



UPPSALA
UNIVERSITET



OpenTox Euro 2013

Chemical decision support in toxicology and pharmacology

Ola Spjuth

ola.spjuth@farmbio.uu.se

Department of Pharmaceutical Biosciences

Faculty of Pharmacy, Uppsala University





UPPSALA
UNIVERSITET

Open standards facilitates integration

- Aim: Take advantage of the OpenTox infrastructure in Bioclipse
 - Look up services (e.g. data and models)
 - Consume services (download, upload, run predictions)
 - Create nice GUI for OpenTox infrastructure
 - Integrate with Bioclipse offline capabilities

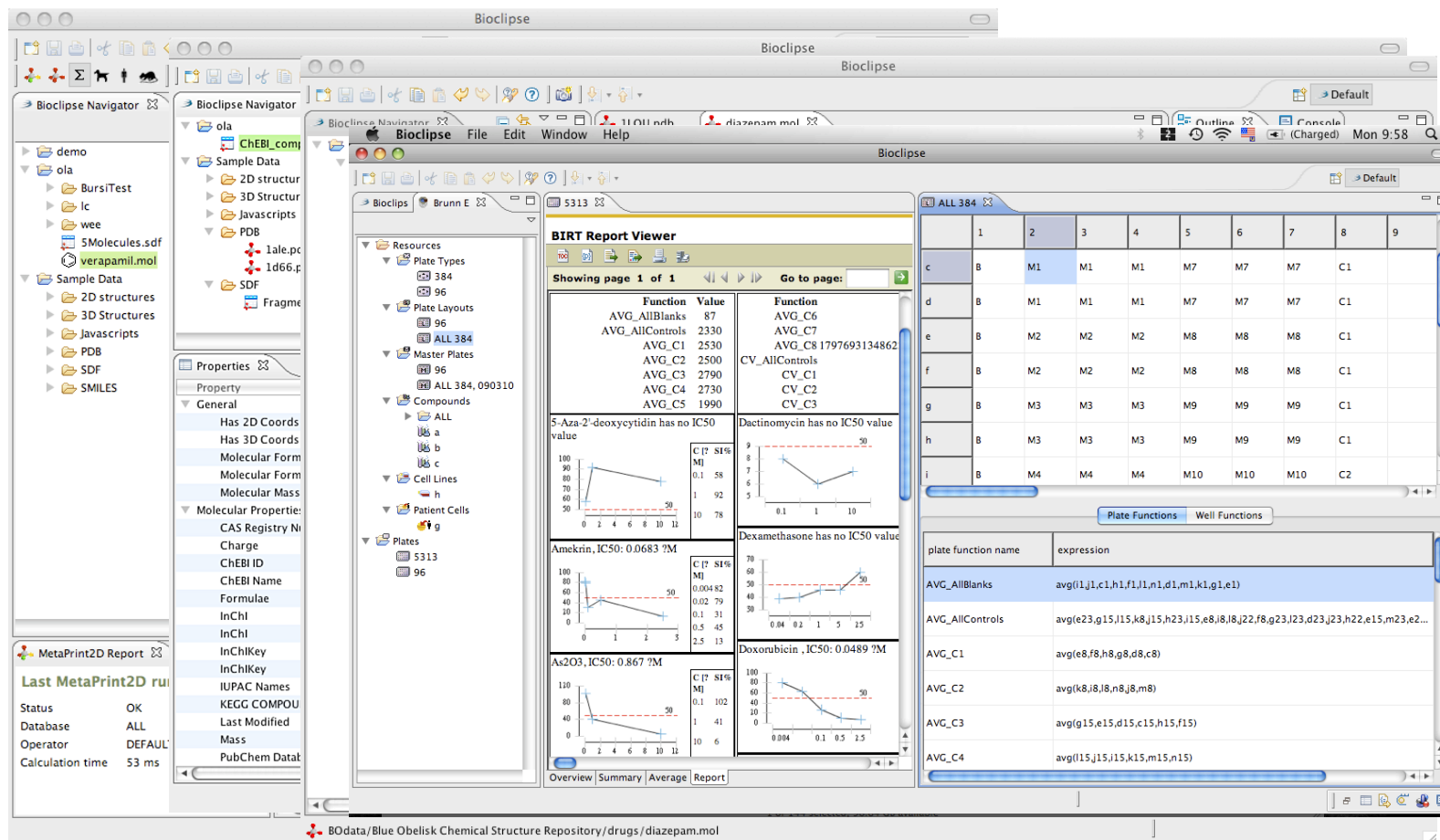




UPPSALA
UNIVERSITET



Bioclipse – an open source workbench for the life sciences



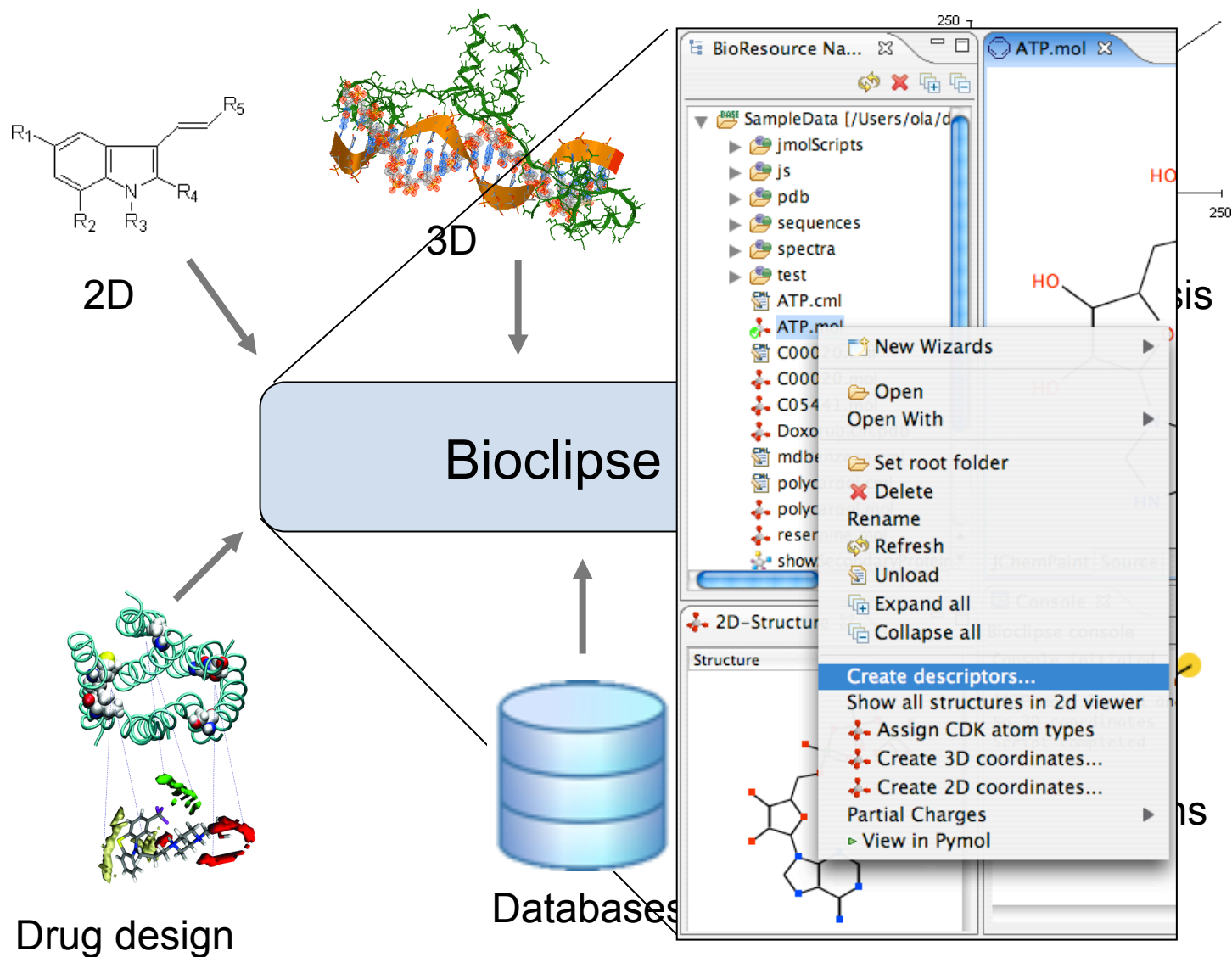
O. Spjuth, J. Alvarsson, A. Berg, M. Eklund, S. Kuhn, C. Mäsak, G. Torrance, J. Wagener, E.L. Willighagen, C. Steinbeck, and J.E.S. Wikberg. *Bioclipse 2: A scriptable integration platform for the life sciences*. BMC Bioinformatics 2009, **10**:397

Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, Wagener J, Murray-Rust P, Steinbeck C, Wikberg JES: *Bioclipse: an open source workbench for chemo- and bioinformatics*. BMC Bioinformatics 2007, **8**:59.



UPPSALA
UNIVERSITET

Component-based architecture





UPPSALA
UNIVERSITET

Bioclipse Scripting Language

The screenshot shows the Bioclipse Javascript Console with two code snippets. Snippet (a) contains three lines of code: `biows.queryEMBL("X56734")`, `biows.queryRefseq("NM_000410")`, and `biows.queryUniProtKB("INSR_HUMAN")`. Snippet (b) contains three lines of code: `seqs = biows.queryEMBL("X56734,X56735");`, `aln = kalignws.alignDNA(seqs);`, and `biojava.sequencesToFASTAfile(aln, "save here");`. To the right of the console is a 'New' wizard dialog box titled 'Select a wizard'. It lists various wizards under categories like General, Cheminformatics, Data, Download, and Scripting. The 'Scripting' category is expanded, showing options like 'Get structure from PubChem' and 'Query WSDbfetch at EBI'. The 'Next >' button is highlighted.

(a)

```
biows.queryEMBL("X56734")
biows.queryRefseq("NM_000410")
biows.queryUniProtKB("INSR_HUMAN")
```

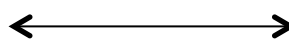
(b)

```
seqs = biows.queryEMBL("X56734,X56735");
aln = kalignws.alignDNA(seqs);
biojava.sequencesToFASTAfile(aln, "save here");
```

(c)



gist.github

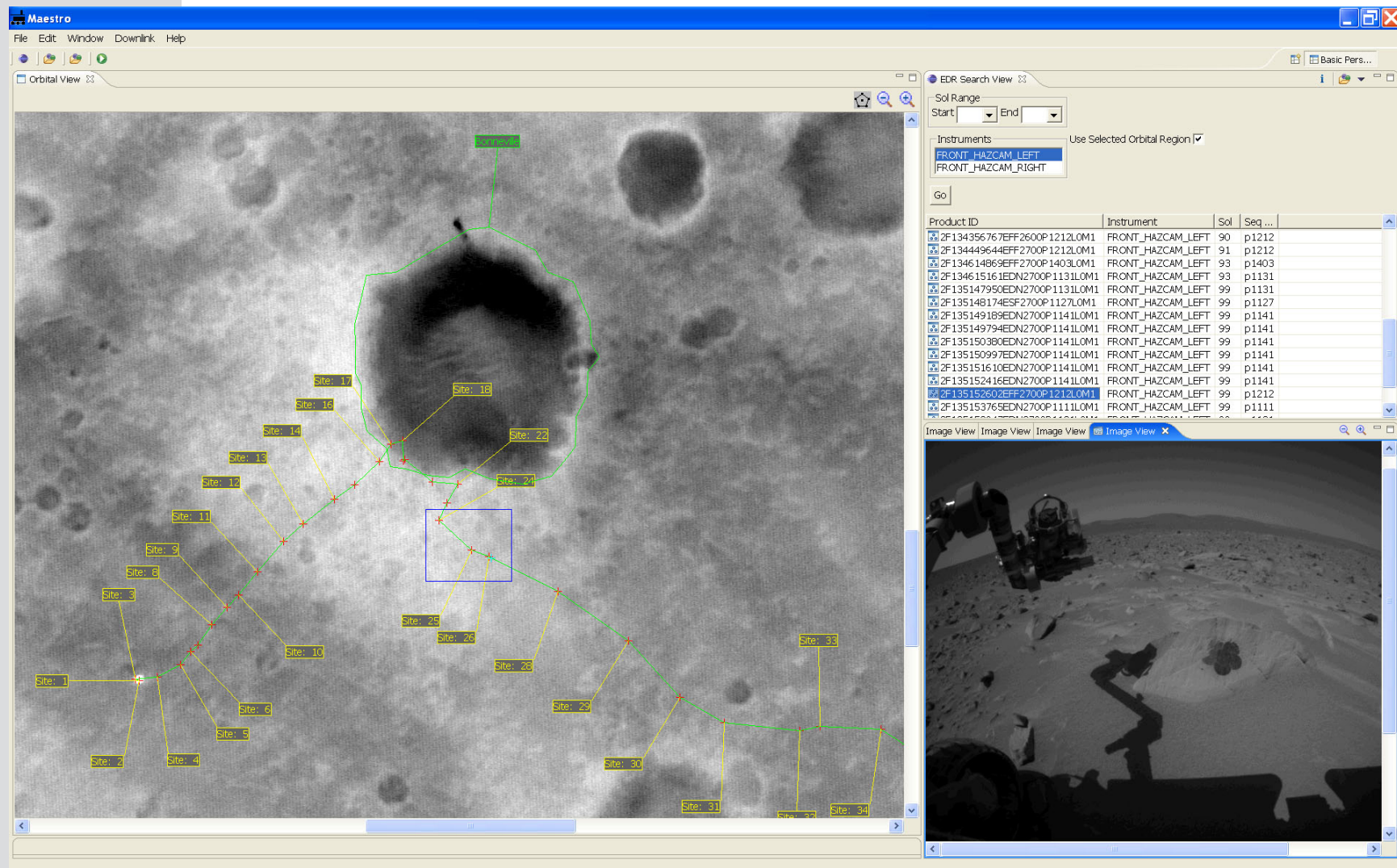


experiment



UPPSALA
UNIVERSITET

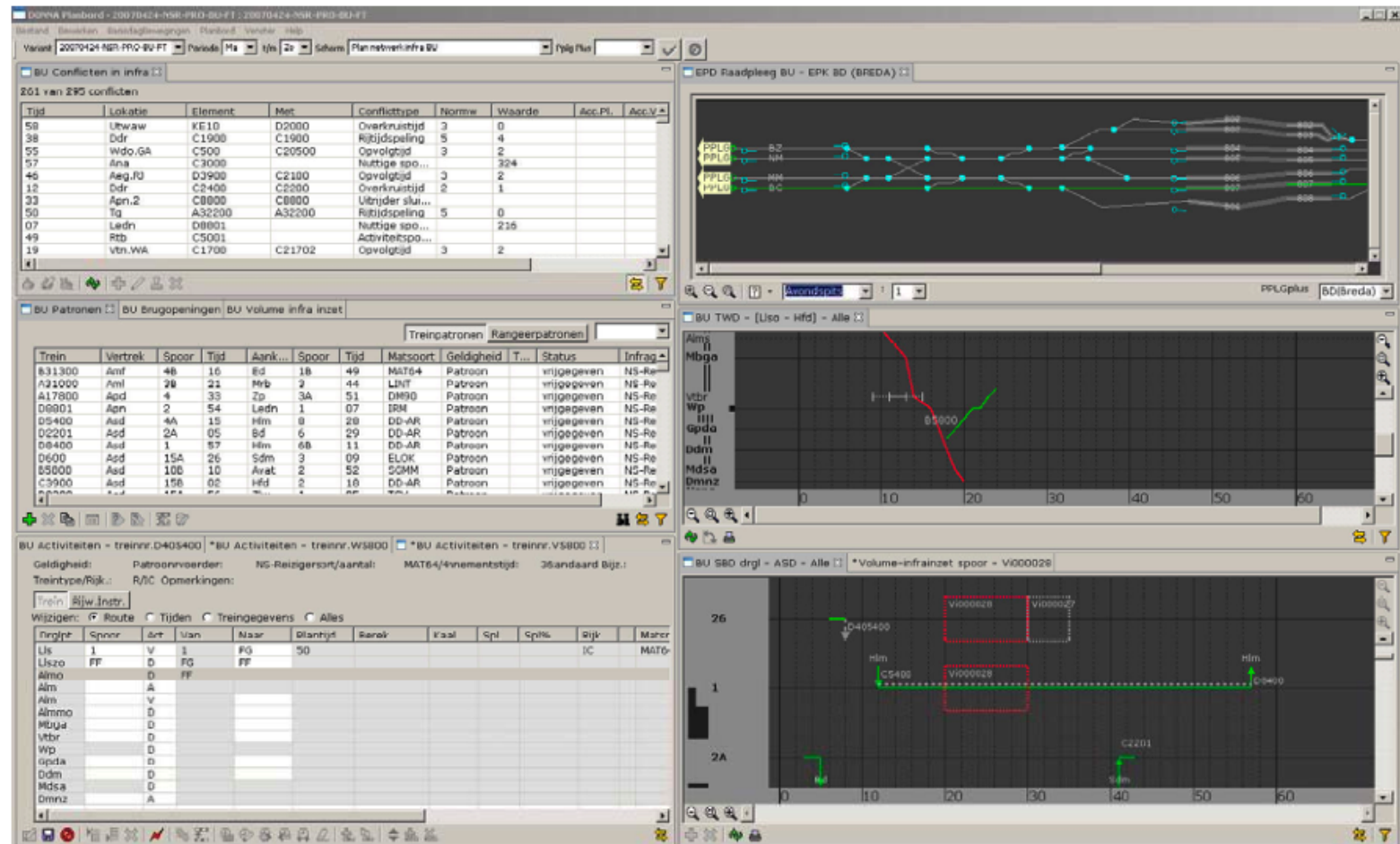
Eclipse: Proven technology





UPPSALA
UNIVERSITET

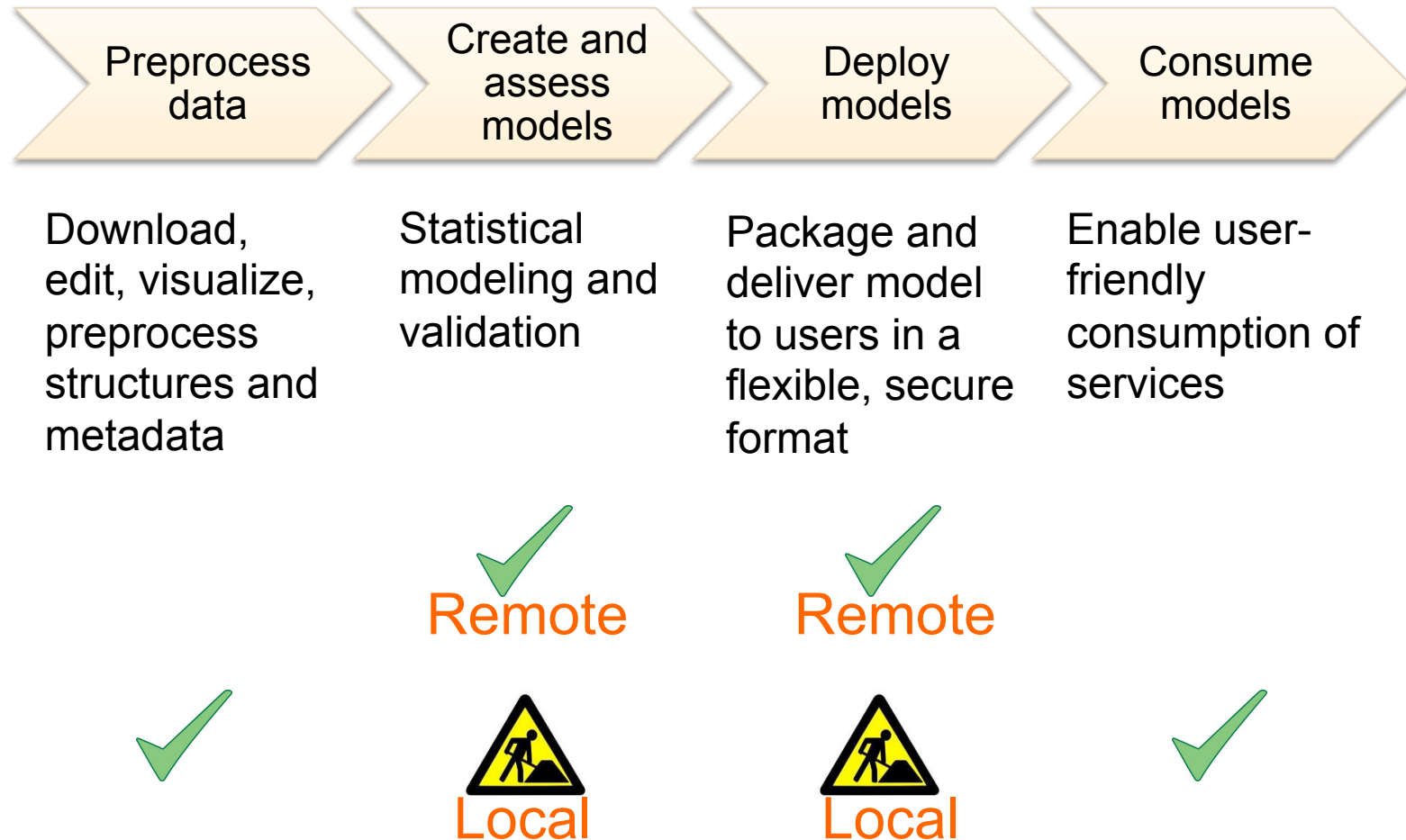
Example: Dutch railways





UPPSALA
UNIVERSITET

From data to predictions



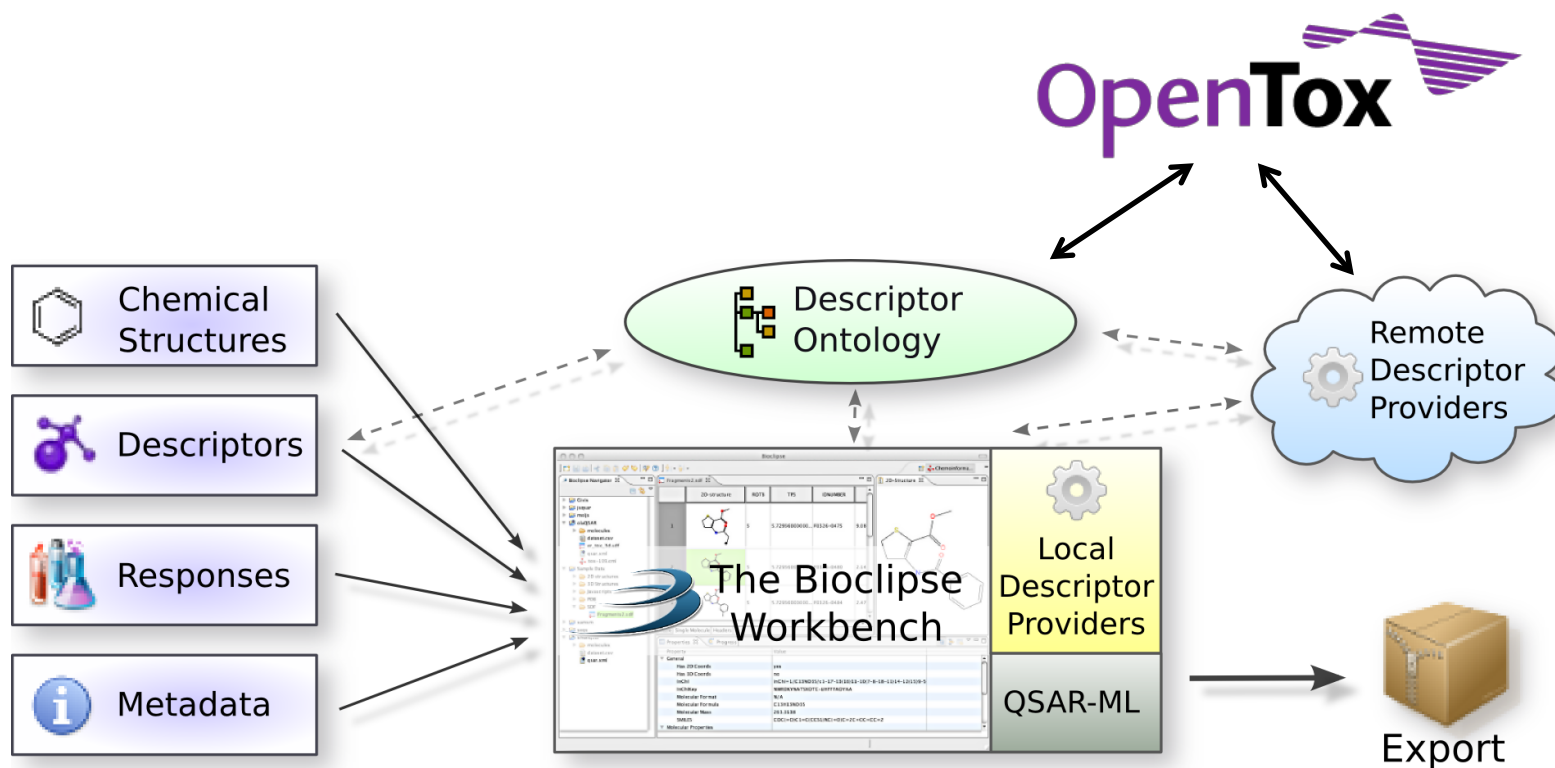
OpenTox





UPPSALA
UNIVERSITET

Bioclipse-QSAR: Reproducible QSAR datasets

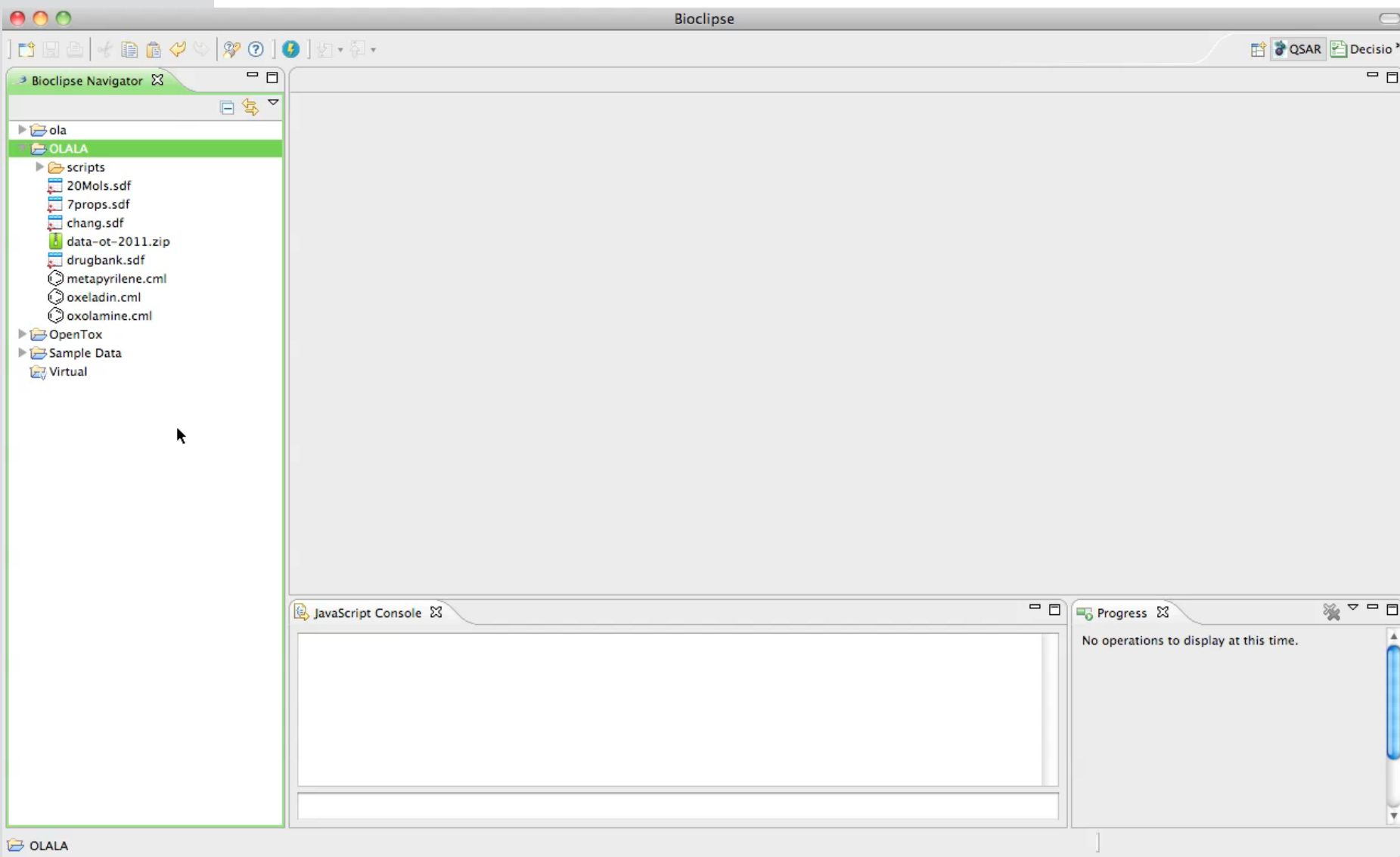


O. Spjuth, E.L. Willighagen, R. Guha, M. Eklund, and J.E.S. Wikberg. *Towards interoperable and reproducible QSAR analyses: Exchange of data sets.* Journal of Cheminformatics 2010, **2**:5



UPPSALA
UNIVERSITET

Bioclipse-QSAR with OpenTox





UPPSALA
UNIVERSITET

Bioclipse Decision Support for assessing chemical liabilities

- Original plan: Use Bioclipse workbench and provisioning system to:
 - Run models locally
 - fast execution, no need for network connection
 - Deliver a customized prediction workbench
 - Just install the plugins you desire
- Evolution: Also take advantage of networked services (e.g. SOAP, XMPP, OpenTox)



UPPSALA
UNIVERSITET

Chemical Decision Support for safety assessment

- Integrate various predictive models
 - Similarity searches (InChi, signatures, fingerprints)
 - Structural alerts (toxicophores)
 - QSAR models (classification, regression)
- Demonstrated on drug safety endpoints:
 - Mutagenicity (AMES)
 - Carcinogenicity (CPDB)
 - AHR inhibition (PubChem BioAssay 2796)

O. Spjuth, L. Carlsson, M. Eklund, E. Ahlberg Helgee, and Scott Boyer.
Integrated decision support for assessing chemical liabilities.
J. Chem. Inf. Model, 2011, 51 (8), pp 1840-1847



UPPSALA
UNIVERSITET

Bioclipse Decision Support

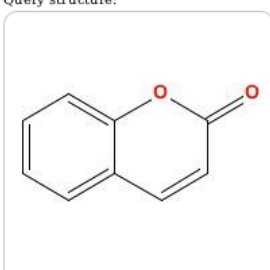
Bioclipse

Decision Sup... Default

20Mols.sdf Report Editor

Report date: 2010/08/26

Query structure:



Properties

Mol.	146.143
H donors:	1
H	1
alogP:	1.031

Number of Consensus: 4
Endpoints: 2 negative, 1 inconclusive

Compound: coumarin
SMILES: O=C1OC2=CC=CC=C2C=C1
Formula: C₉H₆O₂
InChI: InChI=1S/C9H6O2/c10-9-6-5-7-3-1-2-4-8(7)/11-9/h1-6H

Endpoint: AHR **NEGATIVE**

Model: AHR exact matches
Consensus: INCONCLUSIVE

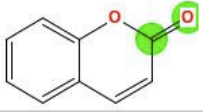
Model: AHR nearest neighbour
Consensus: **NEGATIVE**

Model: AHR Signature Alerts
Consensus: **NEGATIVE**

Model: AHR Signature Significance
Consensus: **NEGATIVE**

Compound: [O](=[C])
Classification: **NEGATIVE**

Details:



1 of 3

Decision Support

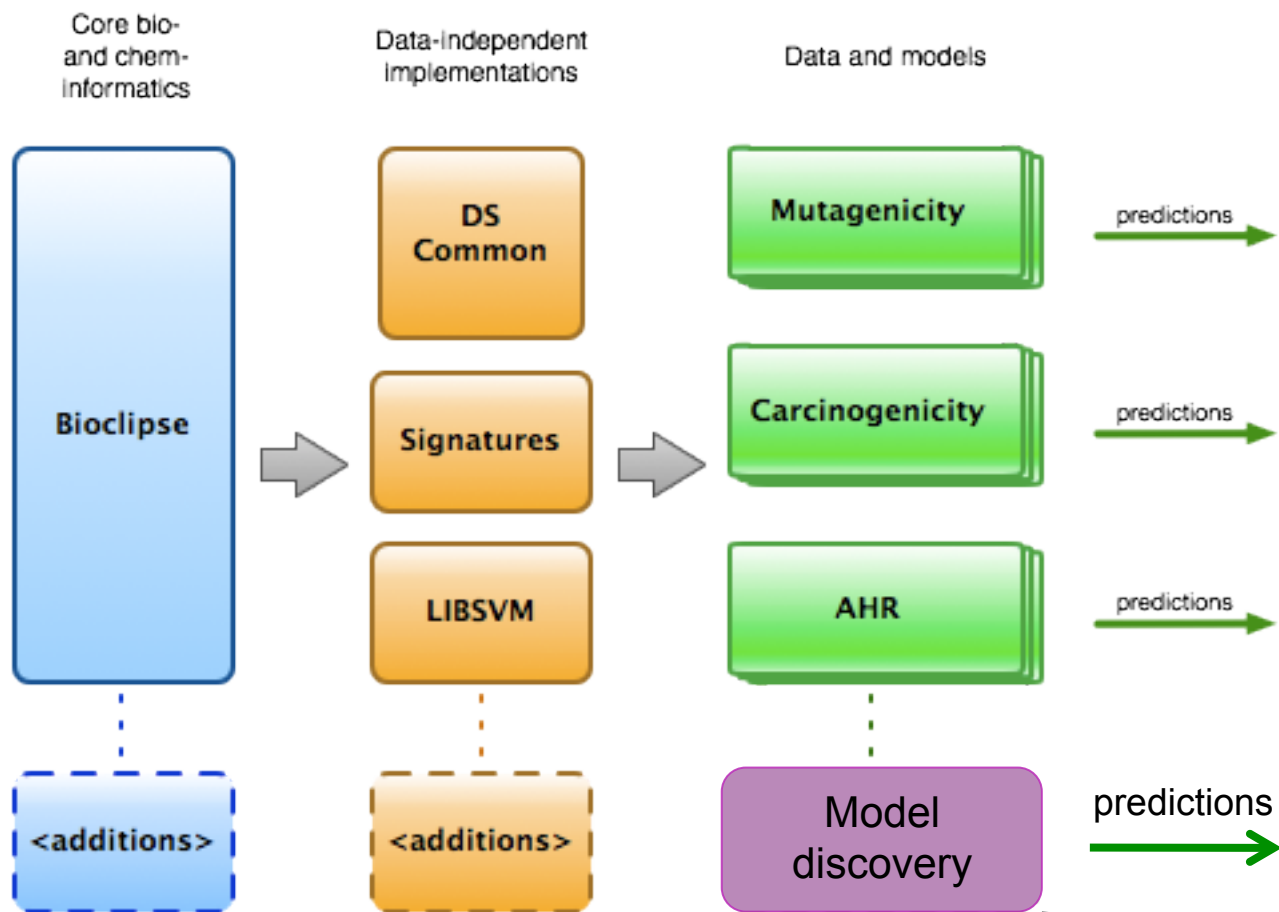
- AHR
 - AHR Signature Alerts
 - AHR Signature Significance [1 neg]
 - [O](=[C])
 - AHR exact matches
 - AHR nearest neighbour
 - Carcinogenicity
 - CPDB Signature Alerts
 - CPDB Signature Significance [1 pos]
 - Result: 1.002
 - CPDB exact matches [1 pos]
 - Index 199
 - CPDB nearest neighbour
 - Mutagenicity
 - Ames Signature Significance [1 neg]
 - [C]([C]([O]=[O]))
 - Ames Structural Alerts
 - Ames exact matches [1 pos]
 - 91-64-5
 - Ames nearest neighbour [3 neg]
 - 90-33-5 [tanimoto=0.78]
 - 2107-76-8 [tanimoto=0.76]
 - 26093-31-2 [tanimoto=0.75]

Consensus: **NEGATIVE**



UPPSALA
UNIVERSITET

Bioclipse Decision Support



OpenTox



-
- Decision Support**
- Carcinogenicity**
 - CPDB Signature Alerts
 - CPDB exact matches
 - CPDB nearest neighbour
 - CPDB regression
 - Mutagenicity**
 - Ames Signature Significance
 - Ames Structural Alerts
 - Ames exact matches
 - Ames nearest neighbour
 - OpenTox**
 - ECOSAR LC50 fish [excluded]
 - IST Carcinogenicity model (Canc) <http://www.w3.org/2001/XMLSchema#boolean>
 - IST Fish Toxicity model (LC50_mmol) <http://www.w3.org/2001/XMLSchema#boolean>
 - IST Salmonella Mutagenicity model (SAL) <http://www.w3.org/2001/XMLSchema#boolean>
 - Lipinski Rule of Five [excluded]
 - MLR model for Exp LogKow <http://www.w3.org/2001/XMLSchema#boolean>
 - MLR model for caco2 <http://www.w3.org/2001/XMLSchema#boolean>
 - MolecularWeight [excluded]
 - OpenTox model created with TUM's J48 model learning v1.0 [excluded]
 - OpenTox model created with TUM's kNNregression model [excluded]
 - OpenTox model created with TUM's kNNregression model [excluded]
 - QSAR SRC KOWWIN fingerprints AD [excluded]
 - START biodegradation and persistence plug-in [excluded]
 - SmartCYP: Cytochrome P450-Mediated Drug Metabolism [excluded]
 - ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity [excluded]
 - ToxTree: Cramer rules [excluded]
 - ToxTree: Eye irritation [excluded]
 - ToxTree: Skin irritation [excluded]
 - ToxTree: Skin sensitisation alerts (M. Cronin) [excluded]
 - ToxTree: Structure Alerts for the in vivo micronucleus assay [excluded]
 - ToxTree: Verhaar scheme (modified) for predicting toxicity [excluded]
 - ToxTree: Verhaar scheme for predicting toxicity mode of action [excluded]
 - XLogP [excluded]
 - pKa [excluded]



UPPSALA
UNIVERSITET

OpenTox in Bioclipse

Bioclipse

Decision Support

Bioclipse N

10mols.sdf metoprolol.cml

ola
OLALA
olaqsar
OpenTox
10mols.sdf
3drugs.sdf
ambit0.sdf
ambit1.sdf
ambit2.sdf
ambit3.sdf
ambit4.sdf
candesartan.cml
danthron.mol
metoprolol.cml
Sample Data
WEEQSAR

Chemical structure of Metoprolol (2D):

CC(C)NCC(O)COc1ccc(COC)cc1

Properties

Property	Value
General	
Has 2D Coords	yes
Has 3D Coords	no
InChI	Failed to calculate
InChIKey	Failed to calculate
Molecular Format	CML (Single 2D Molecule)
Molecular Formula	C15H25NO3
Molecular Mass	267.3645
SMILES	OC(COC1=CC=C(C=C1)CCOC)CNC(C)C

2D-Structure

Decision Support

- Carcinogenicity
 - CPDB Signature Alerts
 - CPDB exact matches
 - CPDB nearest neighbour
 - CPDB regression
- Mutagenicity
 - Ames Signature Significance
 - Ames Structural Alerts
 - Ames exact matches
 - Ames nearest neighbour
- OpenTox
 - ECOSAR LC50 fish
 - IST Carcinogenicity model (Canc)^{http://www.w3.org/2}
 - IST Fish Toxicity model (LC50_mmol)^{http://www.w3.o}
 - IST Salmonella Mutagenicity model (SAL)^{http://www.w}
 - Lipinski Rule of Five
 - MLR model for Exp LogKow^{http://www.w3.org/2001/}
 - MLR model for caco2^{http://www.w3.org/2001/XMLSc}
 - MolecularWeight
 - OpenTox model created with TUM's J48 model learning v
 - OpenTox model created with TUM's kNNregression mode
 - OpenTox model created with TUM's kNNregression mode
 - QSAR SRC KOWWIN fingerprints AD
 - START biodegradation and persistence plug-in
 - SmartCYP: Cytochrome P450-Mediated Drug Metabolism
 - ToxTree: Benigni/Bossa rules for carcinogenicity and mu
 - ToxTree: Cramer rules
 - ToxTree: Eye irritation
 - ToxTree: Skin irritation
 - ToxTree: Skin sensitisation alerts (M. Cronin)
 - ToxTree: Structure Alerts for the in vivo micronucleus as
 - ToxTree: Verhaar scheme (modified) for predicting toxic
 - ToxTree: Verhaar scheme for predicting toxicity mode of
 - XLogP
 - pKa



OpenTox in Bioclipse (2)

The screenshot displays the Bioclipse software interface. On the left, a file explorer shows a project named 'Bioclipse N' containing folders like 'ola', 'OLALA', 'olaqsar', and 'OpenTox'. The 'OpenTox' folder is expanded, showing files such as '10mol.sdf', '3drugs.sdf', and several 'ambit' files. The central workspace is currently empty. At the bottom, there are three tabs: 'Properties' (showing a table with 'Property' and 'Value' columns), 'JavaScript Console', and '2D-Structure'. On the right side, a 'Decision Support' panel is active, displaying a hierarchical list of models categorized by 'Carcinogenicity', 'Mutagenicity', and 'OpenTox'. Each category lists various models and their associated URLs or descriptions.

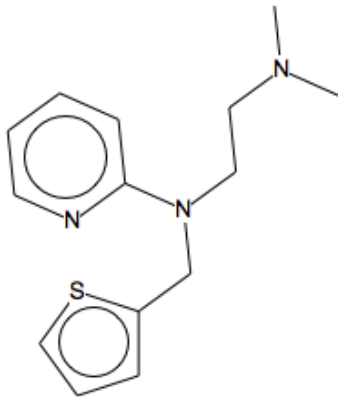


UPPSALA
UNIVERSITET

Report model validation

metapyrilene.cml

Decision Support



▼ Cancer growth inhibition

▼ Carcinogenicity

- CPDB Signature Alerts [no hits]
- CPDB Signature Significance [no hits]
- CPDB exact matches [no hits]
- CPDB nearest neighbour [no hits]

▼ Mutagenicity

- Ames Mutagenicity [1 pos]
 - accuracy: 0.582
- Mutagenicity
 - Ames Signature Significance [1 neg]
 - [C]([N])
 - Ames Structural Alerts [no hits]
 - Ames exact matches [no hits]
- Ames nearest neighbour [1 neg]
 - 148-65-2 [animoto=0.87]

Properties

JavaScript Console

Property	Value
▼ Dataset	
Dataset name	Bursi Mutagenicity Dataset
Descriptors	Signatures (height 0-3)
Observations	4337
URL	http://pubs.acs.org/doi/abs/10.1021/jm040835a
Variables	23226
▼ Model	
Learning model	SVM
Learning parameters	kernel=RBF, c=10, gamma=0.002
Model choice	Maximal accuracy, with 5-fold cross-validated accuracy as objective function
Model name	Ames Signature Significance
Model performance	0.85
Model type	QSAR
Model validation	Accuracy measured on an external test set
▼ Result	
Consensus	NEGATIVE
Prediction time	251 ms
Status	FINISHED



UPPSALA
UNIVERSITET

Scripting OpenTox in Bioclipse

The screenshot displays the Bioclipse application window. The top menu bar includes File, Edit, Window, Install, and Help. The Bioclipse Navigator on the left shows a project structure with folders for Bioclipse-OpenTox Workshop, NanoTox, and OpenTox, and a Test folder. The main editor window shows a JavaScript script in test5.js:

```
1 var ontologyService = "http://apps.ideaconsult.net:8080/ontology/";
2 js.clear();
3
4 sets = opentox.searchDataSets(ontologyService, "EPA");
5 js.say(setsets);
6
```

The JavaScript Console at the bottom shows the output of the script:

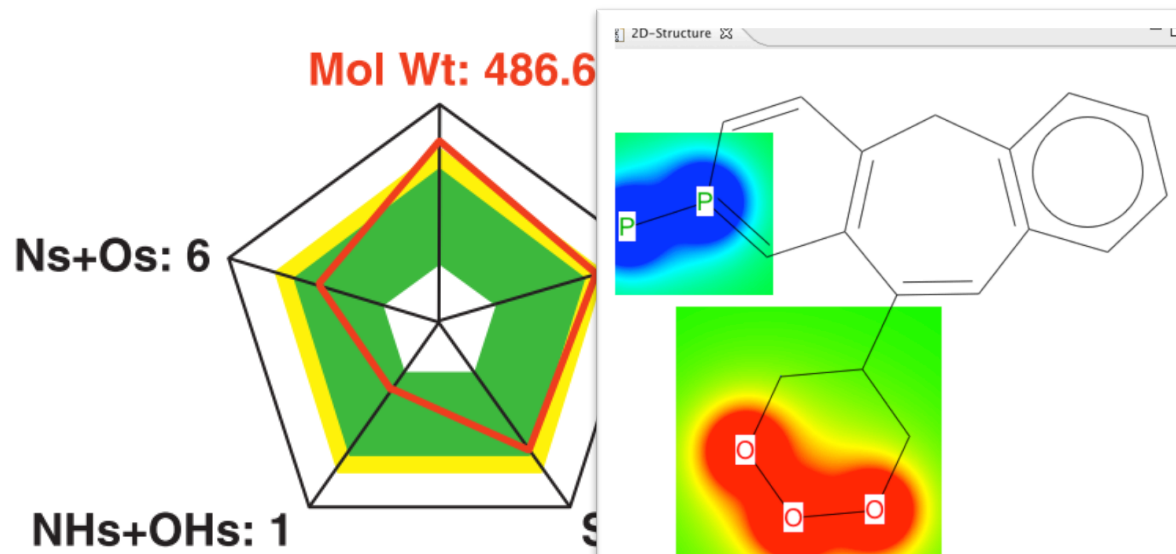
```
[["set","title"],
["http://apps.ideaconsult.net:8080/ambit2/dataset/11","DBPCAN: EPA Water
Disinfection By-Products with Carcinogenicity Estimates"],
["http://apps.ideaconsult.net:8080/ambit2/dataset/12","EPAFHM: EPA Fathead
Minnow Acute Toxicity"],
["http://apps.ideaconsult.net:8080/ambit2/dataset/14","IRISTR: EPA Integrated
Risk Information System (IRIS) Toxicity Review Data"],
["http://apps.ideaconsult.net:8080/ambit2/dataset/13","KIERBL: EPA Estrogen
Receptor Ki Binding Study (Laws et al.)"]
]
```

The status bar at the bottom indicates 'Writable', 'Insert', and '6 : 1'.



Related work

- Modeling in Bioclipse
 - Integrated solutions (e.g. libsvm)
 - Modeling with R
- Publish Bioclipse models using p2
- Visualizations



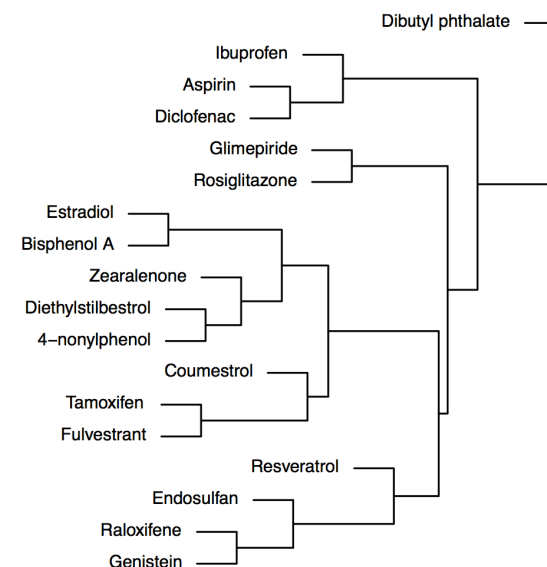


UPPSALA
UNIVERSITET

Related work 2

- Secure provisioning of offline models
- Model building on HPC/cloud resources
- Secondary pharmacology predictions
- Biological similarity: QuantMap

Schaal W, Hammerling U, Gustafsson MG, Spjuth O.
**Automated QuantMap for rapid quantitative
molecular network topology analysis.**
Bioinformatics. 2013 Sep 15;29(18):2369-70.





UPPSALA
UNIVERSITET

Next steps: Bioclipse-OpenTox

- Better indication of AA, status, classification, and confidence of results (API improvements?)
- Distinguish local and remote services better
- Highlight OpenTox substructure results
- Categorization of models



UPPSALA
UNIVERSITET

Predicting nanotoxicity using OpenTox from Bioclipse

Bioclipse

Decision Su... <Plug-in De... Java

NM Editor

Chemical Formula	Fe3O4
Type	METALOXIDE
Zeta Potential	-27.0 eV

Decision Support

- OpenTox
 - Burello Conductivity Band Energy [1 neg]
 - Toxic = NO



UPPSALA
UNIVERSITET

Acknowledgement: The Chemistry Development Kit (CDK)

- Most prominent open source Java library for cheminformatics
- Coordinated from Maastricht University (Egon Willighagen) and EBI (Christoph Steinbeck)
- ~200 citations
- Widely Used in academia and pharma industry
 - AstraZeneca, Bayer, Merck-Serono, Sanofi Aventis, Eli Lilly, Novartis



UPPSALA
UNIVERSITET



- Providing commercial support around the topics of Bioclipse and statistical modeling.
- Research, implementations, education, branding/deployment
- www.genettasoft.com



UPPSALA
UNIVERSITET

Acknowledgements

**Dept. Pharmaceutical Biosciences,
Uppsala University, SWE**

Jarl E. S. Wikberg

Martin Eklund

Jonathan Alvarsson

Arvid Berg

Samuel Lampa

Valentin Giorgiev

Wesley Schaal

**Dept Medical Sciences, Uppsala
University, SWE**

Mats Gustafsson

Ulf Hammerling

Karolinska Institutet, Sweden

Roland Grafström

Rebecca Ceder

Pekka Kohonen

Maastricht University, NL

Egon Willighagen

University of Copenhagen, DK

Patrik Rydberg

Michal Rostkowski

Ideaconsult, SLO

Nina Jeliazkova

AstraZeneca R&D, Sweden

Scott Boyer

Lars Carlsson

Ernst Ahlberg

Jan Person

Ola Engqvist

Tobias Noeske

Ulf Norinder

Douglas Connect

Barry Hardy



UPPSALA
UNIVERSITET

- Thank you -

Ola Spjuth

ola.spjuth@farmbio.uu.se

Department of Pharmaceutical Biosciences
Faculty of Pharmacy, Uppsala University

