



On-line chemoinformatics tools for SAR/QSAR model development

Igor V. Tetko
Institute of Structural Biology
Helmholtz Zentrum Muenchen

13th October 2016, on-line course of the BIGCHEM project

Outline

Overview of existing on-line services on the web

- Individual groups web servers
- Projects web servers
- Model repositories
- Model development platforms

OCHEM

- Typical uses of OCHEM
- Model development
- Model interpretation
- Model evaluation



Individual groups web servers

<http://cdb.ics.uci.edu>

ChemDB Chemoinformatics Portal

Home

Molecules

ChemicalSearch

Find a molecule by its name, structure, or similarity to another molecule and filter the results.

Virtual Chemical Space

Interactively deconstruct a target molecule into possible chemical precursors and reassemble them into a combinatorial library of real or virtual molecules around the target.

MOLpro

Calculate or predict molecular properties other than 3D structure.

AquaSol

Predict aqueous solubility of small molecules using deep learning and ensembles.

COSMOS (downloadable)

Predict 3D molecular structures using open crystallography libraries.

COSMOS

Predict 3D molecular structures.

Reactions

Reaction Explorer

Learn and practice reactions, syntheses, and mechanisms interactively with support for: automated generation of problems, curved-arrow mechanism diagrams, and inquiry-based learning.

Reaction Predictor

Predict reaction outcomes and mechanisms using machine learning.

Tools

Smi2Depict

Generate 2D images from SMILES.

Babel

Convert between molecule file formats.

Reaction Processor

Generate product libraries.

Pattern Match Counter

Count functional groups (sub-structures).

Pattern Count Screen

Screen molecules by functional group count.

MSFragment

Fragment molecules for mass spec analysis.

Mass2Structure

Search ChemDB by monoisotopic mass and substructure filtering.

Protein Target Predictor

Predict activities of small molecules against a large set of protein targets

Datasets

Chemical datasets

Datasets for training and testing machine learning and other algorithms.

Prof. P. Baldi, UCI

<http://infochim.u-strasbg.fr/webserv/VSEngine.html>



Powered by ChemAxon

Welcome, **demo**

[QSAR-based Property Predictions](#)

[Add your compounds to ScreenDB](#)

[ScreenDB Query Tool](#)

[Conformer Generation](#)

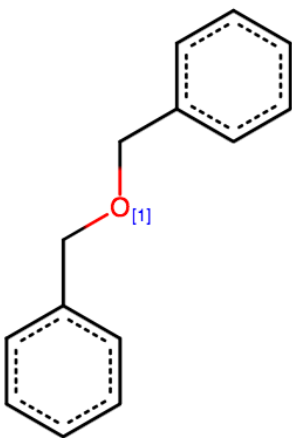
[Project Deletion](#)

Your 6 stored projects:

- [aaa](#)
- [new](#)
- [new_test](#)
- [TTT](#)
- [test](#)
- [tttt](#)

☒ sorted by creation date
☐ sorted by name

[ReFresh](#)



1 [Get Structure](#)

PhysicoChemical Properties

- MichaelReactProfiler : Feasibility Profile Predictor of Michael Reactions: Enter potential Michael addition products
- HBAcceptor : [Hydrogen Bond Acceptor Strength Predictor of user-labeled acceptors \(pKBHX\)](#)
- HBAcceptAuto : [Hydrogen Bond Acceptor Strength Predictor of automatically detected acceptors \(pKBHX\)](#)
- HalogenBond : Halogen Bond (Charge Transfer Complex with Diiodine) Strenght Predictor of user-labeled acceptors (pKI2)
- HalogenBondAuto : Halogen Bond (Charge Transfer Complex with Diiodine) Strenght Predictor of automatically detected acceptors (pKI2)

Select property to predict :

MichaelReactProfiler

HBAcceptor

HBAcceptAuto

HalogenBond

☐ Enable generation of HTML result pages with compounds/page and [GO!](#)

Available Prediction Results

Hbond_acceptor [.csv format](#) [.html format](#)

Prof. A. Varnek, Strasbourg

http://tox.charite.de/tox



PROTOX - Prediction of Rodent Oral TOXicity

PROTOX

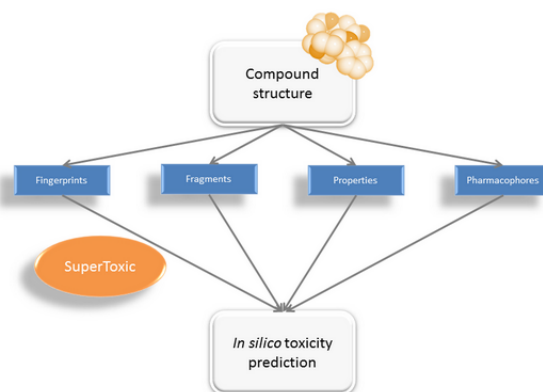
Welcome to **PROTOX**, a webserver for the prediction of oral toxicities of small molecules in rodents! The prediction of compound toxicities is an important part of the drug design development process. Computational toxicity estimations are not only faster than the determination of toxic doses in animals, but can also help to reduce the amount of animal experiments. To read more about reducing animal testing, go to [Animal Ethics 3R](#). Our webserver is based on chemical similarities between compounds with known toxic effects and the presence of toxic fragments. To predict the oral toxicity of your compounds, please click [here](#). For a description of the server and tutorials, go to [FAQ](#). To see statistics about our training set as well as the cross-validation results, please go to [Statistics](#). Should you have further questions, do not hesitate to [contact us](#)!

Citations

Drwal M.N., Banerjee P., Dunkel M., Wettig M.R., Preissner R.: **ProTox: a web server for the *in silico* prediction of rodent oral toxicity** Nucleic Acids Res (Web server issue 2014); [NAR](#)

Schmidt U., Struck S., Gruening B., Hossbach J., Jaeger I.S., Parol R., Lindequist U., Teuscher E., Preissner R.: **SuperToxic: a comprehensive database of toxic compounds**. Nucleic Acids Res 37 (Database issue 2009): D295-9; [NAR](#)

Predict compound toxicity



PROTOX enables uncomplicated predictions of oral rodent toxicities. To conduct a toxicity prediction, please click on the image.

Toxic doses and toxicity classes

Toxic doses are often given as LD50 values in mg/kg body weight. The LD50 is the median lethal dose meaning the dose at which 50% of test subjects die upon exposure to a compound.

Toxicity classes are defined according to the globally harmonized system of classification of labelling of chemicals (GHS):

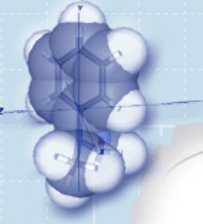
- Class I: fatal if swallowed ($LD50 \leq 5$ mg/kg)
- Class II: fatal if swallowed ($5 < LD50 \leq 50$ mg/kg)
- Class III: toxic if swallowed ($50 < LD50 \leq 300$ mg/kg)
- Class IV: harmful if swallowed ($300 < LD50 \leq 2000$ mg/kg)
- Class V: may be harmful if swallowed ($2000 < LD50 \leq 5000$ mg/kg)
- Class VI: non-toxic ($LD50 > 5000$ mg/kg)

Dr. R. Preissner, Charite University of Medicine

A stylized, light blue robot arm graphic is positioned in the upper right corner of the slide. It features a circular joint at the top, a long rectangular segment, and a circular end effector with a grey center.

Projects web servers

<http://www.vcclab.org>



Home
About
Partners
Software
Articles
Servers
Jobs
Web Services
How to cite?
Contact

<http://www.vcclab.org>

Virtual Computational Chemistry Laboratory

Home About Partners Software Articles Servers Jobs Web Services How to cite? Contact

Welcome! Willkommen! Bienvenue! Benvenuto! Bienvenida! Laskavo prosymo! Dobro pozhalovat'!

This site provides free on-line tools, which we hope you will find helpful in performing computational chemistry, ADME/T and chemoinformatics tasks including the building and visualisation of chemical structures, the calculation of molecular properties and the analysis of relationships between chemical structure and properties. All these tools are developed, provided and supported by the VCCLAB partners.

Please feel free to browse our site and take advantage of the various on-line cheminformatics services provided. However, the calculation algorithms are running on remote servers, communicating via the Internet, so we are unable to control the speed of access to these services. An overview of on-line WWW resources for Phys-Prop calculations is also available.

If you see <http://146.107.217.178> click <http://www.vcclab.org> to correct the address. We could fix it automatically but some users can only access physical address due to DNS problems.

Java security issues: recently Java has dramatically increased security requirements to applets. Thus, please, follow instructions in this [FAQ](#) to correctly setup access to the software.

PREV TOP

Home About Partners Software Articles Servers Jobs Web Services How to cite? Contact

ON-LINE SOFTWARE

ALOGPS 2.1

ASNN

E-BABEL

PNN

PCLIENT

E-DRAGON 1.0

PLS

UFS

SPC

A first QSAR web tool: Polynomial Neural Networks

Virtual Computational Chemistry Laboratory
<http://www.vcclab.org>

Welcome to the PNN program!

© MIPS/VCCLAB PNN parameters Login submit your task

Type of data [ANALYSIS](#) fit data [LOO](#) of the training set no

Maximum [DEGREE](#) of the model + 3-order Number of [ITERATIONS](#) 15

Maximal [NUMBER](#) of terms in model 4 Number of [STORED](#) models 3

[CRITERION](#) to select the best models FPE [VALIDATION](#) set in RR criterion

Details of calculated results ([PRINT](#))

☒ display calculated vs experimental values ☐ save input data in stdout

☐ save calculated values in stdout ☐ save detailed statistics of analysis in stdout

☐ save statistics of input data in stdout ☐ select all options

Specify polynomial neural network parameters. Click the underlined links for more information.

I. V. Tetko, T. I. Aksenova, V. V. Volkovich, T. N. Kasheva, D. V. Filipov, W. J. Welsh, D. J. Livingstone, A. E. P. Villa, *SAR QSAR Environ. Res.* **2000**, *11*, 263-280.

<http://knimewebportal.cosmostox.eu>

COSMOS KNIME WEBPORTAL

Integrated *In Silico* Models for the Prediction of Human
Repeated Dose Toxicity of COSMetics to Optimise Safety

Welcome to the COSMOS KNIME WebPortal

COSMOS is a unique collaboration addressing the safety assessment needs of the cosmetics industry, without the use of animals. The main aim of COSMOS is to develop freely available tools and workflows to predict the safety to humans following the use of cosmetic ingredients.

The models developed within COSMOS have been implemented into KNIME workflows. The KNIME Analytics Platform integrates access to chemical databases, data processing and analysis, modelling approaches, profiling of structures and calculation of properties in a flexible way.

The COSMOS KNIME WebPortal versions of these workflows allow users, not experienced with the software, to execute the workflows in a web browser, without local software installation required. The results are downloadable as pdf reports and other formats e.g. Excel sheets.

The models are documented and user guidance is available through COSMOS Space. A list of all workflows available can be found [here](#).

Login

Username

Password

Login

How to get your login

Registration is free. To get your login details from the COSMOS Space follow the following steps:

1. Visit <http://cosmosspace.cosmostox.eu>
2. Click on Login
3. Click on register
4. Fill in your email address and choose a password.

<http://cosmospace.cosmostox.eu/app/space/cosmosshare/ckwdprojects?pn=2>



[ABOUT](#) [CONTACT](#) [LOGOUT](#)

[Home](#) [COSMOS Share](#) [COSMOS DB v1 dump](#) [COSMOS KNIME Workflows](#) [My COSMOS Space](#) [My Gallery](#) [My Profile](#)

COSMOS Share - COSMOS KNIME Workflow Documentations

List of all available workflows with links to the descriptions and user guidance.

The workflows are freely available for direct execution in a web browser at [COSMOS KNIME WebPortal](#) - login with your COSMOS Space credentials.

Title	Tags
Hepatotoxicity Alerts (WebPortal version)	Hepatotoxicity, Molecular Initiating Events (MIEs), grouping chemicals, structural alert
LXR binding prediction (PLS-DA/RDKit) (Desktop version)	LXR; QSAR; nuclear receptor; classification
Mitochondrial Toxicity Alerts (WebPortal version)	Mitochondrial toxicity, Molecular Initiating Events (MIEs), grouping chemicals, structural alert
Potential NR ligands and alerts towards hepatosteatosis (Desktop Version)	NR, endocrine disruptor, fatty liver, hepatosteatosis, nuclear receptor
Chemical Space Analysis (Report)	Chemical Space; Explorative Analysis; PCA
Covalent Protein Binding Alerts (WebPortal version)	Molecular Initiating Events (MIEs), Protein binding, grouping chemicals, structural alert
Covalent DNA Binding Alerts (WebPortal version)	DNA binding, Molecular Initiating Events (MIEs), grouping chemicals, structural alert
Chemical Space Analysis	Chemical Space; Explorative Analysis; PCA
PAMPA logPm predictor (Desktop version)	GIA, PAMPA permeability, QSAR, logPm
Ranking (WebPortal version)	ranking; priority setting; screening; prediction

<http://www.vega-qsar.eu/>

VEGA TRY THE NEW TOXREAD DOWNLOAD VEGA ABOUT QSAR NEWS COMMUNITY

EUToxRisk ToxBank NIHS caleidos PROSIL

Our Vision

Our Mission

Our System

News & Updates

June, 2016
Working on behalf of UBA [The Platform for prioritising possible persistent, bioaccumulative and toxic \(PBT\) chemicals relies on VEGA](#)

June, 2016

On site

June, 2016
How to interpret VEGA results [Detailed explanation on how to interpret results has been added](#)

June, 2016
An example on how to use VEGA has been provided within the guidance on "how to use and report QSARs" [released by ECHA](#)

Our Community

Sponsored by

- ANTARES Project
- CALEIDOS Project
- CAESAR Project
- ORCHESTRA Project
- ToxBank Project
- PROSIL Project
- EUToxRisk Project
- LIFE-Edesia Project
- Ministero della Salute

Developed by

- IRCCS - Istituto di Ricerche Farmacologiche Mario Negri
- CRS4
- KnowledgeMiner Software
- Kemijärvi Institut
- Kode S.R.L.
- Politecnico di Milano
- US EPA
- UK FERA

Acknowledgements

- ARCHE Consulting
- Aarhus University
- Centro REACH S.r.l.
- Direcção-Geral da Saúde
- EquiTox
- INERIS
- INFOTOX
- Technology for Growth
- Universitat Rovira i Virgili
- Umweltbundesamt GmbH
- Ministry of health, Labour and Welfare, Japan
- ISPRA
- Ministero dell'Ambiente

http://toxpredict.org

Unable to connect

Firefox can't establish a connection to the server at apps.ideaconsult.net:8080.

- The site could be temporarily unavailable or too busy. Try again in a few moments.
- If you are unable to load any pages, check your computer's network connection.
- If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web.

Try Again



Model repositories

http://qsar.db.jrc.it/qmrf/



(Q)SAR Model Reporting Format Inventory



Log in Register

Home Search documents Search structures

All published QMRF documents **(109)** are available for download and can be searched either through free text queries or by several predefined fields. All substances, available in the QMRF Database, can be searched by exact or similar structure.

What is QMRF Database?

Do you need to register to use the QMRF Database?

Please register only if you wish to submit a QMRF. Registration is not necessary if you only wish to search the database and access information on QMRFs.

[Help](#)

How to create an QMRF Document?

- log in into QMRF Database and use the *New document* tab;
- by [QMRF editor](#) : once started, it will create shortcut on your desktop and can be started later even offline.

Most recent QMRF documents

#	QMRF#	Title	Last updated	View	Download
1	Q50-54-55-501	BIOVIA toxicity prediction model – Ames Mutagenicity	2016-6-17 14:58		
2	Q51-54-55-502	BIOVIA toxicity prediction model – rat oral LD50	2016-6-17 14:58		
3	Q50-54-55-503	BIOVIA toxicity prediction model – NTP carcinogenicity call (male rat)	2016-6-17 14:58		

For information about this site please contact JRC-IHCP-COMPUTOX@ec.europa.eu

This page has been accessed 508795 times since 2008-07-03 15:25:48.0

Developed by Ideaconsult Ltd. (2007-2008) on behalf of JRC



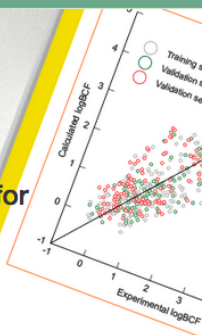
<http://qsardb.org>

REPOSITORY QDB RESOURCES NEWS CONTACTS

Welcome to QsarDB

QsarDB is a smart repository for (Q)SAR/QSPR models and datasets, ready for discovery, exploring, and citing.

[Browse repository](#) [Get started](#) [Downloads](#) [About QsarDB](#)



How to cite **Publications** **Credits** **Policies**

Recent submissions to the repository

07 Oct 2016	Ran, Y.; He, Y.; Yang, G.; Johnson, J. L. H.; Yalkowsky, S. H. Estimation of aqueous solubility of organic compounds by using the general solubility equation. Chemosphere 2002, 48, 487–509.
27 Sep 2016	Li, F.; Wu, H.; Li, L.; Li, X.; Zhao, J.; Peijnenburg, W. J. G. M. Docking and QSAR study on the binding interactions between polycyclic aromatic hydrocarbons and estrogen receptor. Ecotoxicology and Environmental Safety 2012, 80, 273 - 279.
20 Sep 2016	Papa, E.; Pilutti, P.; Gramatica, P. Prediction of PAH mutagenicity in human cells by QSAR classification. SAR and QSAR in Environmental Research 2008, 19, 115–127.
19 Sep 2016	Kar, S.; Deeb, O.; Roy, K. Development of classification and regression based QSAR models to predict rodent carcinogenic potency using oral slope factor. Ecotoxicology and Environmental Safety 2012, 82, 85–95.



Model development platforms

http://chembench.mml.unc.edu



Home	My Bench	Datasets	Modeling	Prediction
------	----------	----------	----------	------------

Accelerating chemical genomics research

Chembench is a free portal that enables researchers to mine available chemical and biological data. Chembench can help researchers rationally design or select new compounds or compound libraries with significantly enhanced hit rates in screening experiments.



Chembench provides cheminformatics research support to molecular modelers, medicinal chemists and quantitative biologists by integrating robust model builders, property and activity predictors, virtual libraries of available chemicals with predicted biological and drug-like properties, and special tools for chemical library design. Chembench was initially developed to support researchers in the [Molecular Libraries Probe Production Centers Network \(MLPCN\)](#) and the Chemical Synthesis Centers.

Please log in

Username:

Password:

Log in

Or, log in as a guest

[Forgot your password?](#)

[Register an account](#)

Help & Resources

- [Chembench Overview](#)
- [Workflows & Methodology](#)

Site Stats

Total visitors: 1229488

Registered users: 503

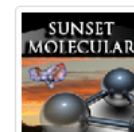
Jobs completed: 59624

Compute time used: 76.94 years

Current users: 5

Running jobs: 34

We thank the following commercial sponsors for their support:



Data storage and model development: <http://ochem.eu>

**Online chemical database**
with modeling environment

v2.4.

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

Home ▾ Database ▾ Models ▾ A+ a-

Welcome to OCHEM! Your possible actions

Explore OCHEM data

Search chemical and biological data: experimentally measured, published and exposed to public access by our users. You can also [upload your data](#).

Create QSAR models

Build QSAR models for predictions of chemical properties. The models can be based on the experimental data published in our database.

Run predictions

Apply one of the available models to predict property you are interested in for your set of compounds.

Screen compounds with ToxAlerts

Screen your compound libraries against structural alerts for such endpoints as mutagenicity, skin sensitization, aqueous toxicity, etc.

Optimise your molecules

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

Tutorials

Check our video tutorials to know more about the OCHEM features.

Our acknowledgements

Check out the properties available on OCHEM

OCHEM contains 1280459 experimental records for about 499 properties collected from 12428 sources

Melting Point

LogL(water) Cbrain/Cblood LogD Cblood/Cair Cbrain/Cair Cfat/Cair

Cliver/Cair Cmuscle/Cair SIF solubility logP(+) logP(-)

Water solubility

LogL(blood) LogL(brain)

LogL(fat) LogL(heart) LogL(kidney) LogL(liver) LogL(lungs)

LogL(muscle) LogL(oil) LogL(plasma) LogBPR LogCSFPR ER

fu(brain) P/Papp Biodistribution(kidney) Biodistribution(liver)

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

Cbrain/Cplasma **IC50** Papp(Caco-2)

Papp(MDCK) P(brain) Oral absorption LIC 50

pK(1/logK) Cliver/Cplasma Clung/Cplasma Cheart/Cplasma

Ckidney/Cplasma Cbrain/Cserum Cfat/Cplasma Cmuscle/Cplasma

Cskin/Cplasma Papp ratio(Caco-2) Papp(MBUA)

Plasma protein binding

Papp(HPBEC)

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

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

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

Papp ratio(SV-ARBEC) Papp ratio(MDCK rat) Papp ratio(MDCK rat)

Latest active users

 Charleshen: Mr. SHEN Charles
seconds ago

 enamine: Dr. Ivan Ivanov
seconds ago

 Reshmi: Mrs. D Reshmi
about 11 hours ago

 Guangchao Chen: Mr. Guangchao Chen
about 14 hours ago

 tacristy: Mr. Tim Cristy
about 15 hours ago

 bfrindt: Mr. Benjamin Frindt
about 16 hours ago

Latest published models

 Melting Point model published by itetko
2 months ago

 Melting Point model published by romney
2 months ago

 IC50 HIV model published by nizamibilal1064
5 months ago

 LEL model published by novserj
more than a year ago

 logERRBA (qualitative) model published by aveima
more than a year ago



Typical use of OCHEM



FILTERS

▼ SOURCE

Article/Source [\[select\]](#)

Page Table

▼ PROPERTY

Activity/Property [\[select\]](#)

☐ Hide records without property

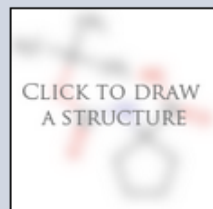
► CONDITIONS

▼ MOLECULE FILTERS

Name / OCHEM ID [?](#) / Inchi-Key

Similarity/substructure search

Draw a structure and search all the molecules containing it or similar to it



Molecular mass [?](#)

between and

► ADVANCED MOLECULE FILTERS

▼ MISCELLANEOUS

Current set [?](#)

Show all

Data origin and quality:

Data introducers: All users

Data visibility: Public and private

Data from other users: Only approved data



1 - 5 of 1075221



Records



5

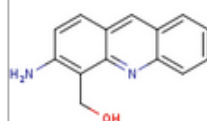
Items on page

1

of 215045

>

>>



molecule profile

● logPow = 1.77 (in Log unit)

Charmantray, F et al

4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...

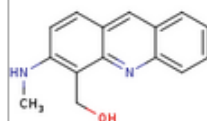
N: 1 P: 970 T: 1

J. Med. Chem. 2003; 46 (6) 967-77

MoleculeID: M11715

Public record

RecordID: R1
09:35, 19 Mar 09 / 10:04, 13 Oct 11
i.tetko [✉](#) / itetko [✉](#)



molecule profile

● logPow = 2.1 (in Log unit)

Charmantray, F et al

4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...

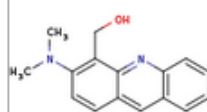
N: 2 P: 970 T: 1

J. Med. Chem. 2003; 46 (6) 967-77

MoleculeID: M9560

Public record

RecordID: R2
09:35, 19 Mar 09
i.tetko [✉](#)



molecule profile

● logPow = 2.2 (in Log unit)

Charmantray, F et al

4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...

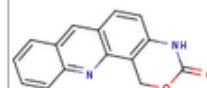
N: 3 P: 970 T: 1

J. Med. Chem. 2003; 46 (6) 967-77

MoleculeID: M10111

Public record

RecordID: R3
09:35, 19 Mar 09
i.tetko [✉](#)



molecule profile

● logPow = 1.64 (in Log unit)

Charmantray, F et al

4-Hydroxymethyl-3-aminoacridine derivatives as a new family ...

N: 4 P: 970 T: 1

J. Med. Chem. 2003; 46 (6) 967-77

MoleculeID: M9777

Public record

RecordID: R4
09:35, 19 Mar 09
i.tetko [✉](#)

Working with ToxAlerts



Online chemical database
with modeling environment

Home ▾ Database ▾ Models ▾ Moderation ▾

Compound properties

- Molecules
- Properties
- Conditions
- Units
- Articles/Books
- Journals
- ToxAlerts** ▸
 - ToxAlerts home
 - View alerts
 - Screen compounds against alerts
 - Upload new alerts
- Pathways ▸
- Baskets
- Tags
- Set area of interest...

our possible actions

data: experimentally
ed to public access by
your data.

Welcome to ToxAlerts!

Structural alerts (also known as "toxicophores") are molecular patterns known to be associated with particular type of toxicity. The studies performed last decade has shown that structural alerts is an efficient technique to detect potentially toxic chemicals. Screening chemical compounds against known structural alerts can be a good practice to complement the QSAR models and to help interpreting their predictions.

ToxAlerts is a platform for screening chemical compounds against structural alerts. The platform allows to search structural alerts, introduce your own alerts and screen chemical libraries for alert-hitting compounds.

View available alerts

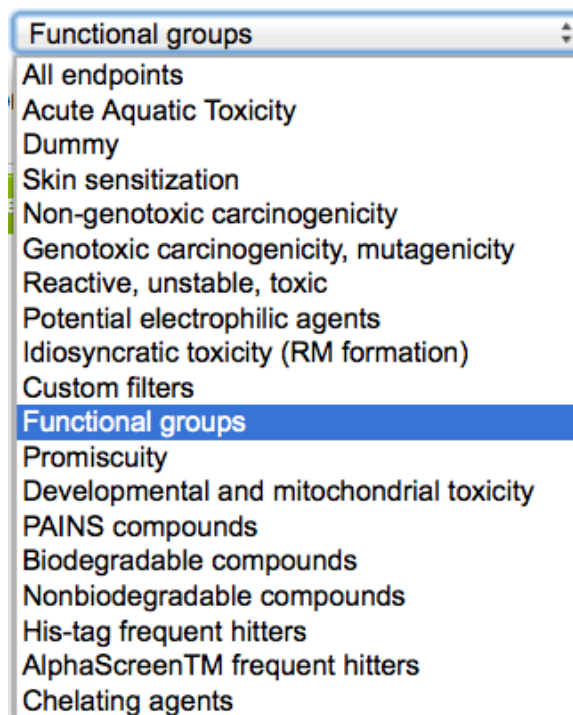
Upload new alerts

Screen your molecules

In case of any questions, ideas, or problems with the software, feel free to [drop us a message](#). We highly appreciate any feedback from you!

ToxAlerts

- Screening of compounds against published toxicity alerts, groups, frequent hitters
- Filter alerts by endpoints or publications
- Create or upload custom SMARTS rules



Sushko et al, JCIM, 2012, 52(8):2310-6.

Analysis of virtual libraries

ToxAlerts: Screening results
The compounds that matched any alerts grouped by endpoints, publications and by alerts themselves

ENDPOINTS X

- ☐ Functional groups 1890 compounds

PUBLICATIONS

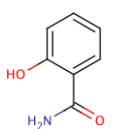
- ☐ 2005 CheckMol 1880 compounds
- ☐ 2012 Salmina 1890 compounds

DETECTED ALERTS

Alert	Source	Count
☐ Hydroxy compounds: alcohols or phenols	2005 CheckMol	416 compounds
☐ Phenols	2005 CheckMol	211 compounds
☐ Carboxylic acid derivatives	2005 CheckMol	542 compounds
☐ Carboxylic acid amides	2005 CheckMol	101 compounds
☐ Carboxylic acid primary amides	2005 CheckMol	14 compounds
☐ Aromatic compounds	2005 CheckMol	993 compounds
☐ Arenes	2005 CheckMol	915 compounds
☐ Nonmetals	2012 Salmina	1890 compounds
☐ Chalcogens (oxygen group)	2012 Salmina	1433 compounds
☐ Pnictogens (nitrogen group)	2012 Salmina	800 compounds
☐ Tetragens (carbon group)	2012 Salmina	1890 compounds
☐ Nitriles	2005 CheckMol	55 compounds
☐ Carboxylic acid amidines	2005 CheckMol	4 compounds
☐ Heterocyclic compounds	2005 CheckMol	309 compounds
☐ Five-membered heterocycles (LS)	2012 Salmina	143 compounds
☐ Five-membered heterocycles with two heteroatoms (LS)	2012 Salmina	61 compounds
☐ Unsaturated five-membered heterocycles with two heteroatoms (LS)	2012 Salmina	6 compounds
☐ Five-membered heterocycles (HS)	2012 Salmina	69 compounds
☐ Five-membered heterocycles with two heteroatoms (HS)	2012 Salmina	31 compounds
☐ Unsaturated five-membered heterocycles with two heteroatoms (HS)	2012 Salmina	5 compounds
☐ Imidazolines (HS)	2012 Salmina	2 compounds
☐ 2-Imidazolines (HS)	2012 Salmina	2 compounds
☐ Carbonyl compounds: aldehydes or ketones	2005 CheckMol	177 compounds
☐ Aldehydes	2005 CheckMol	53 compounds
☐ Thioethers	2005 CheckMol	14 compounds
☐ Dialkylthioethers	2005 CheckMol	5 compounds
☐ Carbonyl compounds: aldehydes and ketones	2005 CheckMol	170 compounds
☐ Amines	2005 CheckMol	316 compounds
☐ Primary aliphatic amines	2005 CheckMol	59 compounds
☐ Primary amines	2005 CheckMol	180 compounds

View records for the filtered compounds Tag the 1890 filtered molecules Export the screening results

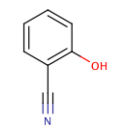
1 - 15 of 1890 15 items on page 1 of 126 > >>



molecule profile

MoleculeID: M10446

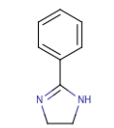
Hydroxy compounds: alcohols or phenols (for Functional groups in 2005 CheckMol)
 Phenols (for Functional groups in 2005 CheckMol)
 Carboxylic acid derivatives (for Functional groups in 2005 CheckMol)
 Carboxylic acid amides (for Functional groups in 2005 CheckMol)
 Carboxylic acid primary amides (for Functional groups in 2005 CheckMol)
 Aromatic compounds (for Functional groups in 2005 CheckMol)
 Arenes (for Functional groups in 2005 CheckMol)
 Nonmetals (for Functional groups in 2012 Salmina)
 Chalcogens (oxygen group) (for Functional groups in 2012 Salmina)
 Pnictogens (nitrogen group) (for Functional groups in 2012 Salmina)
 Tetragens (carbon group) (for Functional groups in 2012 Salmina)



molecule profile

MoleculeID: M1364

Hydroxy compounds: alcohols or phenols (for Functional groups in 2005 CheckMol)
 Phenols (for Functional groups in 2005 CheckMol)
 Nitriles (for Functional groups in 2005 CheckMol)
 Aromatic compounds (for Functional groups in 2005 CheckMol)
 Arenes (for Functional groups in 2005 CheckMol)
 Nonmetals (for Functional groups in 2012 Salmina)
 Chalcogens (oxygen group) (for Functional groups in 2012 Salmina)
 Pnictogens (nitrogen group) (for Functional groups in 2012 Salmina)
 Tetragens (carbon group) (for Functional groups in 2012 Salmina)



molecule profile

MoleculeID: M24113

Carboxylic acid amidines (for Functional groups in 2005 CheckMol)
 Aromatic compounds (for Functional groups in 2005 CheckMol)
 Arenes (for Functional groups in 2005 CheckMol)
 Heterocyclic compounds (for Functional groups in 2005 CheckMol)
 Nonmetals (for Functional groups in 2012 Salmina)
 Pnictogens (nitrogen group) (for Functional groups in 2012 Salmina)
 Tetragens (carbon group) (for Functional groups in 2012 Salmina)
 Five-membered heterocycles (LS) (for Functional groups in 2012 Salmina)
 Five-membered heterocycles with two heteroatoms (LS) (for Functional groups in 2012 Salmina)
 Unsaturated five-membered heterocycles with two heteroatoms (LS) (for Functional groups in 2012 Salmina)
 Five-membered heterocycles (HS) (for Functional groups in 2012 Salmina)
 Five-membered heterocycles with two heteroatoms (HS) (for Functional groups in 2012 Salmina)
 Unsaturated five-membered heterocycles with two heteroatoms (HS) (for Functional groups in 2012 Salmina)
 Imidazolines (HS) (for Functional groups in 2012 Salmina)
 2-Imidazolines (HS) (for Functional groups in 2012 Salmina)

An easy fast way to identify possibly problematic compounds
Comprehensive grouping of compounds according to a given feature

Functional groups

Online chemical database
with modeling environment

Welcome, Guest! [Logout](#)

Home Database Models

ToxAlerts: Structural alerts browser
Here you can browse structural alerts for various toxicological endpoints

FILTERS

Article:
2005 CheckMol

Endpoint / Filter type:
Functional groups

Name / Alert ID:

☐ Show only approved alerts

Upload new alerts Screen compounds

101 - 200 of 204

100 Items on page 2 of 3

Thioxohetarenes
R = H, alkyl, aryl; any heteroaromatic compound with a C=S structure
SMARTS: [S](=C)[nH]1cncnc1
Endpoint: Functional groups
[CheckMol](#)
[List of functional groups recognized by checkmol...](#)
2005;
Alert ID: TA1207

Iminohetarenes
R1, R2 = H, alkyl, aryl; any heteroaromatic compound with a C=N structure
SMARTS: [N]=C[nH]1cncnc1
Endpoint: Functional groups
[CheckMol](#)
[List of functional groups recognized by checkmol...](#)
2005;
Alert ID: TA1208

Orthocarboxylic acid derivatives
R = H, alkyl, aryl; Y = OH, alkoxy, aryloxy, (substituted) amino, etc.
SMARTS: [Y]C(=O)R
Endpoint: Functional groups
[CheckMol](#)
[List of functional groups recognized by checkmol...](#)
2005;
Alert ID: TA1209

Carboxylic acid orthoesters
R1 = H, alkyl, aryl; R2, R3, R4 = alkyl, aryl
SMARTS: RC(=O)OC(OR2)OR3
Endpoint: Functional groups
[CheckMol](#)
[List of functional groups recognized by checkmol...](#)
2005;
Alert ID: TA1210

Carboxylic acid amide acetals
R1, R2, R3 = H, alkyl, aryl; R4, R5 = alkyl, aryl
SMARTS: RC(=O)OC(OR4)OR5
Endpoint: Functional groups

Screening Artefacts

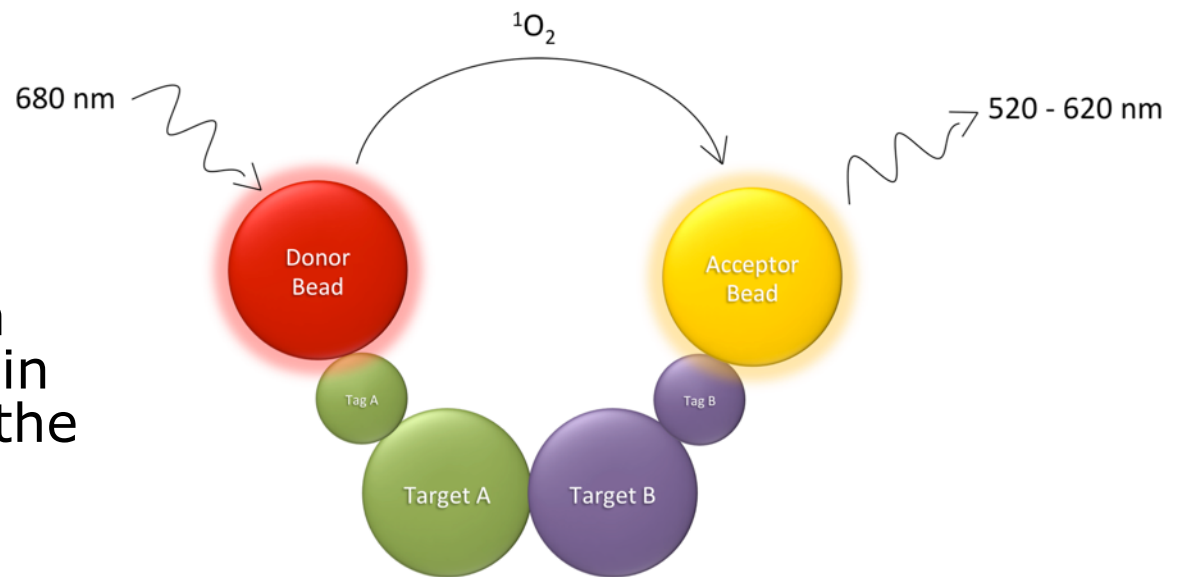
singlet oxygen
quenching

color quenching

auto-fluorescence

disruption of the
interaction between
the tag of the protein
and binding site of the
detection system

AlphaScreenTM



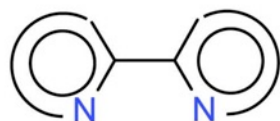
Pan- Assay Interference Compounds (PAINS) filters by Baell and Holloway, 2010.

HIS tag frequent hitters, Schorpp, K. et al *J. Biomol. Screen.* 2014, 19(5):715-726.

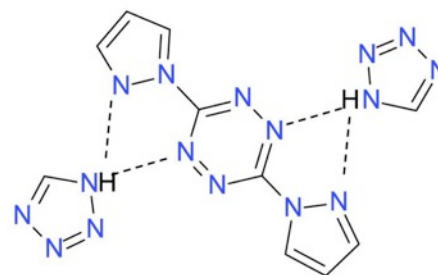
GST tag frequent hitters, Brenke, J.K. et al *J. Biomol. Screen.* 2016, 21(6):596-607.

MOA of AlphaScreenTM-HIS-FH

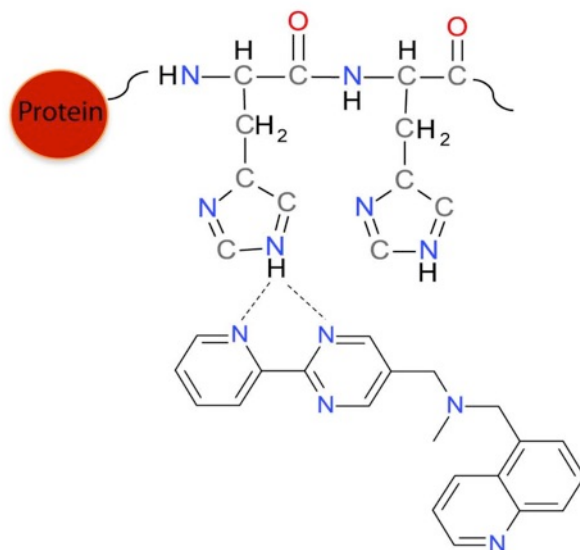
A



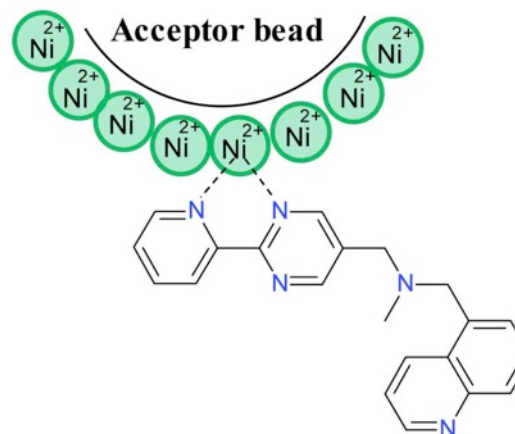
B



C



D



Schoorp et al, J Biomol Screen. 2014, 19(5):715-726.

SetCompare tool for data analysis

 **Online chemical database**
with modeling environment

Home ▾ Database ▾ **Models ▾** Moderation ▾

Welcome

Explore OC
Search chemical
measured, publish
our users. You can

Create QSAR models
Build QSAR models for predictions of chemical
properties. The models can be based on the
experimental data published in our database.

- Create a model
- Apply a model
- Create multiple models
- Open predictor**
- View pending tasks
- Calculate descriptors
- SetCompare utility**

 **SetCompare:** Select the sets to compare

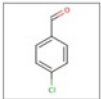
The SetCompare utility is experimental. It allows you to compare two sets of molecules based on their structural features.
Please, provide the two sets available options below.

1 Select the compounds in the **first set**:

Upload compounds from a file
(SDF/MOL2/SMILES/Excel sheet)

☐ Provide a Name/CAS-RN/SMILES

☒ Draw Molecule
(click on depiction to the right to draw)



☐ Choose a previously prepared set: [...]

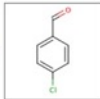
☐ Select molecules by a tag: [...]

2 Select the compounds in the **second set**:

Upload compounds from a file
(SDF/MOL2/SMILES/Excel sheet)

☐ Provide a Name/CAS-RN/SMILES

☒ Draw Molecule
(click on depiction to the right to draw)



☐ Choose a previously prepared set: [...]

☐ Select molecules by a tag: [...]

Examples of scaffolds analysis



SetCompare: Comparison results

The comparison summary of the two selected sets

The following table shows the features (molecular descriptors) that were significantly overrepresented in one of the two. It includes appearance counts of the features in each set and the p-Value of such a distribution.

[Export results as a CSV file](#)

1 - 15 of 253

15 it

Descriptor	In set 1 (13785 molecules)	In set 2 (228174 molecules)	Enrichment factor	p-Value
	3232 (23.4%)	26196 (11.5%)	2.0	1.33E-315
Pnictogens Group 15: the nitrogen family N P As Sb Bi	13235 (96.0%)	202449 (88.7%)	1.1	1.42E-198
	3257 (23.6%)	79741 (34.9%)	1.5	-1.25E-172
	280 (2.0%)	352 (0.2%)	13.2	4.21E-172
	2063 (15.0%)	18761 (8.2%)	1.8	2.23E-140

Pyrolysis vs. melting point



SetCompare: Comparison results

The comparison summary of the two selected sets

The following table shows the features (molecular descriptors) that were significantly overrepresented in one of the two. It includes appearance counts of the features in each set and the p-Value of such a distribution.

[Export results as a CSV file](#)

1 - 15 of 128

Descriptor	In set 1 (141 molecules)	In set 2 (20007 molecules)	Enrichment factor	p-Value
	11 (7.8%)	4 (0.0%)	390.2	1.77E-21
	9 (6.4%)	3 (0.0%)	425.7	6.72E-18
	7 (5.0%)	0 (0.0%)	Inf.	7.07E-16
	7 (5.0%)	2 (0.0%)	496.6	2.52E-14
	6 (4.3%)	0 (0.0%)	Inf.	1.06E-13

HIV Envelope glycoprotein GP120

Application of models

Home Database Models

Welcome to OChem

Explore OChem
Search chemical databases, published data, or user data. You can also view pending tasks or published tasks.

Create QSAR models
Build QSAR models, calculate descriptors, and store them in the database.

Run predictions
Apply one of the available models to predict property values for your set of compounds.

Screen compounds with ToxAlerts
Screen your compound libraries against structural alerts for such endpoints as mutagenicity, skin sensitization, aqueous toxicity, etc.

Tutorials
Check our video tutorials to know more about the OChem features.

Our acknowledgements

Feedback and help

User's manual
Check an online user's manual

Create a model
Apply a model
Create multiple models
Open predictor
View pending tasks
View published tasks
SetCompare utility
MolOptimiser
Calculate descriptors
Descriptors storage

Check out the properties available on OChem
OChem contains 1666815 experimental records for about 43 properties collected from 23262 sources

Melting Point logPow
logS CYP450 modulation
LC50 aquatic log(IGC50-1)

Boiling Point AMES
DMSO Solubility LogKoc
BCF MPProperty 9 chembl_IC50

chembl_Ki chembl_pIC50
chembl_pKB chembl_KB chembl_ID50 chembl_Ratio

chembl_Kd chembl_inhibition chembl_Efficacy
chembl_EC50 chembl_Activity

chembl_pEC50
chembl_pKi chembl_pKd
Dummy pIC50 pKi
pEC50 pKd
UMLS_PROPERTY
Caco2 permeability
Caco2 apparent permeability
Caco2 permeability rate

Model Applier

Provide the compound(s) to predict
Please provide compounds for which you want to predict the target property. Several options are available:

☒ Upload compounds from a file (SDF/MOL2/SMILES/Excel sheet) MoleculesToPredict.sdf

☐ Provide a Name/CAS-RN/SMILES

☐ Draw Molecule (click on depiction to the right to draw)

☐ Choose a previously prepared set: [...]

☐ Select molecules by a tag: [...]

Additional options
Prediction scenario: Use predictions only
☐ Disable prediction cache

Models applier browser

The complete list of models at OChem available for you is displayed below. If you are new here, you can also switch to a simplified OChem predictor

Select a model from the list

Model name or model ID: and property name: or by article id: Models visibility: Public and private Order by: creation time [refresh]

1 - 6 of 6

☐		Aqueous toxicity - a demo model	apply the model	predicts log(IGC50-1) using T Pyriformis tutorial dataset (training) (856) ASNN validated by T Pyriformis tutorial dataset (test) (437)	2014-02-13
--------------------------	--	---------------------------------	-----------------	---	------------

OChem predictor - results

Here you can browse the predictions for your compounds and export them in a variety of formats

☒ Export results in a file (Excel, CSV or SDF)

☒ Advanced applicability domain charts>>

Sorting: none

Accuracy estimates for the set
log(IGC50-1) for 192 compounds
RMSE = 0.54 ± 0.06
MAE = 0.42 ± 0.05

1 - 15 of 192 items on page 1 of 13

	log(IGC50-1) (Aqueous toxicity - a demo model) = 2.1 -log(mmol/L) ± 1.15 (ASNN-STDEV = 0.30, estimated RMSE = 0.59)
	log(IGC50-1) (Aqueous toxicity - a demo model) = 0.83 -log(mmol/L) ± 1.15 (ASNN-STDEV = 0.46, estimated RMSE = 0.59) OUT OF AD
	log(IGC50-1) (Aqueous toxicity - a demo model) = 1.9 -log(mmol/L) ± 1.15 (ASNN-STDEV = 0.58, estimated RMSE = 0.59) OUT OF AD
	log(IGC50-1) (Aqueous toxicity - a demo model) = 0.2 -log(mmol/L) ± 1.15 (ASNN-STDEV = 0.26, estimated RMSE = 0.59)



Modeling framework

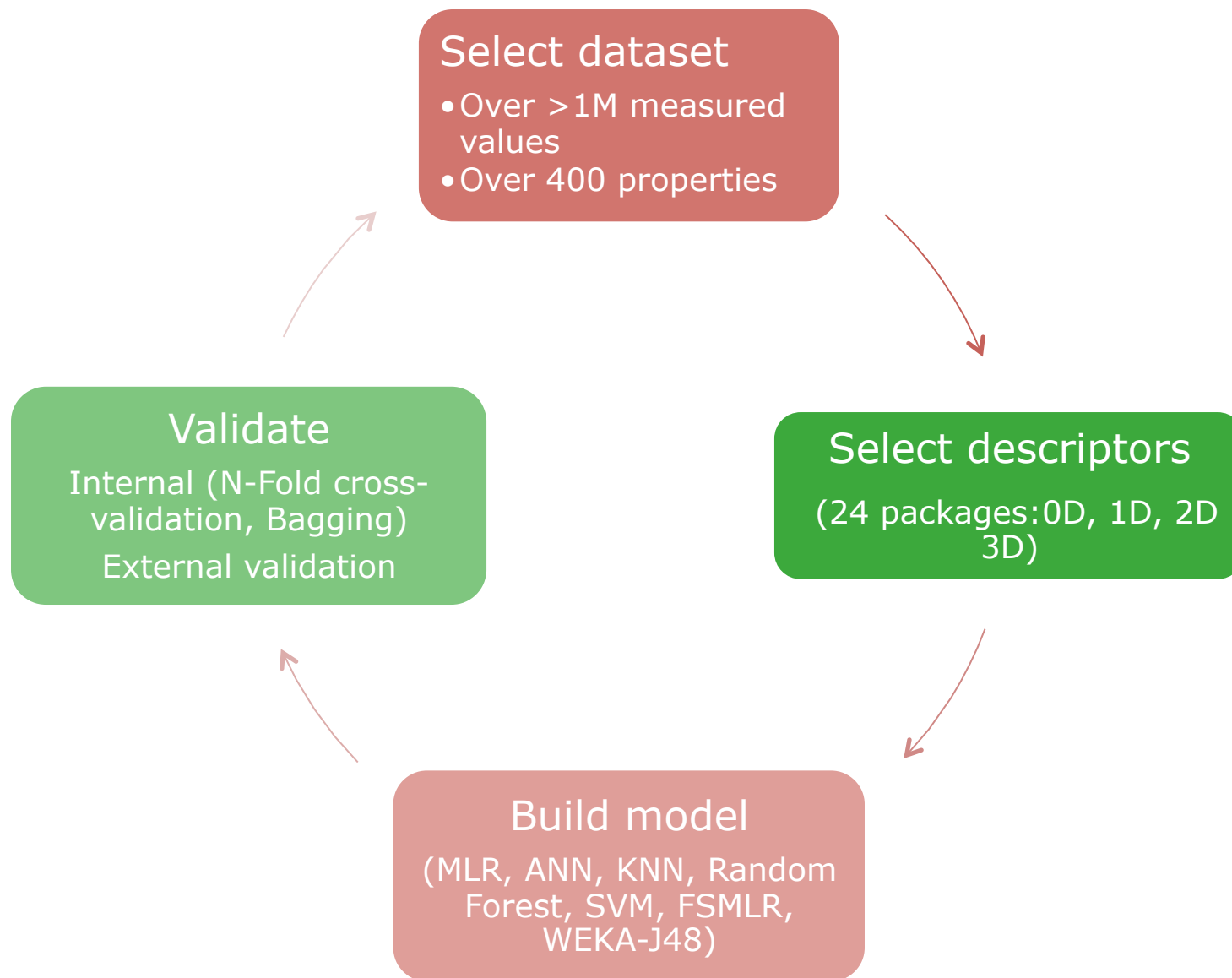
Creation of QSAR/QSPR models

- It is easy to create new models from uploaded data
- More than 10 Machine learning algorithms
 - Neural networks, KNN, SVM, MLR, PLS, random forests, J48
- More than 20 descriptor packages
 - 0D to 3D descriptors, academic and commercial
 - New descriptor packages can be easily integrated
 - Packages can be easily combined (mix & match)
 - Dragon, CDK, ADRIANA.CODE, Chemaxon, E-state, MERA, ...

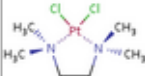
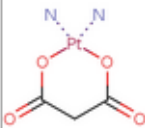
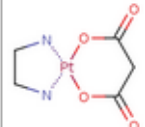
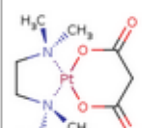
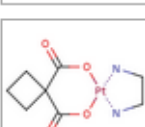
Models can be updated without complete rebuilt of model

Estimation of accuracy for each prediction

Modeling iterative workflow



LogP of Pt(II) + Pt(IV) complexes

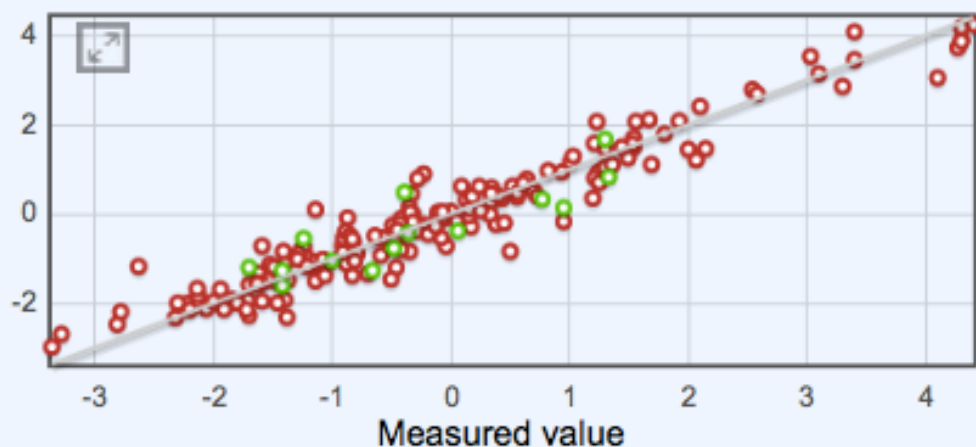
 molecule profile	● logPow = -0.85 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... P: 1204 T: 3 J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M4325 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13669 09:37, 19 Mar 09 / 12:22, 2t Koerner sa / tanzeem sa
 molecule profile	● logPow = -2.32 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M7024 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13671 09:37, 19 Mar 09 / 12:22, 2t Koerner sa / tanzeem sa
 molecule profile	● logPow = -2.19 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M87460 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13672 09:37, 19 Mar 09 / 12:22, 2t Koerner sa / tanzeem sa
 molecule profile	● logPow = -1.17 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... P: 1204 T: 3 J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M13735 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13673 09:37, 19 Mar 09 / 12:22, 2t Koerner sa / tanzeem sa
 molecule profile	● logPow = -1.7 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M3768443 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13676 09:37, 19 Mar 09 / 12:22, 2t Koerner sa / tanzeem sa

Tetko et al., J. Inorg. Biochem., 2016, 156(3), 1-13.

Model statistics

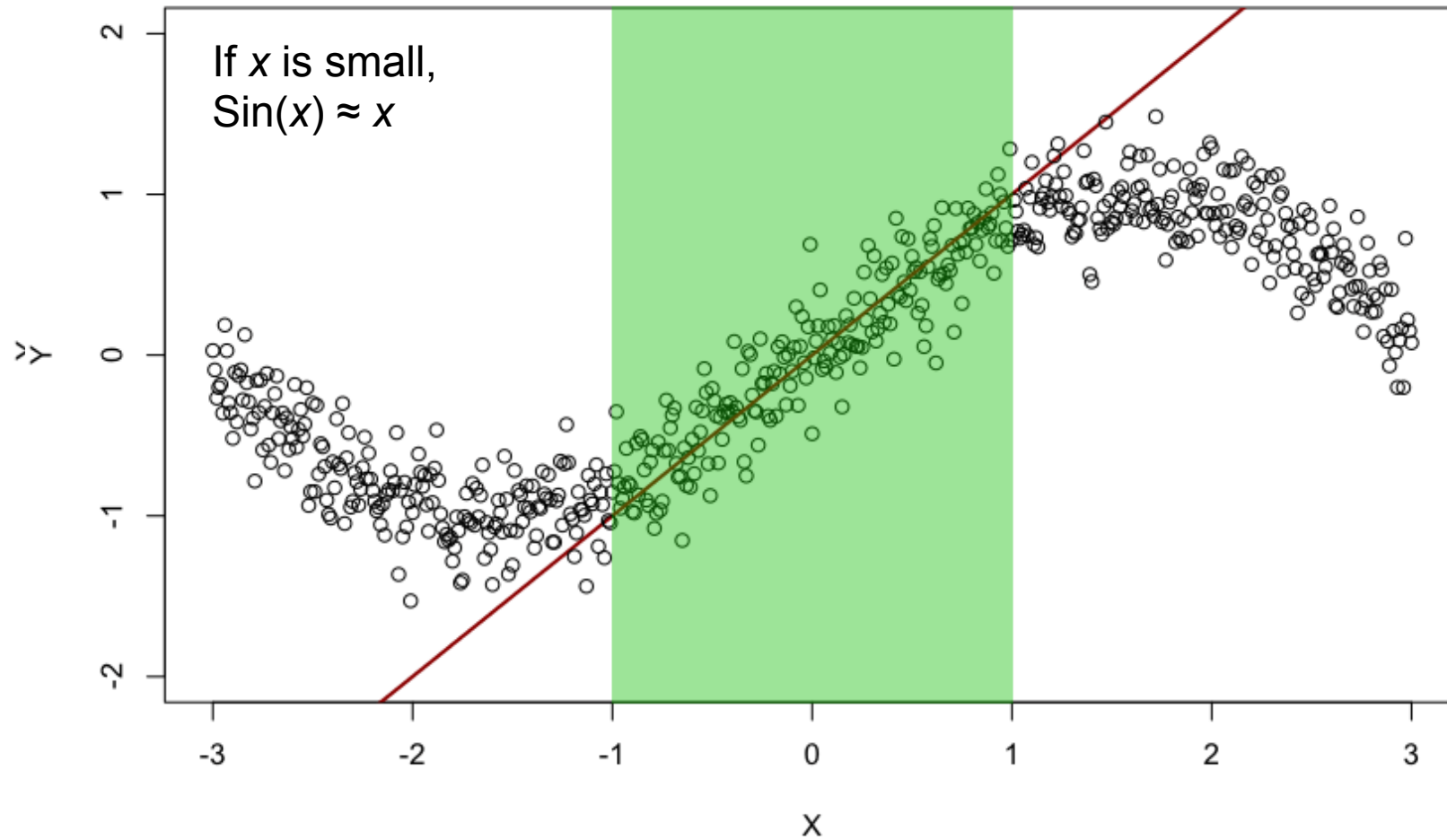
Predicted property: **logPow**
Training method: Consensus

Data Set	#	R ²	q ²	RMSE	MAE
Training set: LogPt final	187 records	0.93 ± 0.01	0.93 ± 0.01	0.41 ± 0.03	0.3 ± 0.02
Test set: Hristo	14 records	0.76 ± 0.1	0.76 ± 0.1	0.5 ± 0.06	0.43 ± 0.07

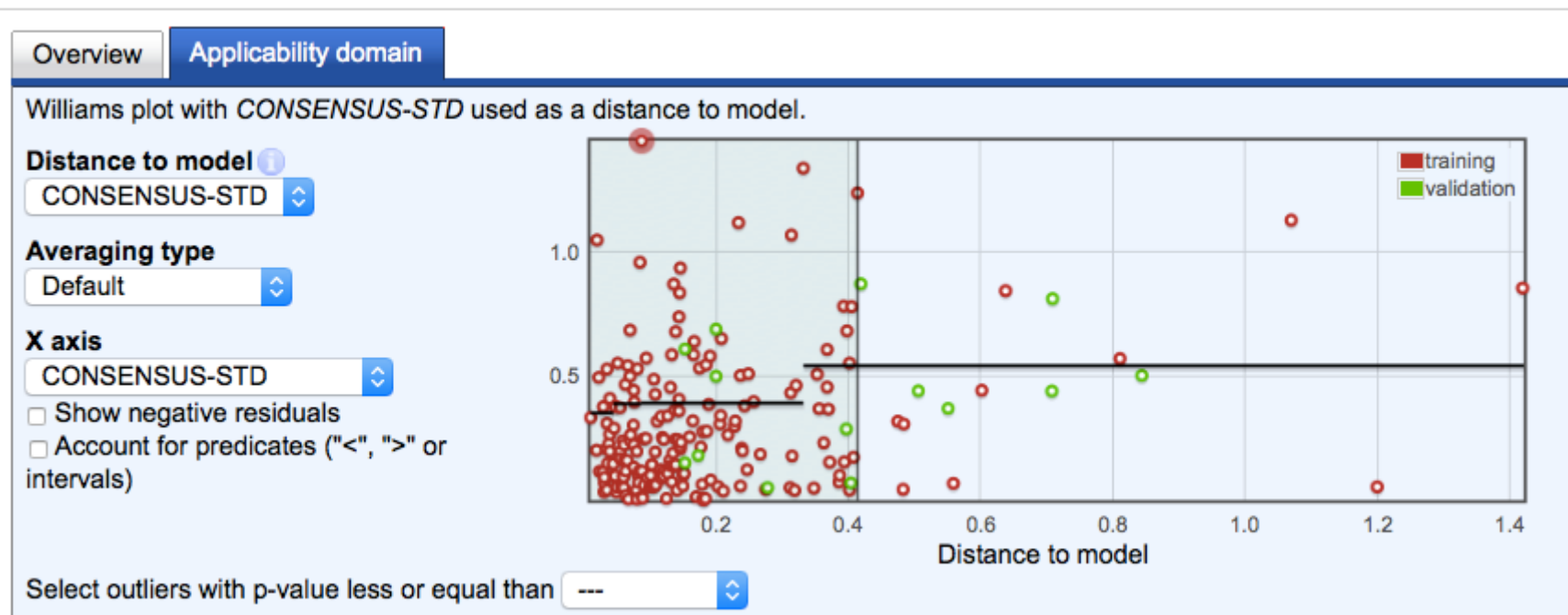


- Multiple statistical measures + confidence intervals
- Export for offline inspection
- Interactive scatter plot

Accuracy of prediction

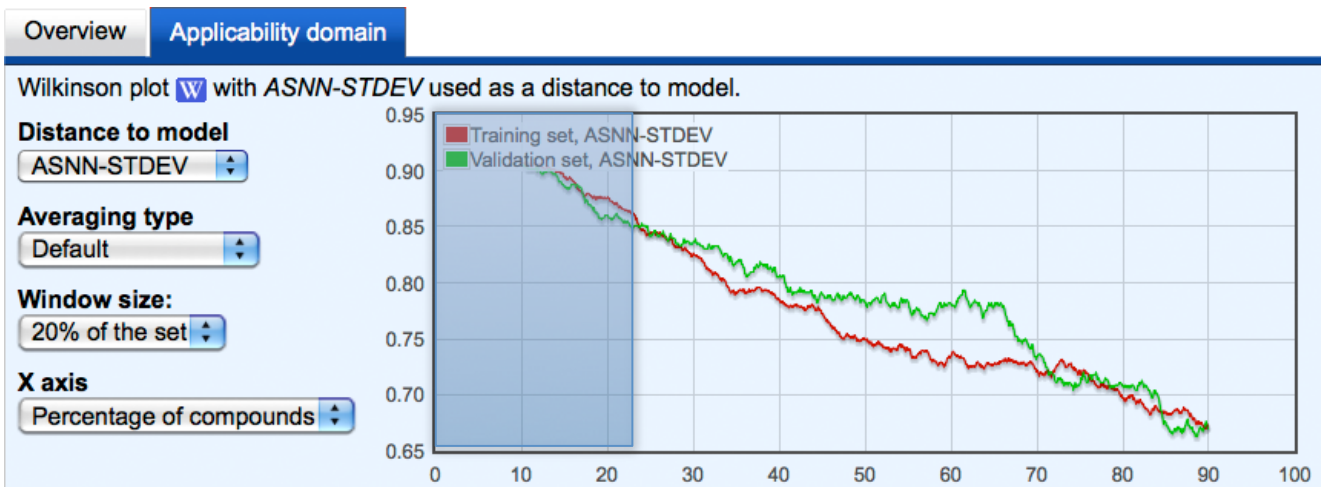
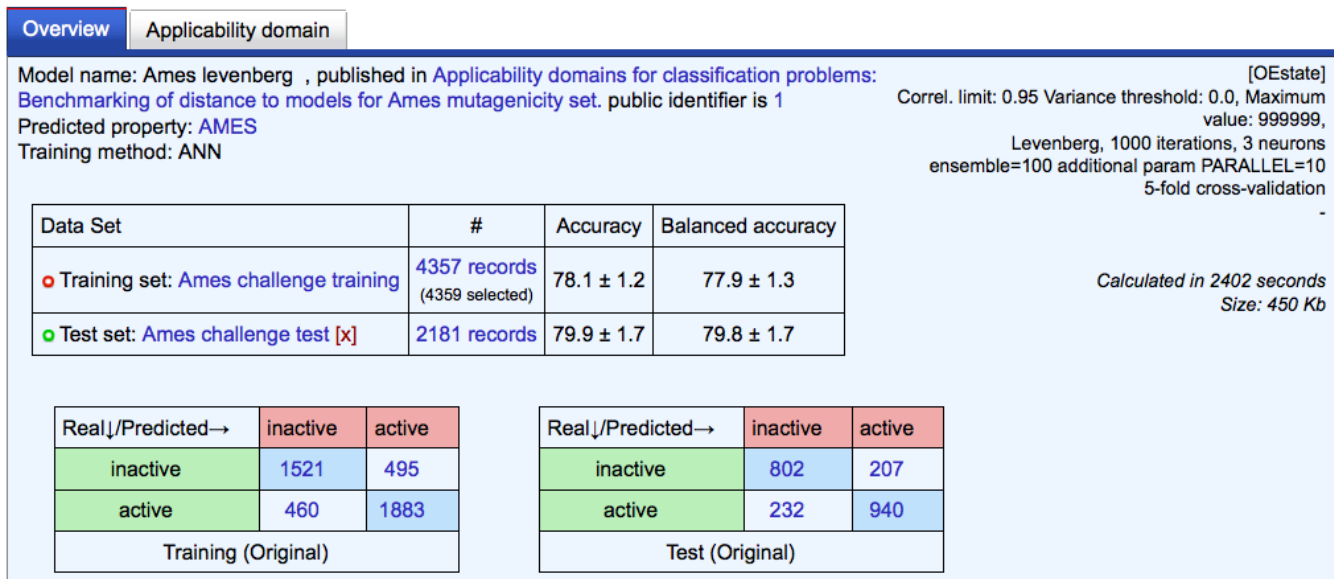


Outlying point on the Applicability Domain plot

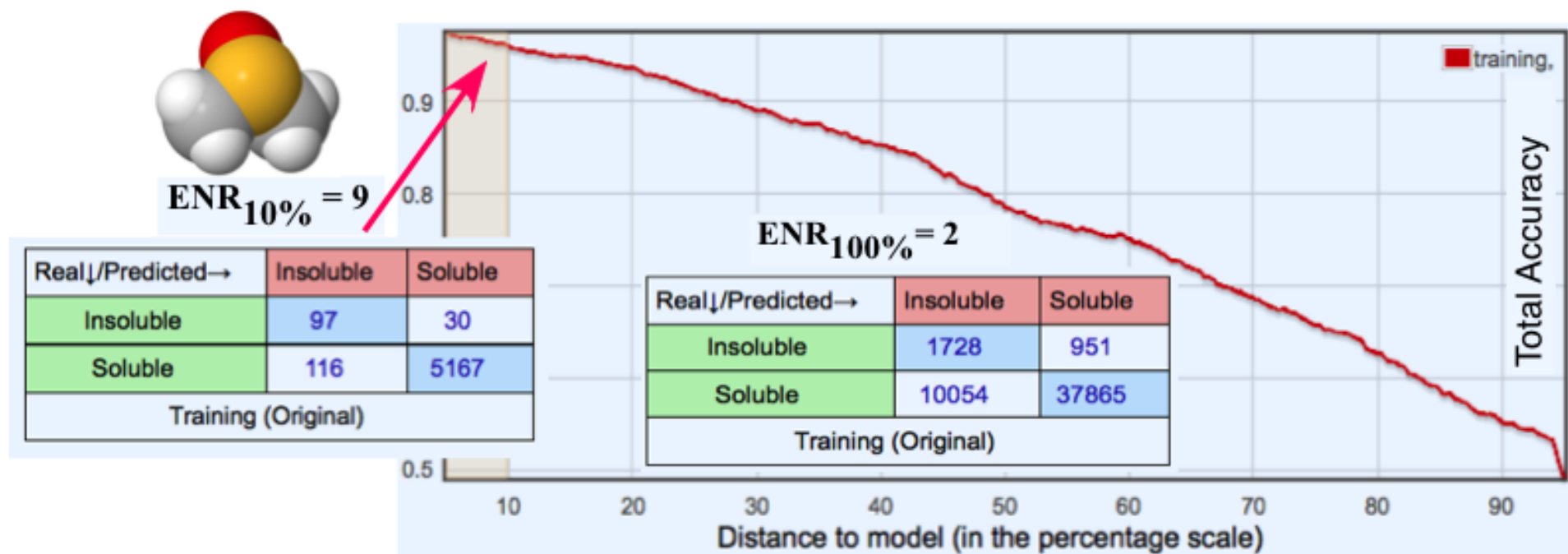


- Estimation of applicability domain of models
- Identification of outliers

Accuracy of predictions for classification model



Solubility in DMSO classification

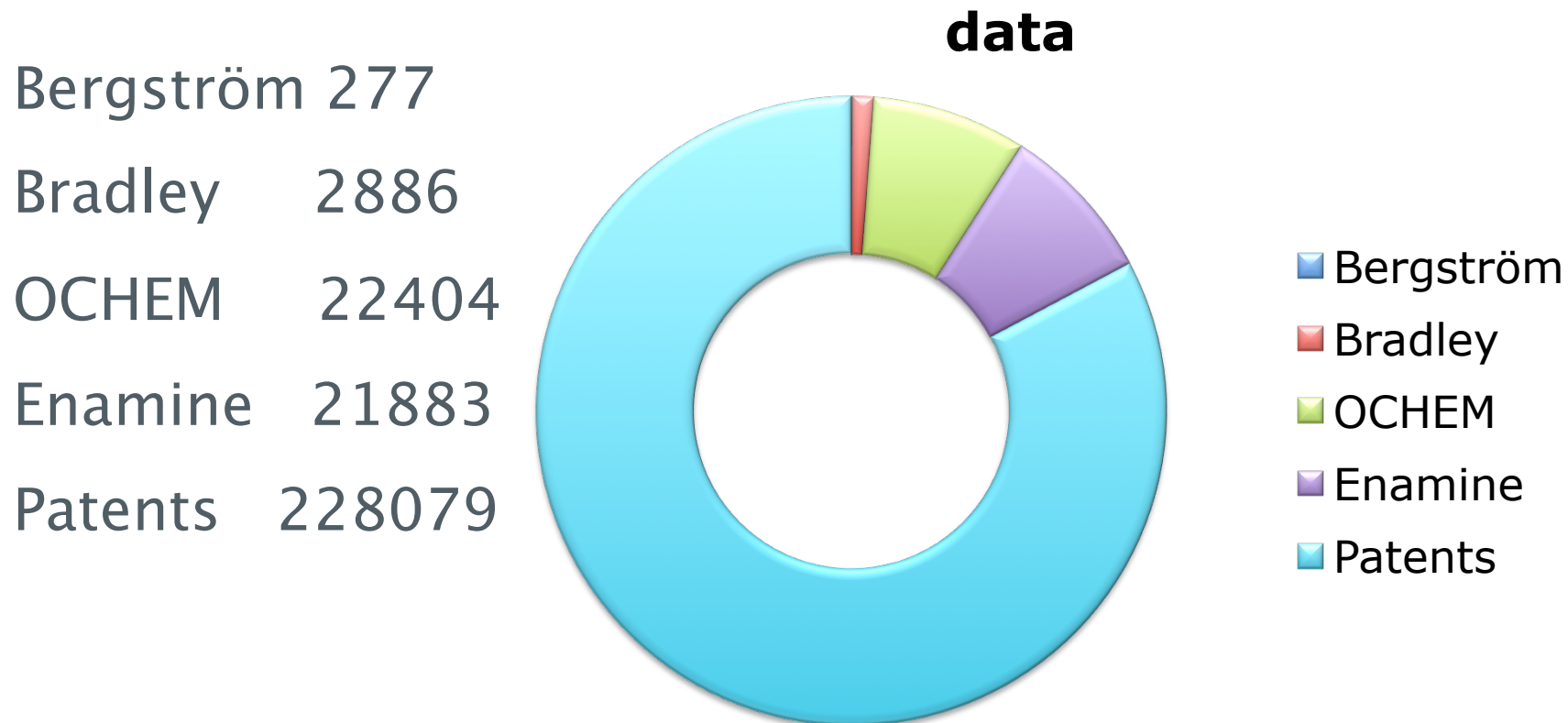


Example of a published solubility in DMSO model

- ✓ based on >163k compounds (provided by UCB and Enamine)
- ✓ nine folds decrease number of non-soluble in DMSO compounds

Tetko et al *J. Chem. Inf. Model.* 2013, 53(8):1990-2000.

300k Melting Point Datasets



Comprehensive modeling

Package name	Type of descriptors	Number of descriptors	Matrix size, billions	Non zero values, millions	Sparseness
Functional Groups	integer	595	0.18	3.1	33
QNPR	integer	1502	0.45	6.3	49
MolPrint	binary	688634	205	8.1	7200
Estate count	float	631	0.19	10	14
Inductive	float	54	0.02	11	1
ECFP4	binary	1024	0.31	12	25
Isida	integer	5886	1.75	18	37
ChemAxon	float	498	0.15	23	1.5
GSFrag	integer	1138	0.34	24	5.7
CDK	float	239	0.07	27	2
Adriana	float	200	0.06	32	1.3
Mera, Mersy	float	571	0.17	61	1.1
Dragon	float	1647	0.49	183	1.5

Comprehensive modeling

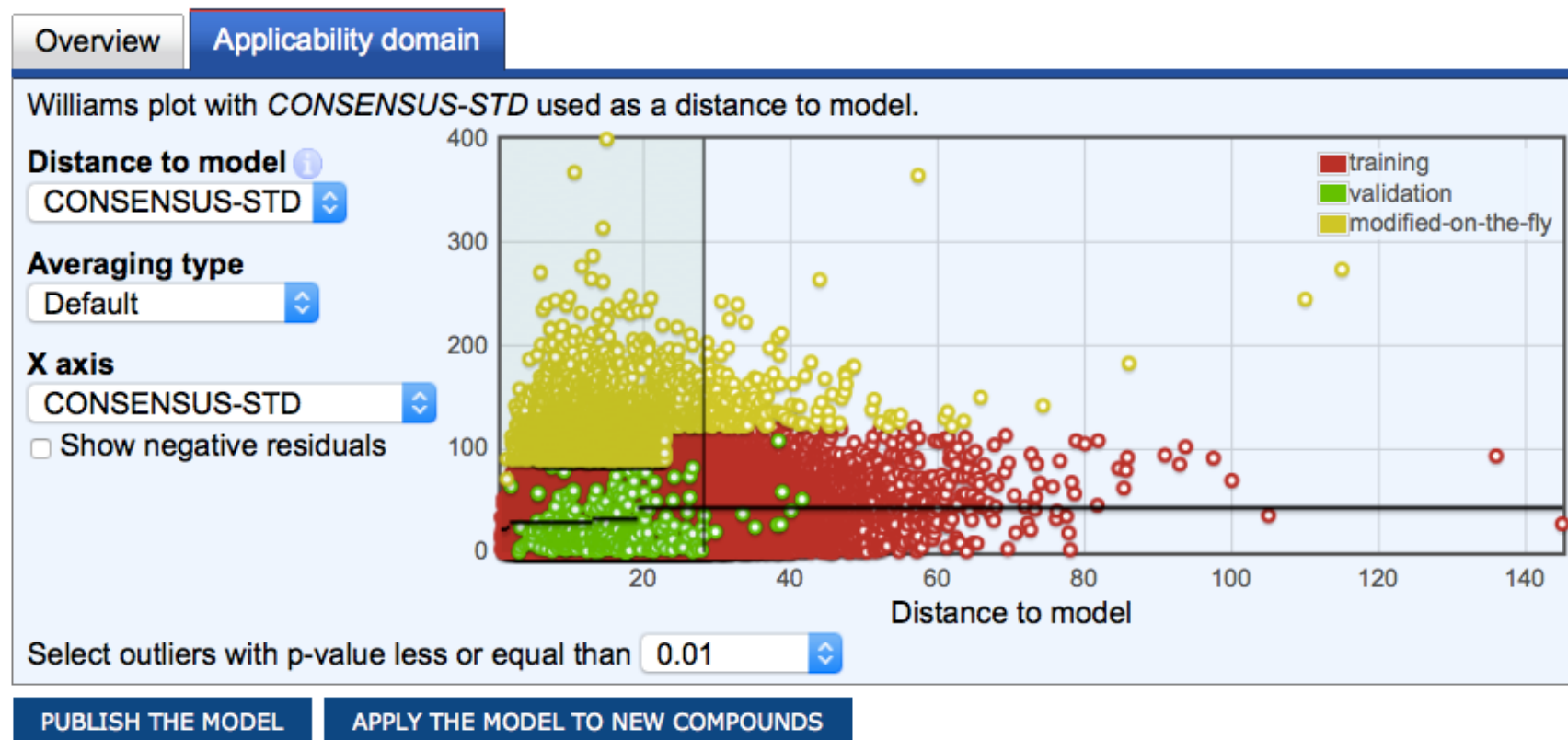
Predicted property: [Melting Point](#)

Training set: [patents - decomposition](#) (4 different versions detected) [i](#)

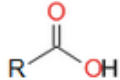
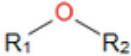
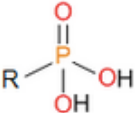
Metrics [RMSE - Root Mean Square Error](#) for [Training set](#) Validation: [All validation protocols](#)

	LibSVM
GSFrag	41.85
ChemaxonDescriptors (7.4)	39.63
InductiveDescriptors	49.54
MolPrint (Length 2)	61.2
StructuralAlerts	41.92
ECFP (4_1024)	61.46
OEstate	37.8
Fragmentor (Length 2 - 4)	38.54
QNPR (SMILES - length 1 - 3 threshold 5)	39.75
CDK (constitutional, topological, geometrical, electronic, hybrid)	38.45
Mera, Mersy	42.8
Dragon (blocks: 1-20)	43.04
Adriana	41.91
SIRMS (LABELING=ELM noH type=EXTENDED)	40.36
MolPrint (Length 3)	60.89
	Consensus
Misc.	36.01

Outliers identified with applicability domain (AD) plot



Functional group analysis of pyrolysis and MP data

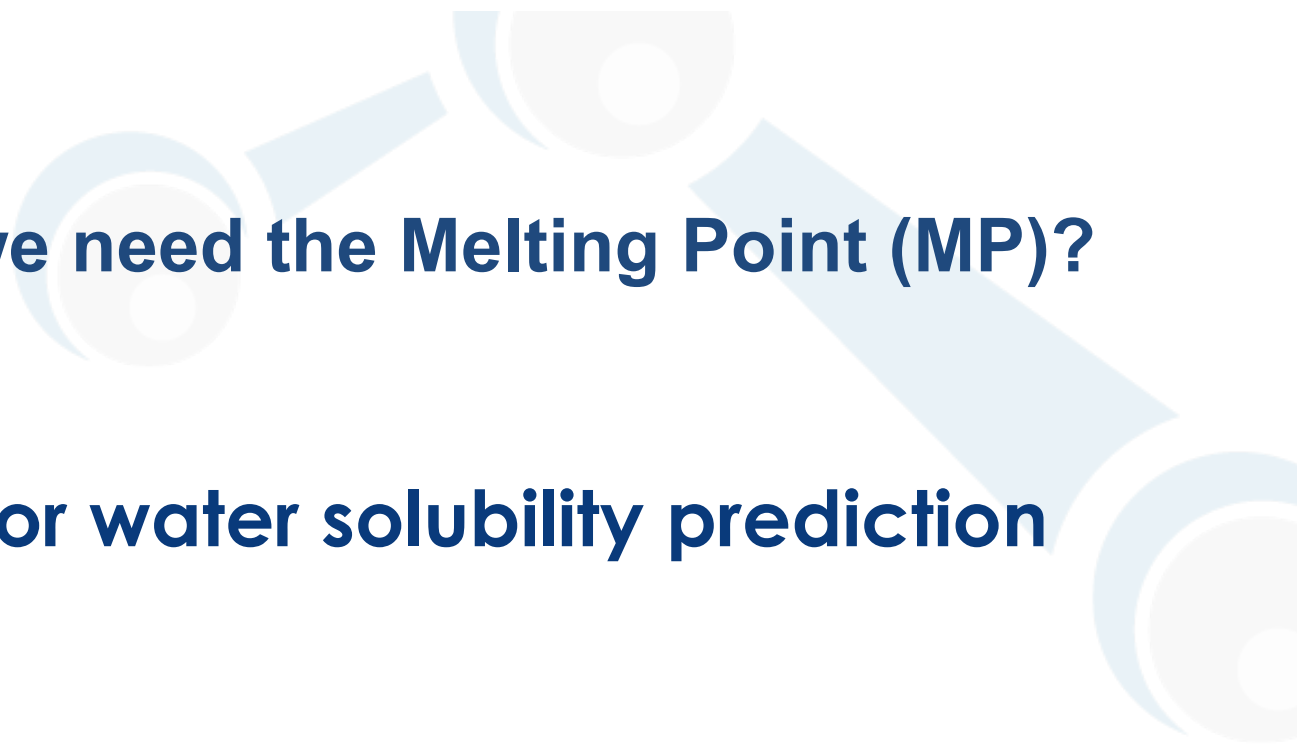
Descriptor	In set 1 (13785 molecules)	In set 2 (228174 molecules)	Enrichment factor	p-Value
	3232 (23.4%)	26196 (11.5%)	2.0	1.33E-315
Pnictogens Group 15: the nitrogen family 	13235 (96.0%)	202449 (88.7%)	1.1	1.42E-198
	3257 (23.6%)	79741 (34.9%)	1.5	-1.25E-172
	280 (2.0%)	352 (0.2%)	13.2	4.21E-172

Outliers have higher percentage of predicted decomposing compounds

20% of ~5k outliers are predicted as pyrolysis

vs.

17% of ~223k PATENTS compounds



Why do we need the Melting Point (MP)?

Use of MP for water solubility prediction

Yalkowsky equation:

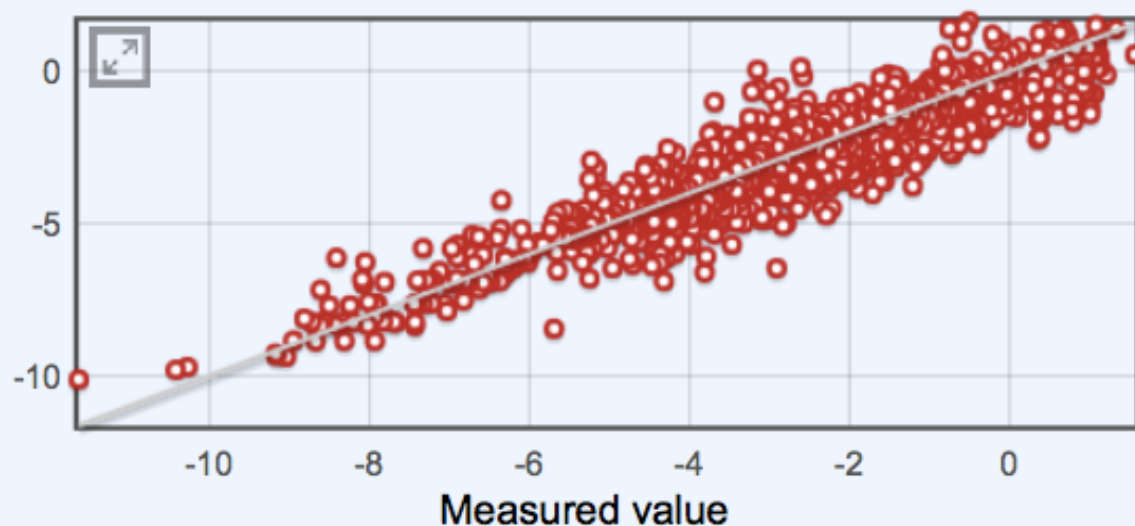
$$\log S = 0.5 - 0.01 (\text{MP} - 25) - \log K_{ow}$$

Prediction of Huuskonen set using ALOGPS logP and MP based on 230k measurements

Predicted property: **Aqueous Solubility** modeled in log(mol/L)

Training method: MLRA

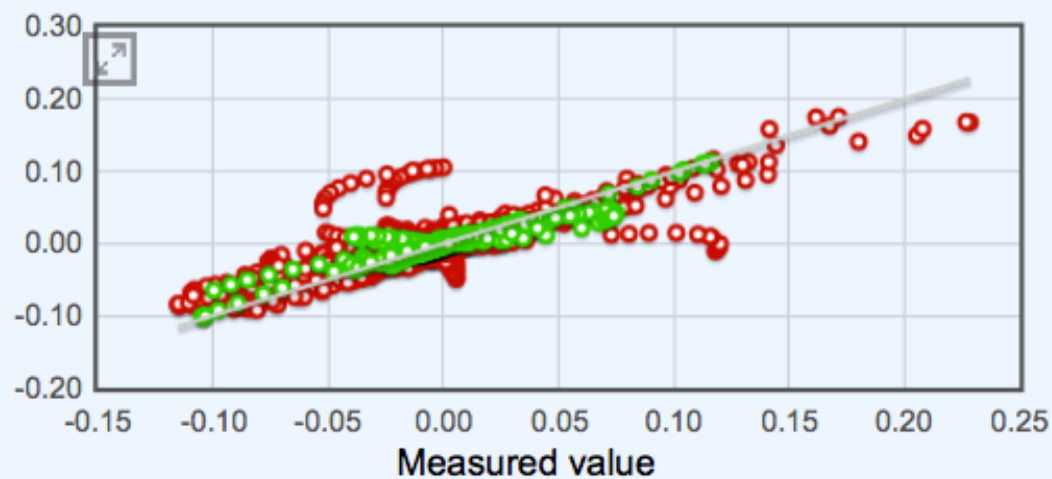
Data Set	#	R2	q2	RMSE	MAE
Training set: logS set	1311 records	0.842 ± 0.009	0.83 ± 0.01	0.84 ± 0.02	0.64 ± 0.02



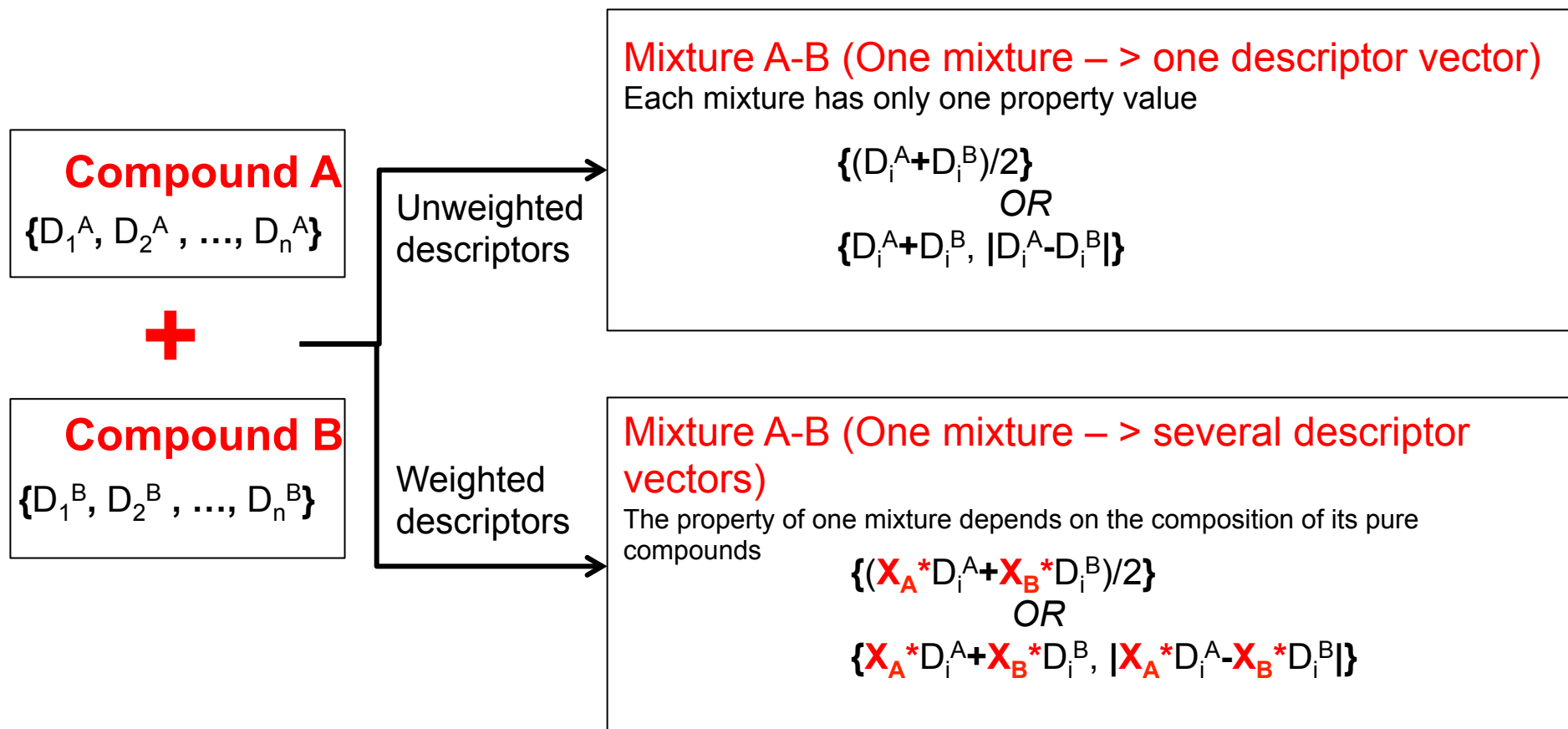
Best models based on this set have RMSE ~ 0.6

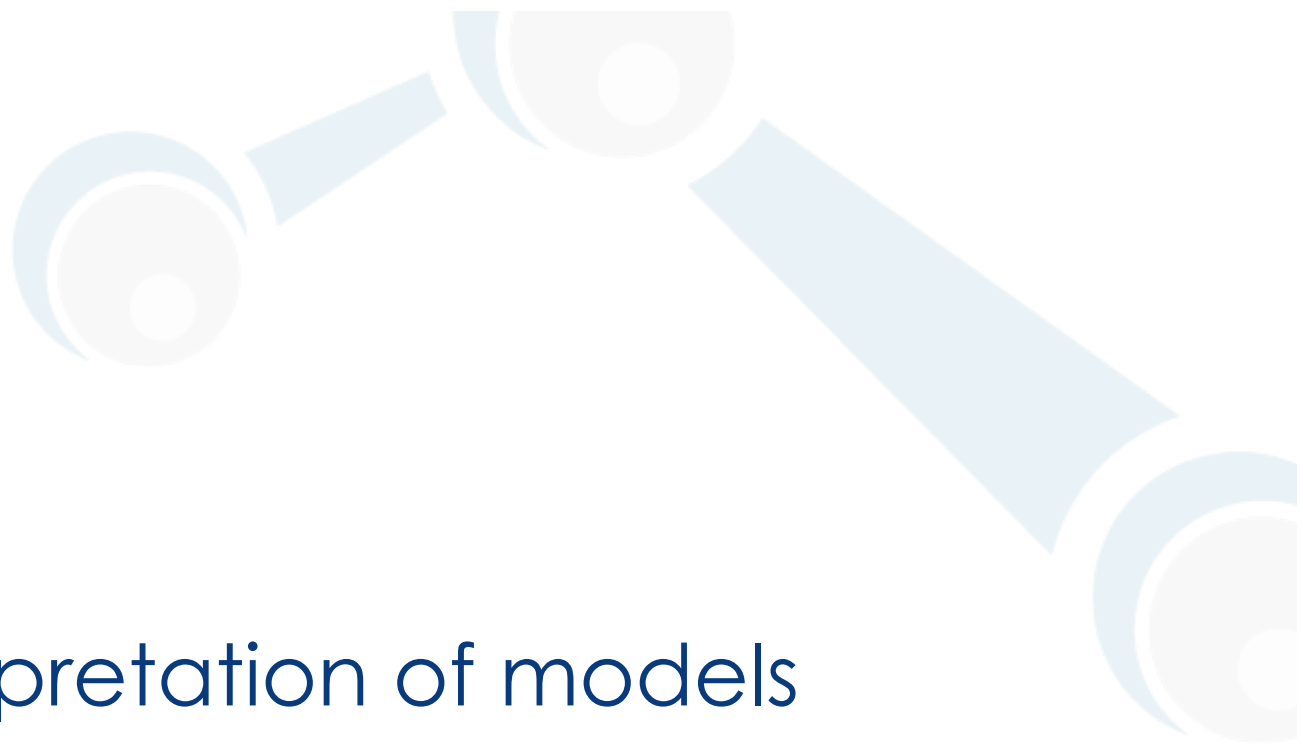
Modeling of mixtures

Data Set	#	R ²	q ²	RMSE	MAE
● Training set: Ajmani_delta_train	3857 records	0.69 ± 0.06	0.69 ± 0.06	0.014 ± 0.001	0.0064 ± 0.0004
● Test set: Ajmani_delta_test [x]	672 records	0.88 ± 0.04	0.87 ± 0.04	0.0089 ± 0.001	0.0041 ± 0.0006



Mixtures' descriptors



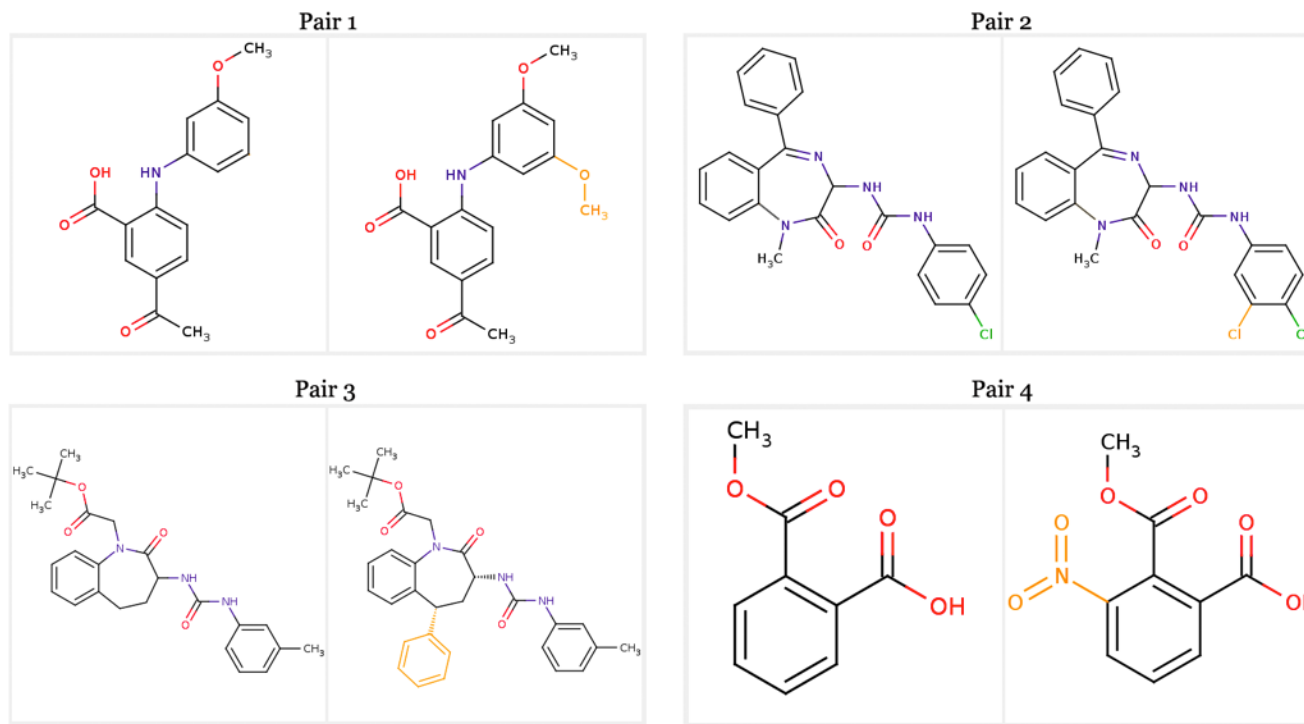


Interpretation of models

Molecular Matched Pairs

A **molecular matched pair** (MMP) is a pair of molecules that have only a (minor) single-point difference.

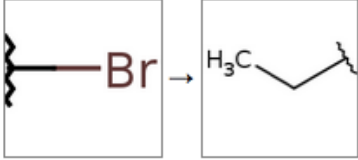
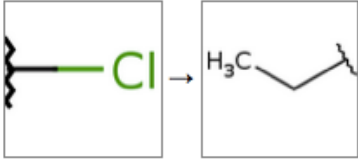
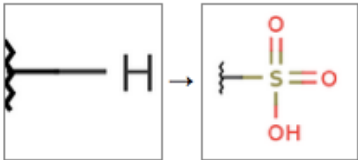
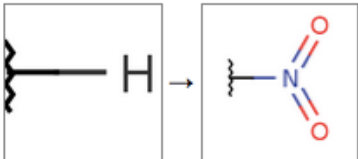
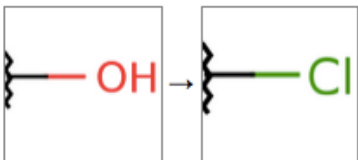
The typical way is to define a minor difference as a changed molecular fragment with less than 10 atoms.



MMPs for classification data (AMES mutagenicity)

Transformations: minimum pairs, p-value

1 - 5 of 32 items on page of 7 > >>

	12 matched pairs ↑ 0 pairs, ↓ 7 pairs 2 negative and 3 positive molecules unaffected SMIRKS: <chem>*Br -> CC*</chem>
	21 matched pairs ↑ 0 pairs, ↓ 9 pairs 8 negative and 4 positive molecules unaffected SMIRKS: <chem>*Cl -> CC*</chem>
	17 matched pairs ↑ 1 pairs, ↓ 8 pairs 5 negative and 3 positive molecules unaffected SMIRKS: <chem>*[H] -> OS(*) (=O) =O</chem>
	204 matched pairs ↑ 57 pairs, ↓ 8 pairs 26 negative and 113 positive molecules unaffected SMIRKS: <chem>*[H] -> *N (=O) =O</chem>
	69 matched pairs ↑ 21 pairs, ↓ 4 pairs 25 negative and 19 positive molecules unaffected SMIRKS: <chem>O* -> *Cl</chem>

1 - 5 of 32

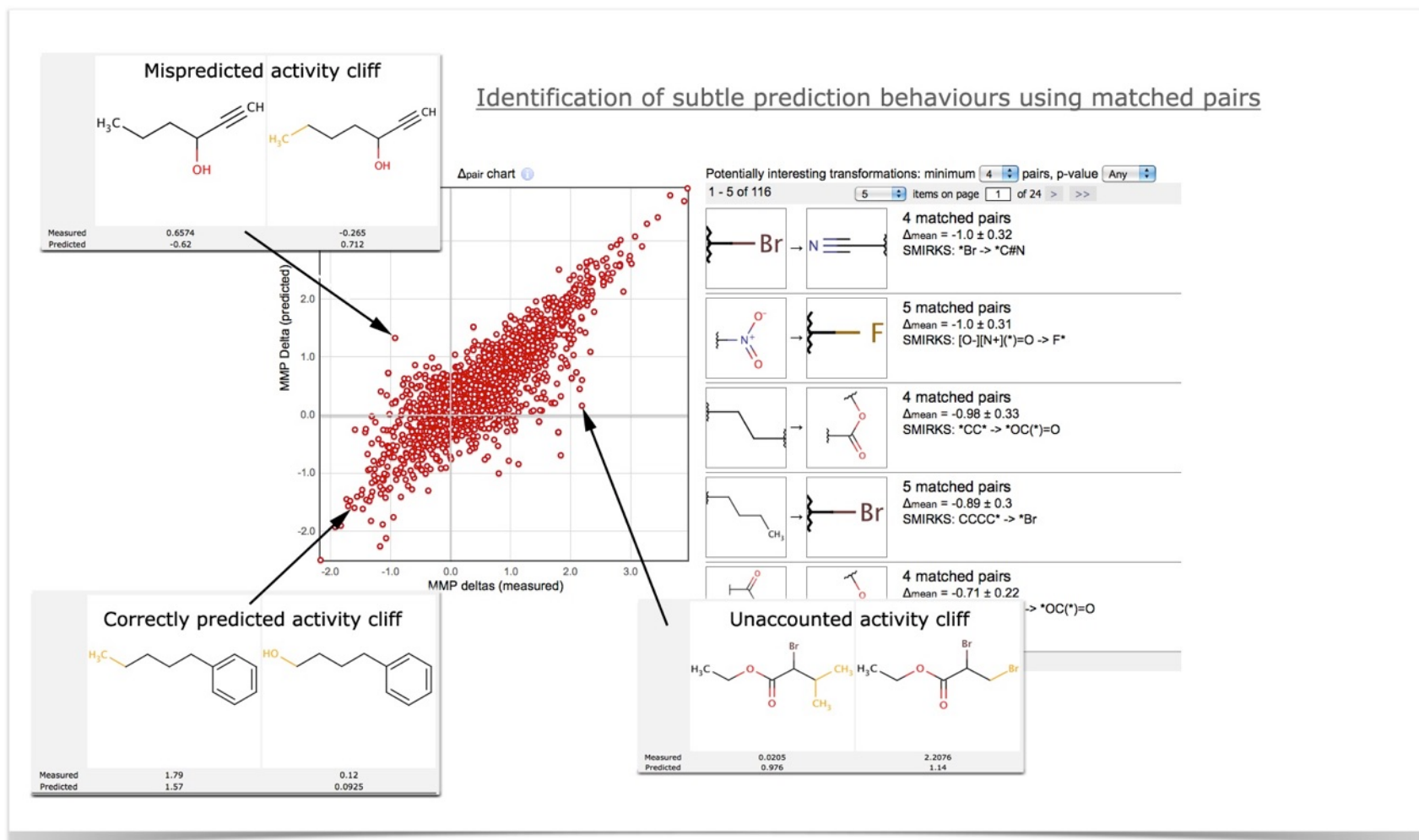
Resubmit all molecules

Show fragment graph

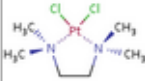
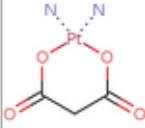
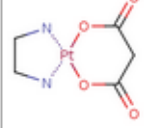
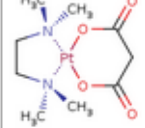
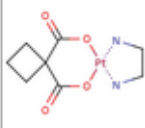
Save transformations

Export pairs

Analysis of rules that were learnt by models



LogP of Pt(II) + Pt(IV) complexes

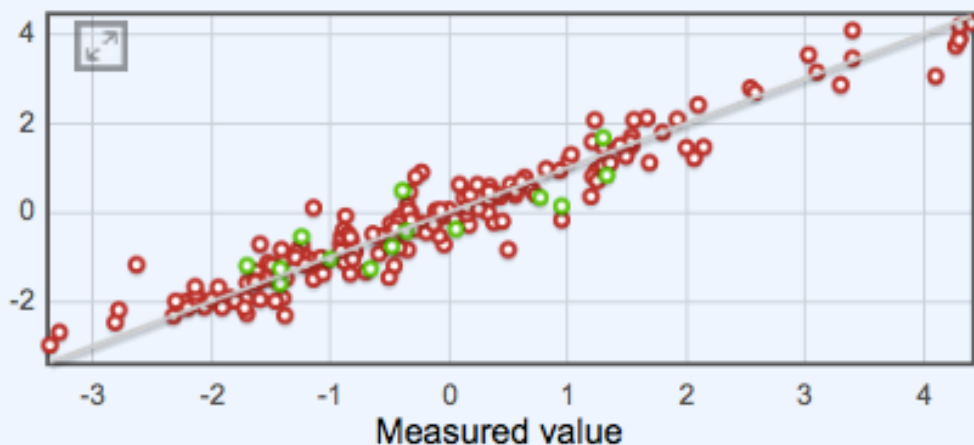
 molecule profile	● logPow = -0.85 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... P: 1204 T: 3 J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M4325 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13669 09:37, 19 Mar 09 / 12:22, 20 Koerner sa / tanzeem sa
 molecule profile	● logPow = -2.32 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M7024 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13671 09:37, 19 Mar 09 / 12:22, 20 Koerner sa / tanzeem sa
 molecule profile	● logPow = -2.19 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M87460 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13672 09:37, 19 Mar 09 / 12:22, 20 Koerner sa / tanzeem sa
 molecule profile	● logPow = -1.17 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... P: 1204 T: 3 J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M13735 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13673 09:37, 19 Mar 09 / 12:22, 20 Koerner sa / tanzeem sa
 molecule profile	● logPow = -1.7 (in Log unit) Platts, JA et al The RP-HPLC measurement and QSPR analysis of logP(o/w) value... J. Inorg. Biochem. 2006; 100 (7) 1199-207 MoleculeID: M3768443 Public record	logPow Method = RP-HPLC logPow Buffer = methanol RecordID: R13676 09:37, 19 Mar 09 / 12:22, 20 Koerner sa / tanzeem sa

Model for LogP for Pt complexes

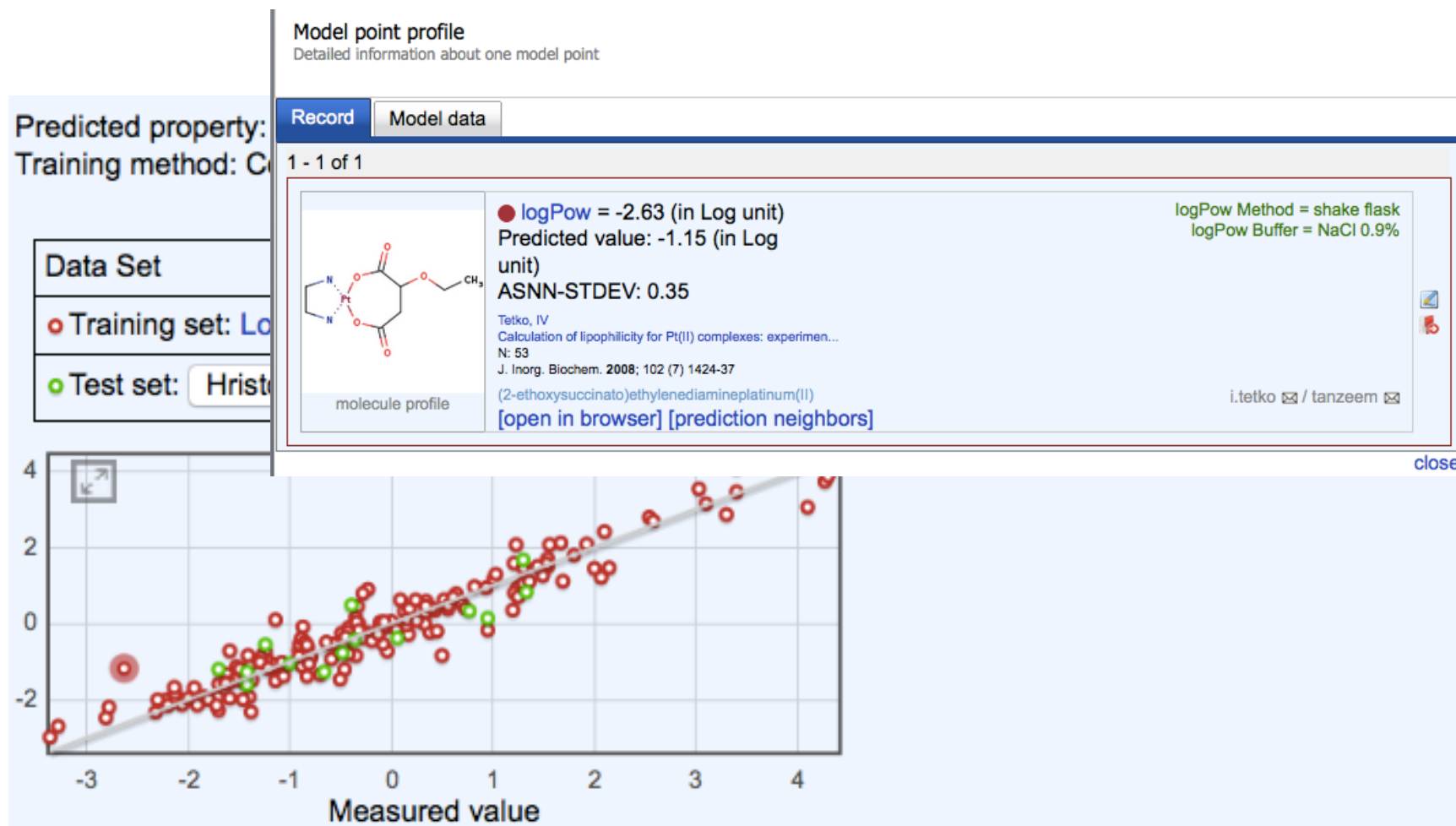
Predicted property: **logPow**

Training method: Consensus

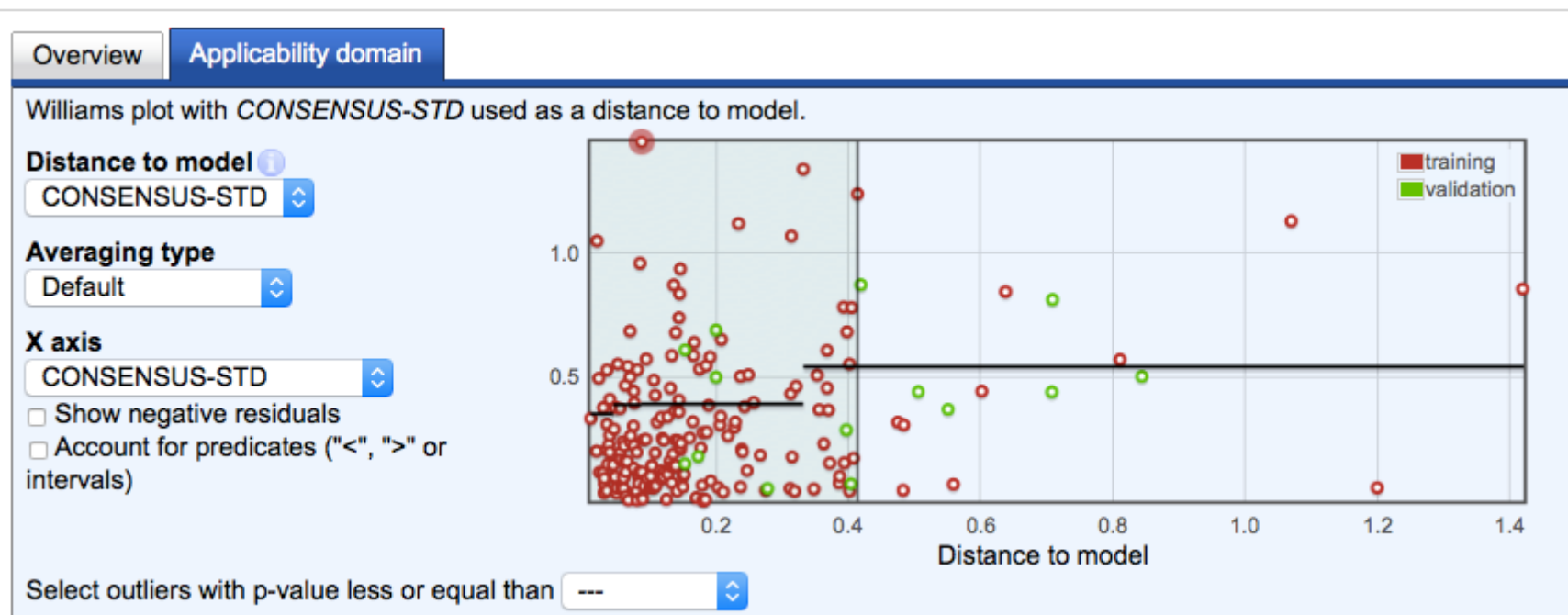
Data Set	#	R ²	q ²	RMSE	MAE
Training set: LogPt final	187 records	0.93 ± 0.01	0.93 ± 0.01	0.41 ± 0.03	0.3 ± 0.02
Test set: Hristo	[x] 14 records	0.76 ± 0.1	0.76 ± 0.1	0.5 ± 0.06	0.43 ± 0.07



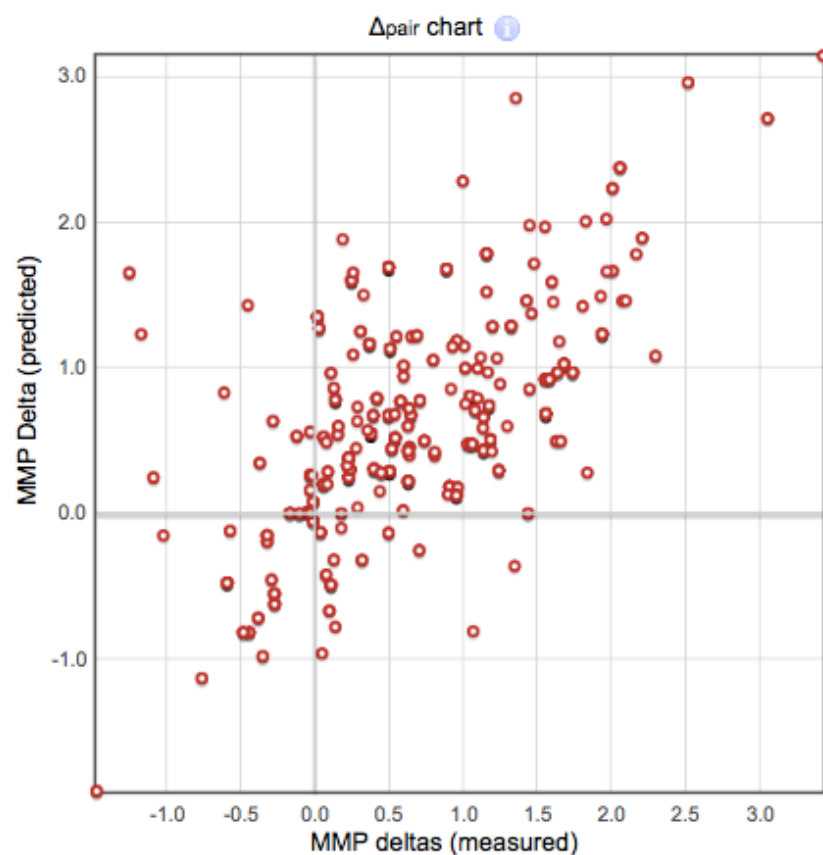
Model for LogP for Pt complexes



Outlying point on the Applicability Domain plot



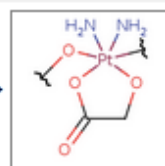
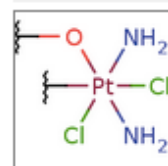
Prediction of logP for Pt complexes



Transformations: minimum 4 pairs, p-value Any

1 - 5 of 60

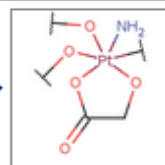
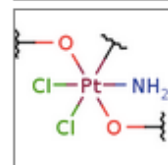
5 items on page 1 of 12 > >>



6 matched pairs

$\Delta_{\text{mean}} = -0.27 \pm 0.19$

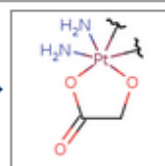
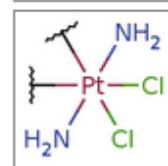
SMIRKS: N[Pt](N)(*)(Cl)(Cl)O* -> N[Pt]1(N)(*)(O*)OCC(=O)O1



6 matched pairs

$\Delta_{\text{mean}} = -0.27 \pm 0.19$

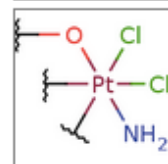
SMIRKS: N[Pt](*)(Cl)(Cl)(O*)O* -> N[Pt]1(*) (O*)(O*)OCC(=O)O1



7 matched pairs

$\Delta_{\text{mean}} = -0.24 \pm 0.2$

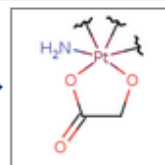
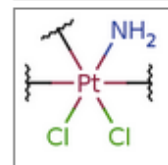
SMIRKS: N[Pt](N)(*)(*)(Cl)Cl -> N[Pt]1(N)(*)(*)OCC(=O)O1



7 matched pairs

$\Delta_{\text{mean}} = -0.24 \pm 0.2$

SMIRKS: N[Pt](*)(*)(Cl)(Cl)O* -> N[Pt]1(*) (O*)(O*)OCC(=O)O1



8 matched pairs

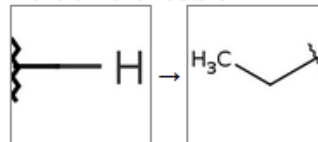
$\Delta_{\text{mean}} = -0.17 \pm 0.26$

SMIRKS: N[Pt](*)(*)(*)(Cl)Cl -> N[Pt]1(*) (*)(*)OCC(=O)O1

Outliers of the model (functional groups descriptors)

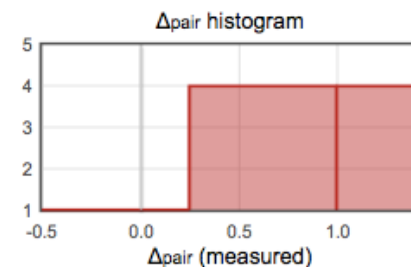


Transformation details:



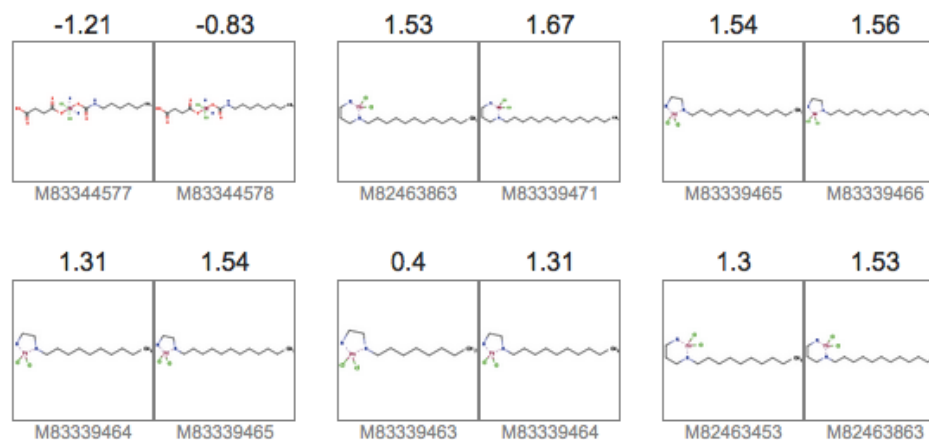
Transformation ID: **TR181**
SMIRKS: *[H] -> CC*

[Back to all transformations](#)

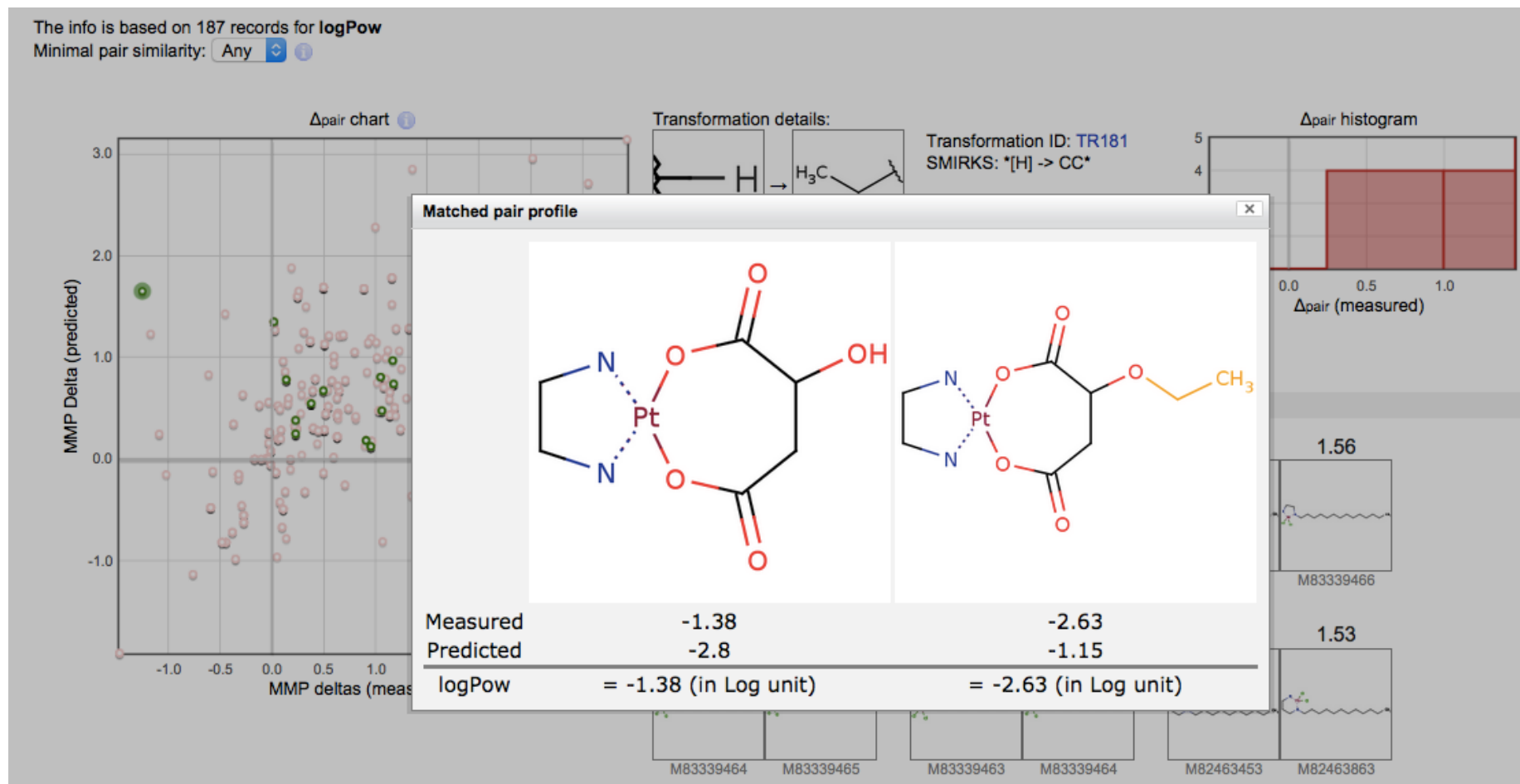


Pairs having the same transformation as the selected one:

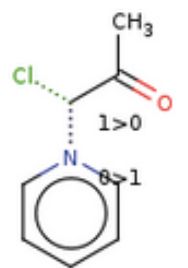
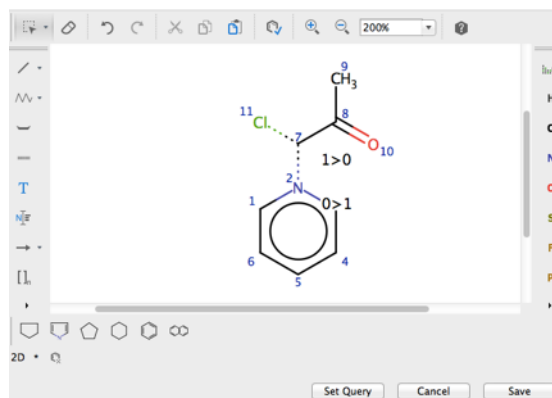
1 - 14 of 14



Outliers of the model (Functional groups descriptors)



Storage of Chemical Reaction as Condensed Graph of Reaction (CRG)



molecule profile

● SN2 reaction

rate constant =

-4.72 (in log(L*mol⁻¹*s⁻¹)))

Nugmanov, R. et al

Development of "structure-property" models in nucleophilic s...

N: 8 P: 270

Journal of Structural Chemistry **2014**; 55 (6) 1026-1032

MoleculeID: M84326300

Private record

Temperature= 55.6 °C

Solvent = Ethanol

RecordID: R19277000

23:25, 4 Jul 15 / 20:32, 6 Jul 15

itetko

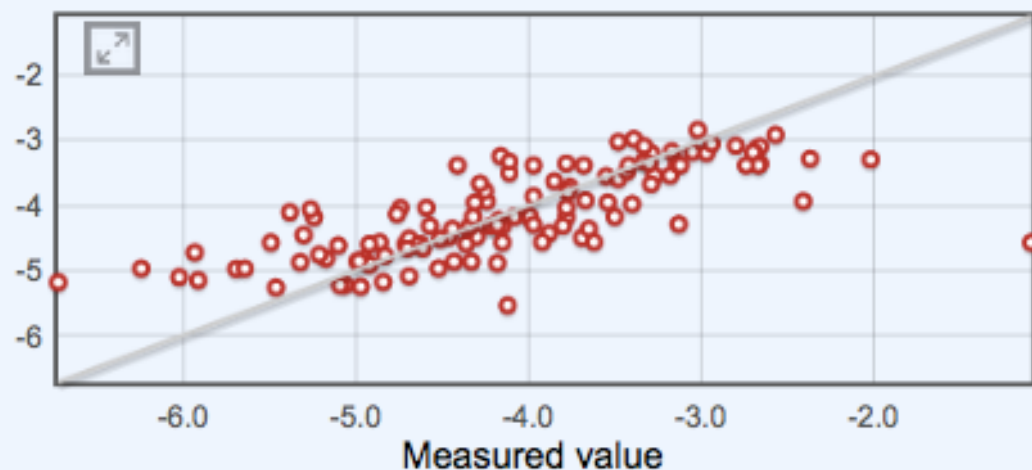
Only visible to itetko

Modeling of SN2 reactions using OCHEM

Predicted property: **SN2 reaction rate constant** modeled in $\log(L \cdot \text{mol}^{-1} \cdot \text{s}^{-1})$

Training method: Consensus

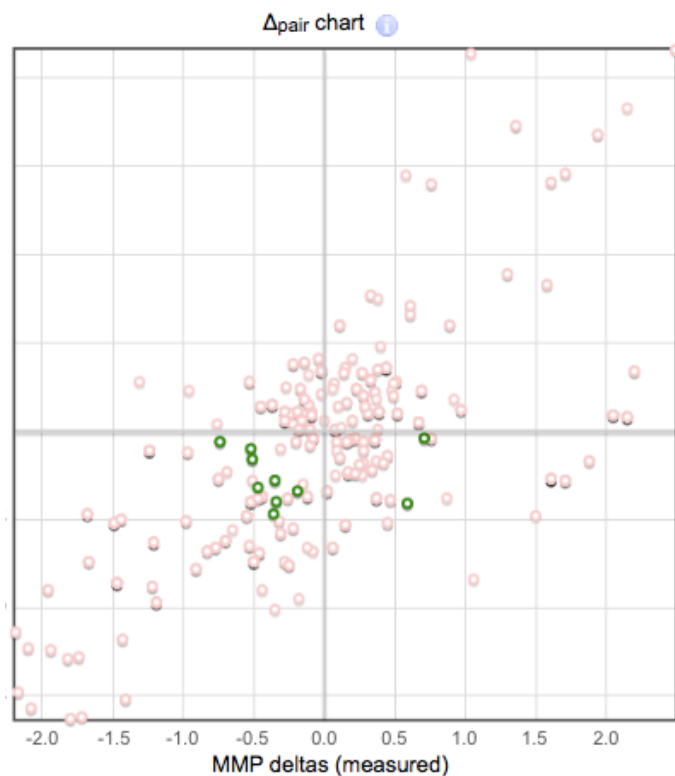
Data Set	#	R2	q2	RMSE	MAE
Training set: SN2 methanol 20-40	125 records	0.54 ± 0.08	0.54 ± 0.08	0.64 ± 0.08	0.44 ± 0.04



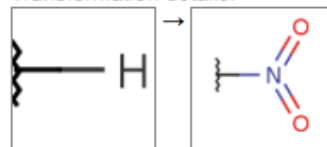
Consistent data: $t = 20-40$, solvent = methanol

based on 125 records for **SN2 reaction rate constant** (excluding 20 not indexed molecules)

pair similarity: ⓘ



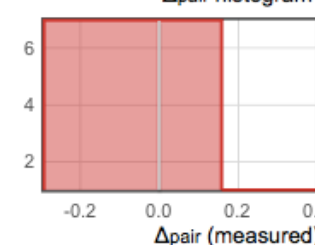
Transformation details:



Transformation ID: **TR291**
SMIRKS: *[H] -> *N(=O)=O

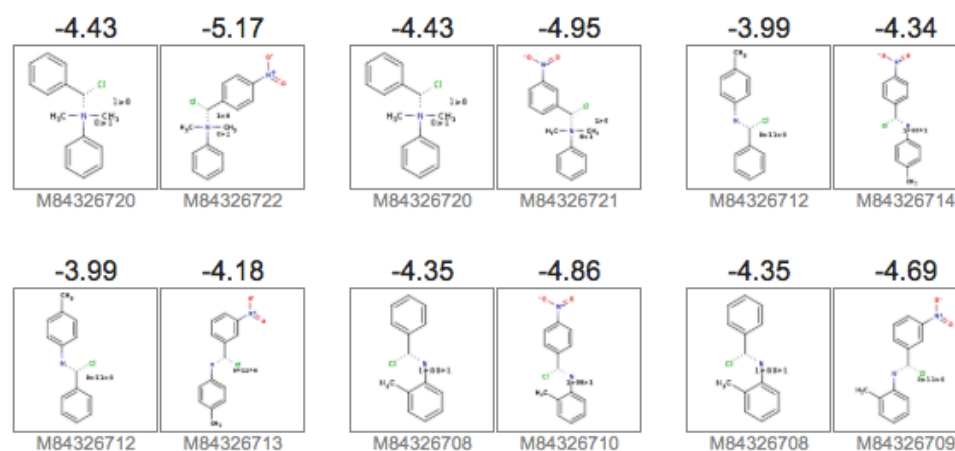
[Back to all transformations](#)

Δ_{pair} histogram

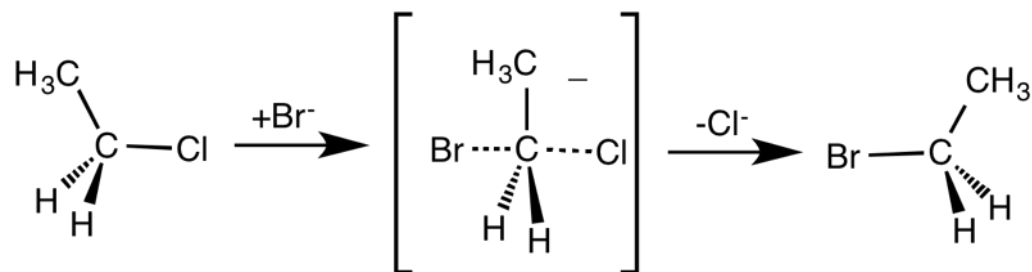


Pairs having the same transformation as the selected one:

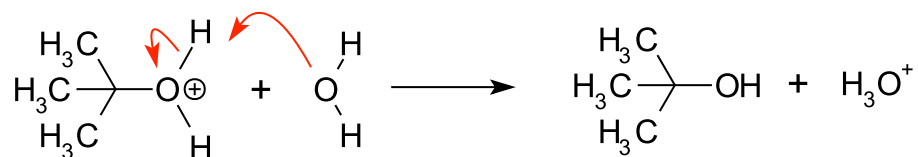
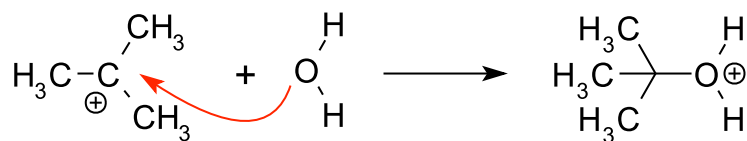
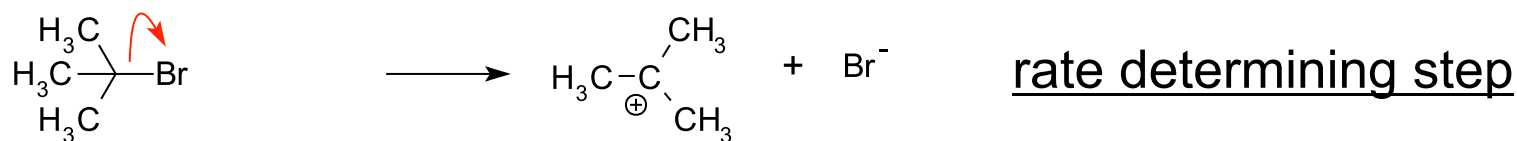
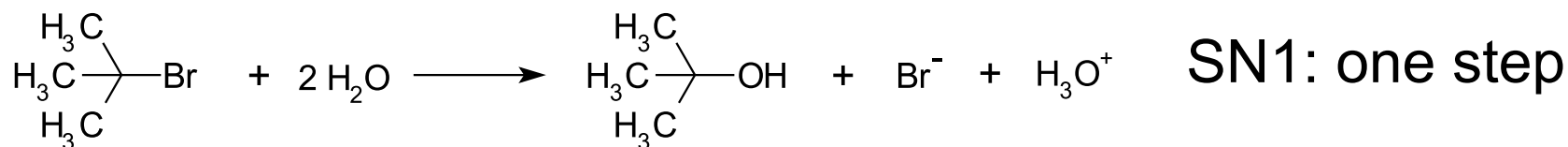
1 - 10 of 10



SN2 vs. SN1 reactions

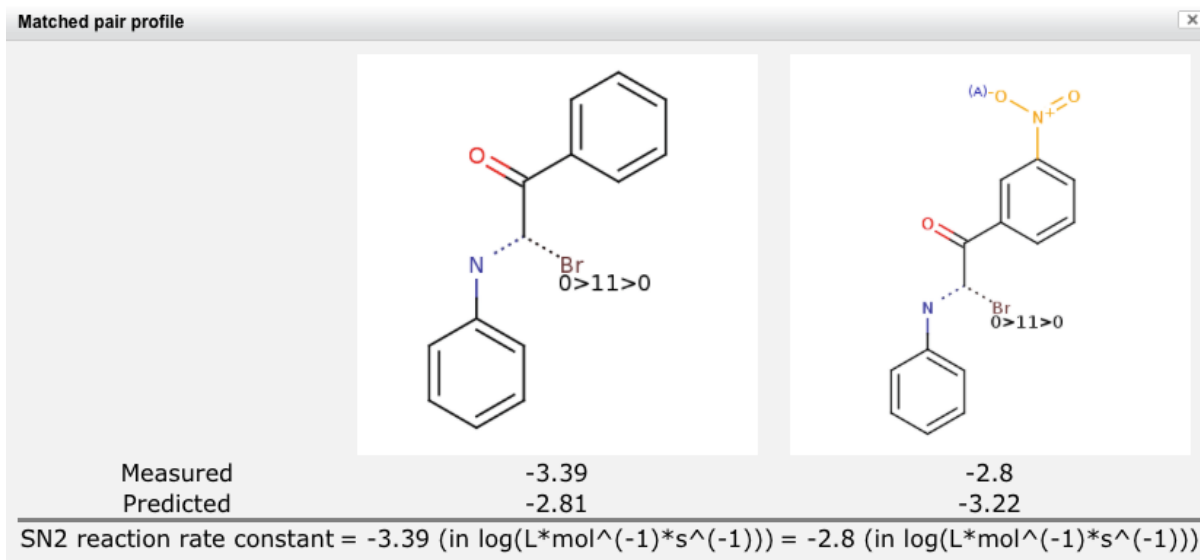


SN2: one step

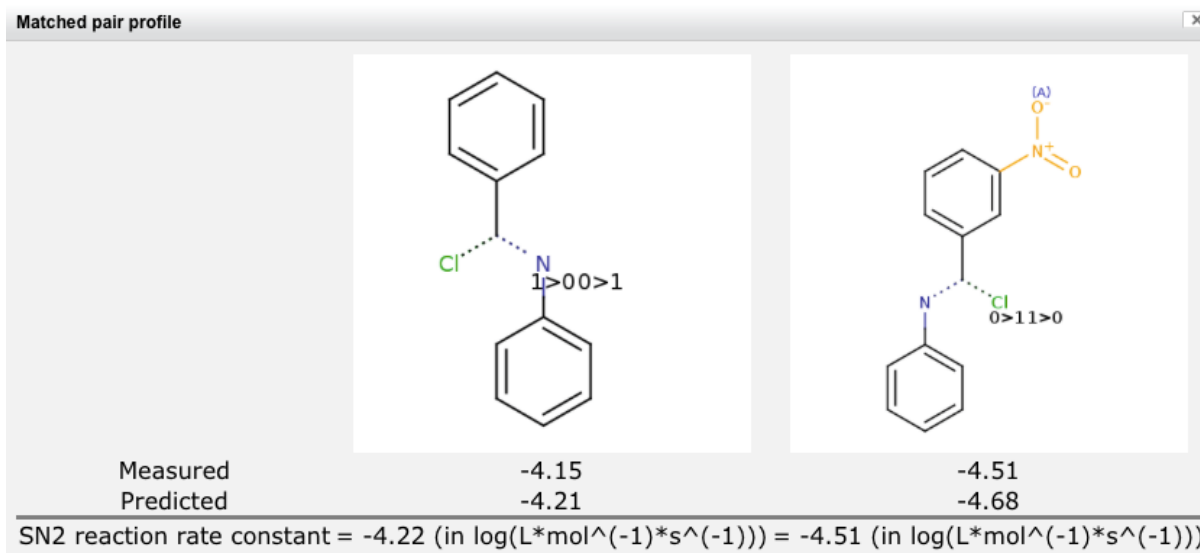


<http://www.wikipedia.org/>

SN2 vs. SN1 reaction



SN2



SN1



Evaluation in challenges

Computational Toxicology Research

[Contact Us](#)

You are here: [EPA Home](#) » [Research & Development](#) » [CompTox](#) » Chemical Data Challenges & Release

Key Links

[CompTox Home](#)
[Basic Information](#)
[Organization](#)
[EPA Exposure Research](#)

[Research Projects](#)
[Chemical Databases](#)
[ToxCast Stakeholder Events](#)
[EPA Chemical Safety Research](#)

[Research Publications](#)
[Scientific Reviews](#)
[Communities of Practice](#)
[ToxCast Data Challenges](#)

[Staff Profiles](#)
[CompTox Partners](#)
[Jobs and Opportunities](#)

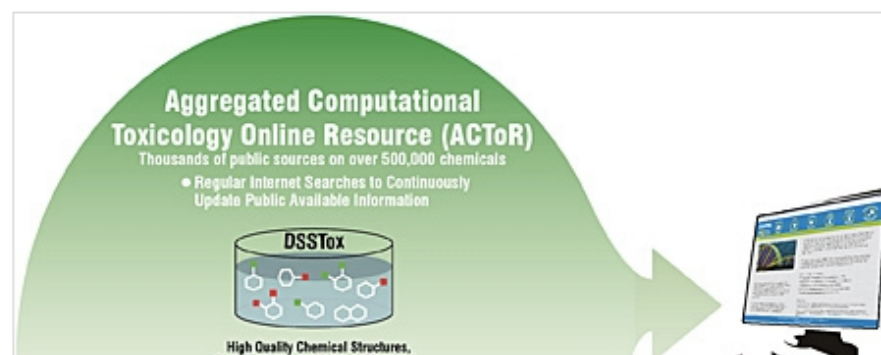
ToxCast Chemical Data Challenges and Release

EPA's high-throughput screening data on 1,800 chemicals is accessible through the interactive Chemical Safety for Sustainability Dashboards (iCSS dashboard). The iCSS dashboard provides user-friendly and customizable access to toxicity data from ToxCast and Tox21 high-throughput chemical screening technologies.

Using the [TopCoder](#) and [InnoCentive](#) crowd-sourcing platform, EPA invited the science and technology community to work with the data and provide solutions for how the new toxicity data can be used to predict potential health effects. The ToxCast data challenges focused on using this data and other publicly available data to predict the lowest effect level from traditional toxicity studies using laboratory animals. Challenge winners received awards for solving this challenge.

Key Links

- [Lowest Effect Level Challenge Results](#) (PDF, 497KB, 18pp)
- [Chemical Safety for Sustainability Dashboards](#)
- [Complete ToxCast Phase II Data & Files](#)
- [TopCoder Challenge](#)
- [InnoCentive Challenge](#)
- [Stakeholder Workshops](#)





Open call ends: November 14, 2014



About the Data



The Challenge

The 2014 [Tox21](#) data challenge is designed to help scientists understand the potential of the chemicals and compounds being tested through the [Toxicology in the 21st Century](#) initiative to disrupt biological pathways in ways that may result in toxic effects.

The goal of the challenge is to "crowdsource"



All challenge winners will receive the opportunity to submit a paper for publication in a special thematic issue of [Frontiers in Environmental Science](#) and recognition on the NCATS website and via social media.

Modeling capabilities

- Top-I rank submission model (May 2014) – entry by Sergii Novotarskyi*
- Two Top-I rank individual sub-challenges and overall best balanced accuracy for all targets (January 2015) – entry by Ahmed Abdelaziz**



*Novotarskyi, S. et al. Chem. Res. Toxicol. 2016, 29, 768-75.

**Abdelaziz, A. et al. Front. Environ. Sci. 2016, 4, 2.

Conclusions

- **On-line tools are important for dissemination**
- OCHEM
 - Can used for chemical data searching
 - ToxAlerts for understanding of hemical data
 - SetCompare can be useful for comparison of sets of molecules
 - Has a powerful modeling framework
 - Easily handles millions of compounds
 - Contribute highly predictive models
 - MMP analysis can complement modeling efforts

Acknowledgements



Dr. Y. Sushko
Dr. S. Novotarskyi
Mr. R. Körner
Mr. A. Abdelaziz
Mrs. E. Salmina

Prof. M. Sattler (HMGU)
Prof. G. Wess (HMGU)

Dr. K. Hadian (TOX, HMGU)
Dr. A. Williams (USA)
Dr. D. Lowe (UK)
Dr. H. Varbanov (Switzerland)
Prof. N. Haider (Austria)



Alexander von Humboldt
Stiftung / Foundation

Thank you for your attention!