structure-Based x +

← → ↻ 🏠 <https://ochem.eu/modelconfigurator/choose.do> ☆

Apps Università WWW Mail Ricerca Vari SoftwareVari NatComp METZ IMP Kahoot! Learning Save to Mendeley TEMP »

 **Online chemical database**  
with modeling environment v.3.0.96.1


Welcome, Dear Prof.Ragno! [✉ My account](#) [Logout](#)

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

**Model creator**  
Select model template and training set




---

**Run model builder**

  
Finished posting ... - Processing task Corina - Waiting for a free server -- 09:37  
[\[cancel\]](#) [\[fetch result later\]](#)

---

[<<Back](#) [Next>>](#)

LigandScout\_4....dmg ^ |  LigandScout\_4\_....exe ^ |  LigandScout\_....tar.gz ^ |  jm7b01342\_si\_0....csv ^ | [Show all](#) x



# Preparing the dataset

The screenshot shows a web browser window with multiple tabs. The active tab is 'Online Chemical' at the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the 'Online chemical database with modeling environment' logo and a welcome message for 'Prof. Ragno!'. A navigation menu contains 'Home', 'Database', and 'Models'. The main content area is titled 'Model creator' and 'Run model builder'. A status message indicates: 'Finished posting ... - Processing task Descriptors - Tasks are sent for calculations -- 09:37'. Below this message are links for '[cancel]' and '[fetch result later]'. At the bottom of the main area are '<<Back' and 'Next>>' buttons. The browser's taskbar at the bottom shows several files: 'LigandScout\_4...dmg', 'LigandScout\_4...exe', 'LigandScout...tar.gz', and 'jm7b01342\_si\_0...csv', along with a 'Show all' button.

# Preparing the dataset

Autenticazione G x | Home | Sapienz: x | HomePage - Citt x | Online Chemical x | Online Chemical x | Journal of Medic x | Structure-Based x | +

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

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning | Save to Mendelej | TEMP >>

**Online chemical database**  
with modeling environment

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

v.3.0.96.1

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

### Model creator

Select model template and training set

### Save the model

Please enter your model's name:

#### Overview

Model name: pKi\_PLS\_[Dragon7 (blocks: 1-30)] - 336948 [\[rename\]](#) [Dragon7 (blocks: 1-30)]  
Temporal Public ID: [11990802](#) - use this link to share the model

Correl. limit: 0.0 Variance threshold: 0.01, Maximum value: 999999,

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

Data Set	#	R2	q2	RMSE	MAE
● Training set: 38 Mols	38 records	0.79 ± 0.06	0.79 ± 0.06	0.38 ± 0.04	0.31 ± 0.04

scale X: STANDARDIZE0 latent variables 3D by Corina  
5-fold cross-validation  
2990 pre-filtered descriptors

scale X: STANDARDIZE4 latent variables  
Y = -0.62 + 9.86E-6\**MW* - 0.00151\**AMW* + 1.08E-4\**Sv* + 7.59E-5\**Se* + 1.51E-4\**Sp* + 1.07E-4\**Si* - 0.0974\**Mv* - 0.346\**Me* - 0.0454\**Mp* + 0.133\**Mi* + 0.129\**GD* + 1.07E-4\**nAT* - 1.04E-4\**nSK* - 0.00147\**nTA* + 1.3E-4\**nBT* + 1.04E-4\**nBO* - 9.95E-4\**nBM* - 1.05E-4\**SCBO* - 4.56E-5\**RBN* - 0.0334\**RBF* - 6.43E-4\**nDB* - 7.49E-4\**nAB* + 3.0E-4\**nH* - 7.54E-5\**nC* + 0.0169\**nN* - 0.00448\**nO* - 3.04E-4\**nCL* + 0.00136\**nHM* - 2.29E-4\**nHet* + 0.00449\**nX* + 5.98E-4\**H%* - 0.00122\**C%* + 0.00224\**N%* - 0.00223\**O%* + 0.00189\**X%* + 7.52E-4\**nCsp3* -

LigandScout\_4....dmg ^ | LigandScout\_4.....exe ^ | LigandScout.....tar.gz ^ | jm7b01342\_si\_0....csv ^ | Show all x

# Preparing the dataset

accounts.google.com | Il mio Drive - Google Drive | Online Chemical Modeling Environment

ochem.eu/pendingtasks/tasks.do

Online chemical database with modeling environment v.3.0.96.2

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models A+ a- Privacy statement

Pending tasks X Task results X

Please enter your model's name:

**Overview**

Model name: pKi\_PLS\_[Dragon7 (blocks: 1-30)] - 337241 [rename] [Dragon7 (blocks: 1-30)]  
Temporal Public ID: 35823821 - use this link to share the model  
Correl. limit: 0.0 Variance threshold: 0.01, Maximum value: 999999,

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

Data Set	#	R2	q2	RMSE	MAE
Training set: 38 Mols	38 records	0.79 ± 0.05	0.78 ± 0.06	0.39 ± 0.04	0.32 ± 0.03

scale X: STANDARDIZE0 latent variables 3D by Corina 5-fold cross-validation 2990 pre-filtered descriptors

scale X: STANDARDIZE5 latent variables

$$Y = -11.2 - 7.65E-6 * MW - 0.00236 * AMW + 8.03E-5 * Sv + 5.85E-5 * Se + 9.21E-5 * Sp + 9.43E-5 * Si - 0.093 * Mv - 0.373 * Me - 0.0664 * Mp + 0.208 * Mi + 0.115 * GD + 8.62E-5 * nAT - 7.19E-5 * nSK - 0.00171 * nTA + 1.12E-4 * nBT + 1.2E-4 * nBO - 7.08E-4 * nBM - 4.88E-5 * SCBO + 1.48E-4 * RBN - 0.0125 * RBF - 7.32E-4 * nDB - 5.02E-4 * nAB + 2.37E-4 * nH - 1.02E-4 * nC + 0.0202 * nN - 0.00489 * nO - 6.76E-5 * nCL - 1.06E-4 * nHM + 2.88E-4 * nHet + 0.00475 * nX + 5.04E-4 * H% - 0.00119 * C% + 0.00284 * N% - 0.00245 * O% + 0.00216 * X% + 4.81E-4 * nCsp3 - 7.08E-4 * nCsp2 + 0.00211 * nCIC + 0.00211 * nCIR + 3.84E-4 * TRS + 3.84E-4 * Rperim + 0.0276 * MCD - 0.116 * RFD - 0.0814 * RCI + 0.00211 * NRS + 0.0422 * NNRS + 0.00286 * nR03 - 4.71E-4 * nR06 - 0.00301 * nBnz - 0.0191 * ARR + 6.28E-5 * D/Dtr03 - 8.59E-5 * D/Dtr06 + 1.15E-5 * D/Dtr09 - 0.00011 * D/Dtr12 + 0.00011 * D/Dtr15 + 0.00011 * D/Dtr18 + 0.00011 * D/Dtr21 + 0.00011 * D/Dtr24 + 0.00011 * D/Dtr27 + 0.00011 * D/Dtr30$$

Save Discard

# Preparing the dataset

The screenshot shows a web browser window with the URL <https://ochem.eu/modelconfigurator/choose.do>. The page header includes the logo for the Online chemical database with modeling environment, the version number v.3.0.96.1, and a welcome message for Prof. Ragno. The main content area displays a confirmation message: "Your model has been saved" followed by "Thank you for your cooperation." Below this, it lists "Your next possible actions are:" with two links: "Apply your model" and "View your model's properties". A blue arrow points to the "View your model's properties" link. The browser's taskbar at the bottom shows several files, including "LigandScout\_4...dmg", "LigandScout\_4...exe", "LigandScout...tar.gz", and "jm7b01342\_si\_0...csv".

# Preparing the dataset

accounts.google.com | Il mio Drive - Google Drive | Online Chemical Modeling Environm

ochem.eu/pendingtasks/tasks.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning G...

Online chemical database with modeling environment v.3.0.96.2

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models

Pending tasks X Task results X

### Model profile

Statistical parameters, tables, charts - all the information related to the model.

#### Overview

Model name: pKi\_PLS\_[Dragon7 (blocks: 1-30)] - 337241 [rename]  
Temporal Public ID: 35823821 - use this link to share the model

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

Data Set	#	R2	q2	RMSE	MAE
Training set: 38 Mols	38 records	0.79 ± 0.05	0.78 ± 0.06	0.39 ± 0.04	0.32 ± 0.04

[Dragon7 (blocks: 1-30)]  
Correl. limit: 0.0 Variance threshold: 0.01,  
Maximum value: 999999,  
scale X: STANDARDIZED0 latent variables 3D  
by Corina  
5-fold cross-validation  
-  
2990 pre-filtered descriptors

scale X: STANDARDIZED5 latent variables  
Y = -11.2 - 7.66E-6\*MW - 0.00236\*AMW +  
8.03E-5\*Sv + 5.86E-5\*Se + 9.21E-5\*Sp +  
9.43E-5\*Si - 0.093\*Mv - 0.373\*Me - 0.086\*Mp  
+ 0.208\*Mi + 0.115\*GD + 8.62E-5\*nAT -  
7.19E-5\*nSK - 0.00171\*nTA + 1.12E-4\*nBT +  
1.2E-4\*nBO - 7.08E-4\*nBM - 4.88E-5\*SCBO +  
1.48E-4\*RBN - 0.0125\*RBF - 7.32E-4\*nDB -  
5.02E-4\*nAB + 2.37E-4\*nH - 1.02E-4\*nC +  
0.0202\*nN - 0.00489\*nO - 6.76E-5\*nCL -  
1.08E-4\*nHM + 2.88E-4\*nHet + 0.00475\*nX +  
5.04E-4\*H% - 0.00119\*C% + 0.00284\*N% -  
0.00245\*O% + 0.00216\*X% + 4.81E-4\*nCsp3  
- 7.08E-4\*nCsp2 + 0.00211\*nClC +  
0.00211\*nClR + 3.84E-4\*TRS + 3.84E-  
4\*Rperim + 0.0276\*MCD - 0.116\*RFD -  
0.0814\*RCI + 0.00211\*NRS + 0.0422\*NNRS +  
0.00286\*nR03 - 4.71E-4\*nR06 - 0.00301\*nBnz  
- 0.0191\*ARR + 6.28E-5\*D/Dtr03 - 8.59E-  
5\*D/Dtr05 - 1.17E-5\*D/Dtr06 - 9.19E-  
5\*D/Dtr10 + 2.18E-5\*ZM1 - 4.66E-5\*ZM1V -

# Preparing the dataset

The screenshot displays the OCHEM web interface. The browser address bar shows `ochem.eu/pendingtasks/tasks.do`. The page header includes the logo "Online chemical database with modeling environment" and a version number "v.3.0.96.2". A navigation menu contains "Home", "Database", and "Models". A user greeting "Welcome, Dear Prof.Ragno!" is visible.

The main content area is divided into two tabs: "Pending tasks" and "Task results". The "Task results" tab is active, showing a scatter plot on the left and a list of model coefficients on the right. The scatter plot has a horizontal axis from 5.5 to 8.0 and a vertical axis from 5.5 to 6.5. A diagonal regression line is drawn through the data points. A blue arrow labeled "Measured value" points to a data point at approximately (6.5, 6.0).

Below the plot are several interactive options: "Download descriptors and model statistics", "MMP-based analysis (experimental)", "Create a copy of this model", "View configuration XML", and "Export configuration XML".

The right panel displays a long list of model coefficients, including terms like  $0.0814 \cdot \text{RCI} + 0.00211 \cdot \text{NRS} + 0.0422 \cdot \text{NNRS} + \dots$ .

# Preparing the dataset

accounts.google.com | Il mio Drive - Google Drive | Online Chemical Modeling Environm

ochem.eu/pendingtasks/tasks.do

Apps | Università | WWW | Mail | Ricerca | Vari | SoftwareVari | NatComp | METZ | IMP | Kahoot! | Learning G...

Online chemical database  
with modeling environment v.3.0.98.2

Welcome, Dear Prof.Ragno! My account Logout

Home Database Models

Pending tasks X Task results X Download descriptors and model statistics X

Data export  
Export the selected data as an Excel, CSV or SDF file

Please, select the items that you want to export:

[select all] [select unrestricted only] [select none]

- Structure (SMILES or SDF)
- CASRN
- RECORDID
- MOLECULEID
- External unique identifier
- Identifier in article (N)
- NAMES
- Introducers of the records
- Last modifiers of the records
- Publication IDs
- 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
- Inchi-key
- Merge information for the same molecule

Select the units to which the exported values will be converted:  
pKi -log(M)

# Preparing the dataset

The screenshot shows a web browser window with the URL `ochem.eu/pendingtasks/tasks.do`. The page title is "Online chemical database with modeling environment" (version 3.0.06.2). The user is logged in as "Prof.Ragno!". The main content area is titled "Download descriptors and model statistics" and contains the following elements:

- Instructions: "Please, select the items that you want to export."
- Selection options: "[select all] [select unrestricted only] [select none]"
- Checklist of exportable items (all are checked):
  - Structure (SMILES or SDF)
  - CASRN
  - RECORDID
  - MOLECULEID
  - External unique identifier
  - Identifier in article (N)
  - NAMES
  - Introducers of the records
  - Last modifiers of the records
  - Publication IDs
  - 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
  - Inchi-key
  - Merge information for the same molecule
- Unit selection: "Select the units to which the exported values will be converted:" with a dropdown menu showing "pKi -log(M)".
- Export buttons: "Get Excel file", "Get CSV file", "Get SDF file", and "Get R script".

A blue arrow points from the unit selection dropdown to the "Get Excel file" button.



# Preparing the dataset

The screenshot shows a web browser window with the URL `ochem.eu/pendingtasks/tasks.do`. The page header includes the logo for "Online chemical database with modeling environment" and a version number "v.3.0.98.2". A navigation bar contains links for "Home", "Database", and "Models". The main content area displays the following information:

File name: `pki_pls_[dragon7_blocks_1_30_]_337241.xls`

Your file is ready and the download will start in a few seconds. If the download does not start, you can use the direct link below.

Two buttons are visible: a green "Download" button and a brown "close this page" button. A blue arrow points to the "Download" button.