



UPPSALA  
UNIVERSITET



OpenTox 2011

# Enabling decision support for chemical liability assessment

*Integrating Bioclipse and OpenTox*

**Ola Spjuth**

*ola.spjuth@farmbio.uu.se*

Department of Pharmaceutical Biosciences

Faculty of Pharmacy, Uppsala University





UPPSALA  
UNIVERSITET



# Bioclipse – an open source workbench for the life sciences

The screenshot displays the Bioclipse software interface. On the left is the 'Bioclipse Navigator' showing a hierarchical tree of resources like 'Sample Data', '2D structures', and 'SDF'. The main window is titled 'BIRT Report Viewer' and shows a table of data with columns for 'Function' and 'Value'. Below the table are several line graphs showing 'IC50' values for various compounds. On the right, a 'Table' view shows a grid of data points. At the bottom, a 'MetaPrint2D Report' is visible, showing status and calculation time.

Function	Value
AVG_AllBlanks	87
AVG_AllControls	2330
AVG_C1	2530
AVG_C2	2500
AVG_C3	2790
AVG_C4	2730
AVG_C5	1990

Function	Value
AVG_C6	1797693134862
AVG_C7	
AVG_C8	
CV_AllControls	
CV_C1	
CV_C2	
CV_C3	

plate function name	expression
AVG_AllBlanks	avg(l,j,l,c1,h1,f,l,j,n1,d1,m1,k1,g1,e1)
AVG_AllControls	avg(e23,g15,i15,k8,j15,h23,i15,e8,i8,j22,f8,g23,i23,j23,h22,e15,m23,e2...
AVG_C1	avg(e8,f8,h8,g8,d8,c8)
AVG_C2	avg(k8,j8,i8,n8,j8,m8)
AVG_C3	avg(g15,e15,d15,c15,h15,f15)
AVG_C4	avg(l15,j15,i15,k15,m15,n15)

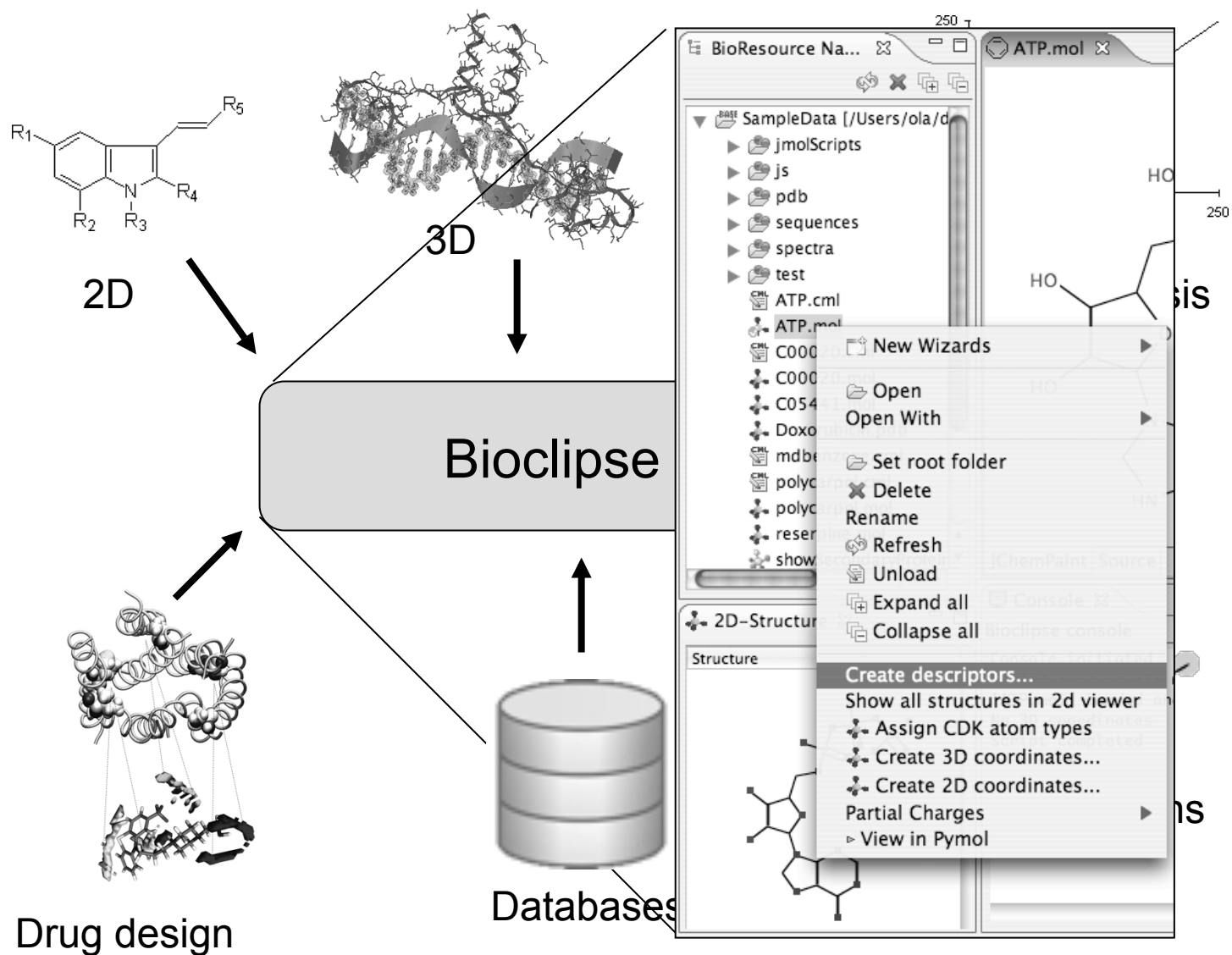
O. Spjuth, J. Alvarsson, A. Berg, M. Eklund, S. Kuhn, C. Mäsak, G. Torrance, J. Wagener, E.L. Willighagen, C. Steinbeck, and J.E.S. Wikberg. *Bioclipse 2: A scriptable integration platform for the life sciences*. BMC Bioinformatics 2009, **10**:397

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 interface with a Javascript Console window. The console contains two code snippets:

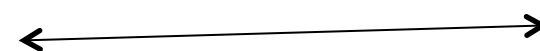
(a)

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

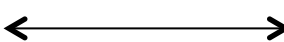
(b)

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

Snippet (a) is enclosed in a box with a 'mol-cdk.t' label on the left. Snippet (b) is enclosed in a box with a 'mol-cdk.t' label on the left. To the right, a 'Select a wizard' dialog box is open, showing a tree view of wizards under the 'Scripting' category. The 'Query WSDbfetch at EBI' wizard is selected. The dialog has 'Back', 'Next', 'Cancel', and 'Finish' buttons.



gist.github



my experiment



UPPSALA  
UNIVERSITET

# Eclipse: Proven technology

Maestro

File Edit Window Downlink Help

Orbital View

Bonneville

Site: 17 Site: 16 Site: 14 Site: 13 Site: 12 Site: 11 Site: 9 Site: 8 Site: 3 Site: 2 Site: 4 Site: 6 Site: 5 Site: 10 Site: 10 Site: 25 Site: 26 Site: 24 Site: 29 Site: 26 Site: 29 Site: 30 Site: 31 Site: 32 Site: 33 Site: 34

EDR Search View

Sol Range  
Start End

Instruments  
FRONT\_HAZCAM\_LEFT  
FRONT\_HAZCAM\_RIGHT

Use Selected Orbital Region

Go

Product ID	Instrument	Sol	Seq ...
2F134356767EFF2600P1212L0M1	FRONT_HAZCAM_LEFT	90	p1212
2F134449644EFF2700P1212L0M1	FRONT_HAZCAM_LEFT	91	p1212
2F134614869EFF2700P1403L0M1	FRONT_HAZCAM_LEFT	93	p1403
2F134615161EDN2700P1131L0M1	FRONT_HAZCAM_LEFT	93	p1131
2F135147950EDN2700P1131L0M1	FRONT_HAZCAM_LEFT	99	p1131
2F135148174ESF2700P1127L0M1	FRONT_HAZCAM_LEFT	99	p1127
2F135149189EDN2700P1141L0M1	FRONT_HAZCAM_LEFT	99	p1141
2F135149794EDN2700P1141L0M1	FRONT_HAZCAM_LEFT	99	p1141
2F135150380EDN2700P1141L0M1	FRONT_HAZCAM_LEFT	99	p1141
2F135150997EDN2700P1141L0M1	FRONT_HAZCAM_LEFT	99	p1141
2F135151610EDN2700P1141L0M1	FRONT_HAZCAM_LEFT	99	p1141
2F135152416EDN2700P1141L0M1	FRONT_HAZCAM_LEFT	99	p1141
2F135152602EFF2700P1212L0M1	FRONT_HAZCAM_LEFT	99	p1212
2F135153765EDN2700P1111L0M1	FRONT_HAZCAM_LEFT	99	p1111

Image View Image View Image View Image View



UPPSALA  
UNIVERSITET

# Example: Dutch railways

The screenshot displays a complex railway planning software interface with several panels:

- Top Panel:** Includes menu options like 'Bestand', 'Beveilen', 'Bereken', 'Bereken', 'Plan', 'Verst', and 'Help'. It also shows filters for 'Variabel' (20070424 NSR-PRO-BU-FT) and 'Periode' (Ma).
- BU Conflicten in infra:** A table listing 201 conflicts. The table has columns: Tijd, Lokatie, Element, Met, Conflicttype, Normw, Waarde, Acc.Pl., and Acc.V.
- BU Patronen:** A table listing train patterns with columns: Trein, Vertrek, Spoor, Tijd, Aank..., Spoor, Tijd, Matsoort, Geldigheid, T..., Status, and Infrag.
- BU Activiteiten:** A table listing activities with columns: Ringnr, Spoor, Art, Van, Naar, Dlantijd, Rerok, K'aal, Cpl, Cpl%, Zijk, and Mater.
- Graphs:** Two line graphs showing data trends over time (0 to 50). The top graph is titled 'EPD Raadpleeg BU - EPK BD (BREDA)' and the bottom one is 'BU SBD drgl - ASD - Alle'.



UPPSALA  
UNIVERSITET

# Bioclipse and OpenTox: Shared goals – predictive toxicology

- OpenTox
  - Defines an API, ontology, and services for predictive toxicology
- Bioclipse
  - Rich GUI, built for taking advantage of networked services
  - Local predictions



→ A good match!



UPPSALA  
UNIVERSITET

# From data to predictions

Preprocess  
data

Download,  
edit, visualize,  
preprocess  
structures and  
metadata



Create and  
assess  
models

Statistical  
modeling and  
validation



Deploy  
models

Package and  
deliver model  
to users in a  
flexible, secure  
format



Consume  
models

Enable user-  
friendly  
consumption of  
services



OpenTox

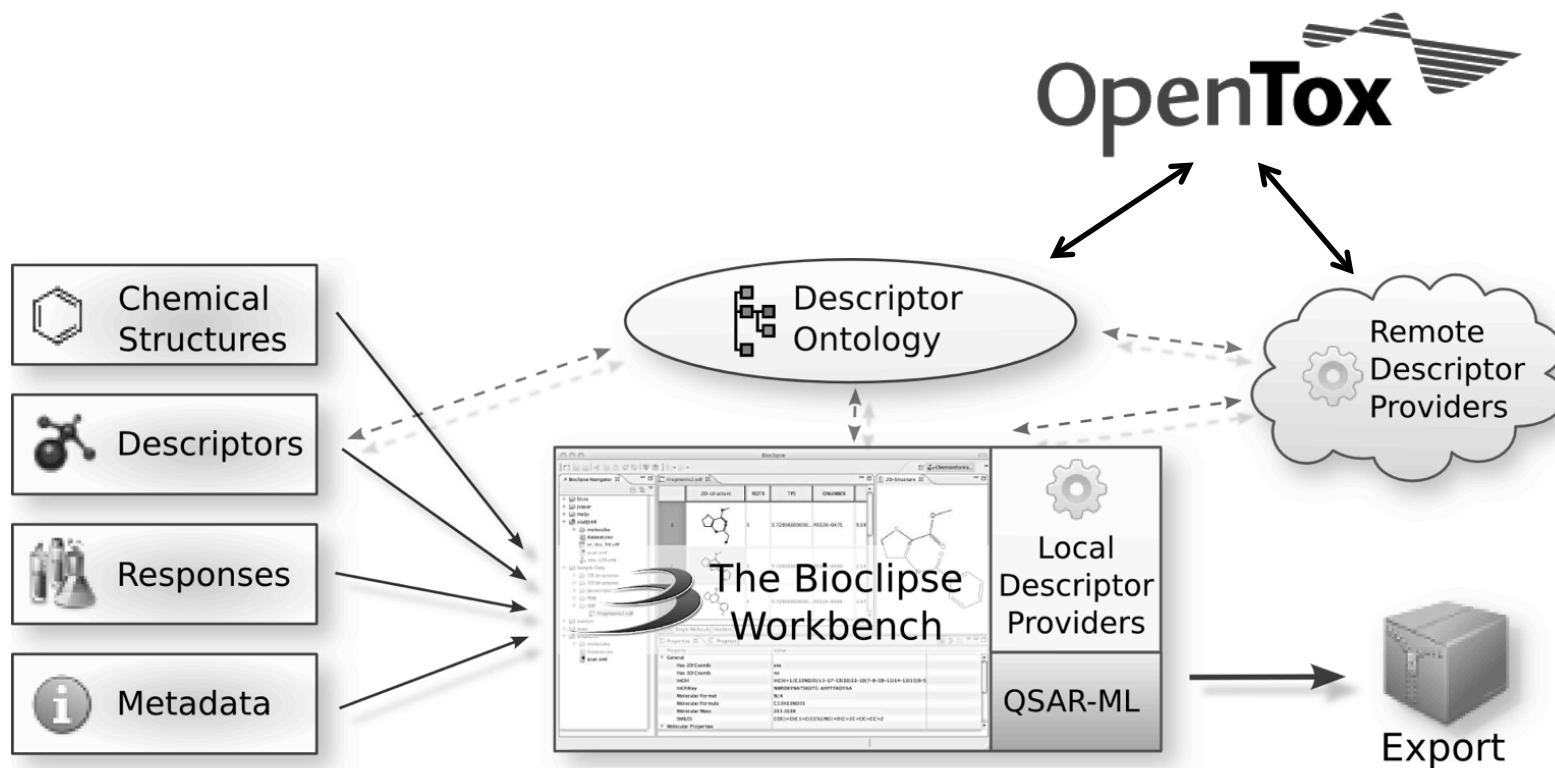






UPPSALA  
UNIVERSITET

# 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

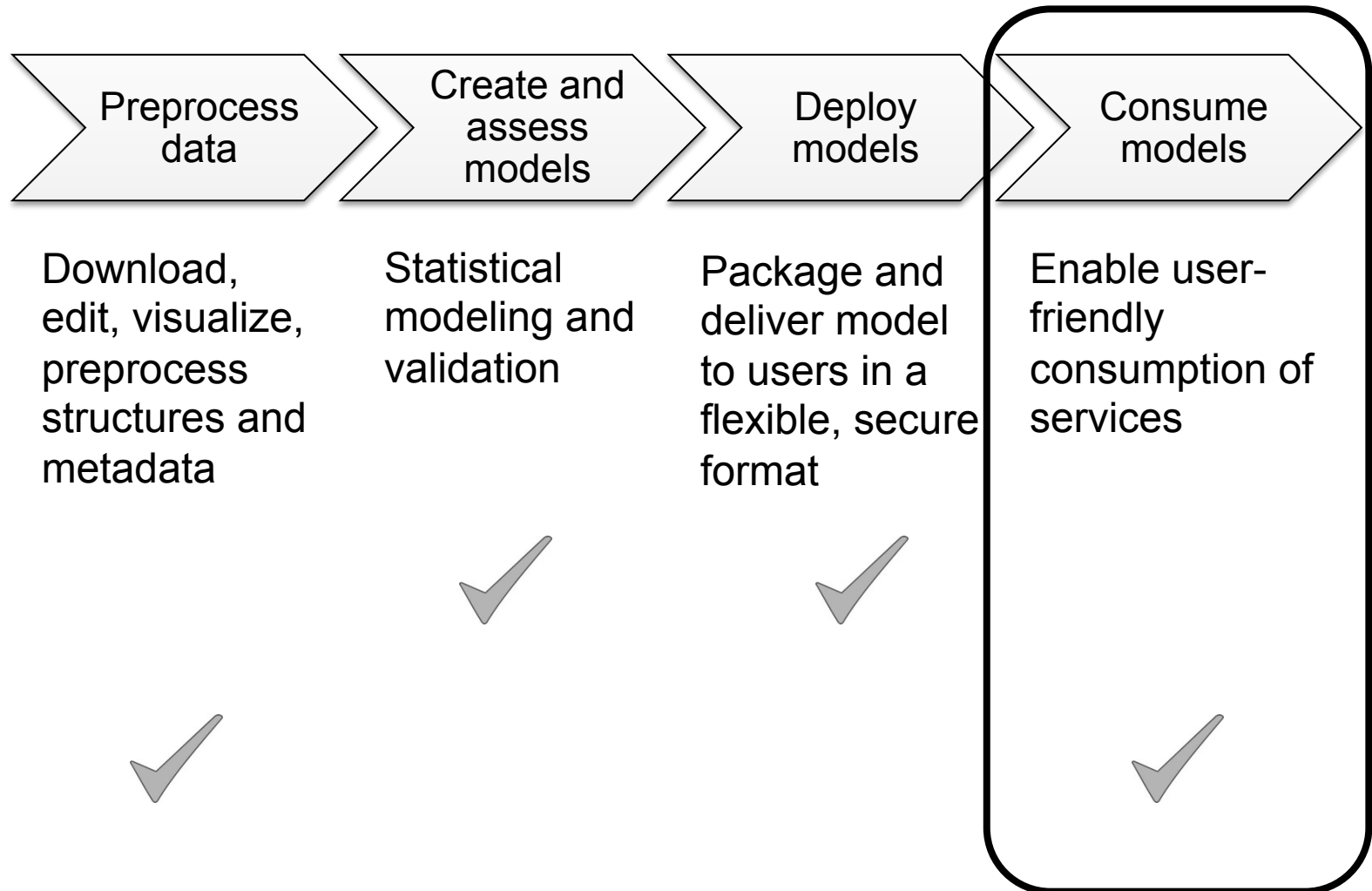
# Bioclipse-QSAR with OpenTox

The screenshot displays the Bioclipse software interface. The main window is titled "Bioclipse" and features a toolbar with various icons for file operations and navigation. On the left, the "Bioclipse Navigator" panel shows a project structure under the name "ola". The structure includes a folder named "OLALA" which contains a "scripts" folder and several files: "20Mols.sdf", "7props.sdf", "chang.sdf", "data-ot-2011.zip", "drugbank.sdf", "metapyrilene.cml", "oxeladin.cml", and "oxolamine.cml". Below the "OLALA" folder are three more folders: "OpenTox", "Sample Data", and "Virtual". The main workspace is currently empty. At the bottom, there are two panels: "JavaScript Console" and "Progress". The "JavaScript Console" panel is empty, and the "Progress" panel displays the message "No operations to display at this time." The status bar at the bottom left shows the project name "OLALA".



UPPSALA  
UNIVERSITET

# From data to predictions



OpenTox 





# 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

# Bioclipse Decision Support for assessing chemical liabilities

- Integrate various predictive technologies
  - Similarity searches (InChi, signatures, fingerprints)
  - Structural alerts (SMARTS + Signature alerts)
  - QSAR models (signatures, SVM)
- Safety endpoints based on open data:
  - 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*. Accepted in J. Chem. Inf. Model.



UPPSALA  
UNIVERSITET

# Bioclipse Decision Support

Bioclipse

Decision Sup... → Default

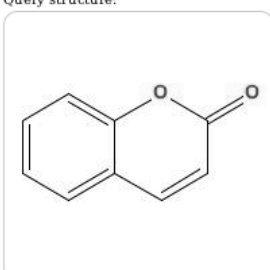
20Mols.sdf Report Editor

20Mols

**Bioclipse Decision Support**  
Chemical Liability Assessment

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

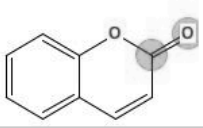
Model: AHR nearest neighbour  
Consensus: ✓ NEGATIVE

Model: AHR Signature Alerts  
Consensus: ✓ NEGATIVE

Model: AHR Signature Significance  
Consensus: ✓ NEGATIVE

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

Details:



Consensus: NEGATIVE

1 of 3

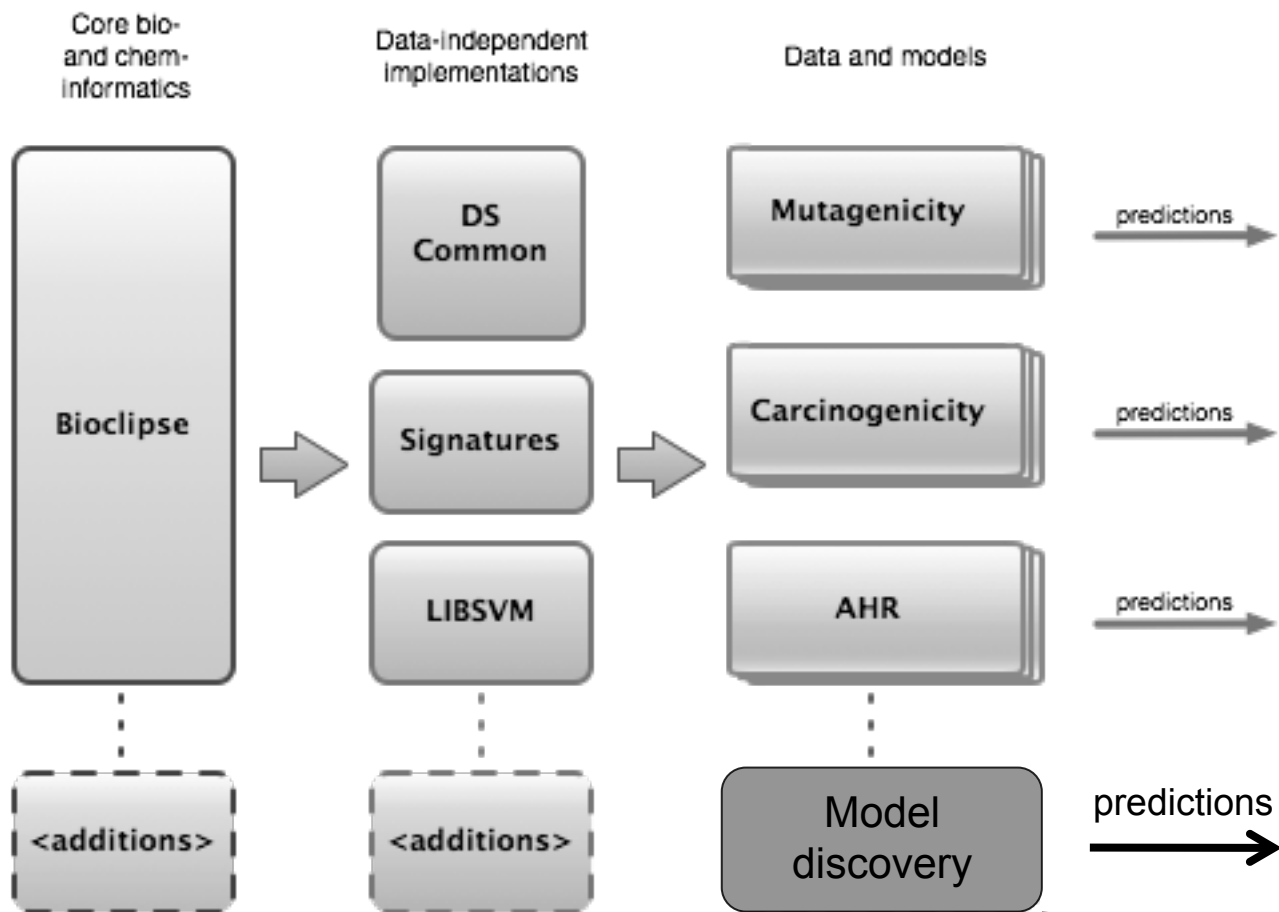
Decision Support

- ▼ 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]



UPPSALA  
UNIVERSITET

# Bioclipse Decision Support



OpenTox 



UPPSALA  
UNIVERSITET

# Rich GUI for predictions

The screenshot displays the Bioclipse software interface. At the top left is the Uppsala University logo. The main window is titled "Bioclipse" and shows a chemical structure of candesartan in the "Changed" panel. The left sidebar contains a file tree with folders like "ola", "OLALA", "olaqsar", and "OpenTox", and files such as "10mols.sdf", "ambit0.sdf", "ambit1.sdf", "ambit2.sdf", "ambit3.sdf", "ambit4.sdf", "candesartan.cml", "danthron.mol", and "metoprolol.cml". The "Decision Support" panel on the right lists various prediction models and their results, including:

- 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)<sup>http://www.w3.org/2</sup>
  - IST Fish Toxicity model (LC50\_mmol)<sup>http://www.w3.o</sup>
  - IST Salmonella Mutagenicity model (SAL)<sup>http://www.w</sup>
  - Lipinski Rule of Five [excluded]
  - MLR model for Exp LogKow<sup>http://www.w3.org/2001/</sup>
  - MLR model for caco2<sup>http://www.w3.org/2001/XMLSc</sup>
  - MolecularWeight [excluded]
  - 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 [excluded]
  - START biodegradation and persistence plug-in [excluded]
  - SmartCYP: Cytochrome P450-Mediated Drug Metabolism
  - ToxTree: Benigni/Bossa rules for carcinogenicity and mu
  - 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 as
  - ToxTree: Verhaar scheme (modified) for predicting toxic
  - ToxTree: Verhaar scheme for predicting toxicity mode of
  - XLogP [excluded]
  - pKa [excluded]

The bottom left shows a "Properties" table with columns for "Property" and "Value", which is currently empty. The bottom right shows a "2D-Structure" panel displaying the same chemical structure as the main window.





UPPSALA  
UNIVERSITET

# OpenTox in Bioclipse

Bioclipse

Decision Su...

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  
WEEEQSAR

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)<sup>http://www.w3.org/2</sup>
  - IST Fish Toxicity model (LC50\_mmol)<sup>http://www.w3.o</sup>
  - IST Salmonella Mutagenicity model (SAL)<sup>http://www.w</sup>
  - Lipinski Rule of Five
  - MLR model for Exp LogKow<sup>http://www.w3.org/2001/</sup>
  - MLR model for caco2<sup>http://www.w3.org/2001/XMLSc</sup>
  - 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





# 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
  - ▼ ● 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	<a href="http://pubs.acs.org/doi/abs/10.1021/jm040835a">http://pubs.acs.org/doi/abs/10.1021/jm040835a</a>
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

# From data to predictions



Download, edit, visualize, preprocess structures and metadata

Statistical modeling and validation

Package and deliver model to users in a flexible, secure format

Enable user-friendly consumption of services

OpenTox 



  
Remote



Local

  
Remote



Local





## Next steps...

- Improve UI
  - Better indication of A&A, encryption, status
  - Selection of local and remote services
  - Highlight substructure results
  - More wizards and editors for OpenTox functionality
- Report accuracy from OpenTox predictions
- Include MetaPrint2D site-of-metabolism predictions, other open models...
- Consensus modeling



# Future prospects

- Deploy Bioclipse models in OpenTox
  - Implement a bridge from OSGi services to OpenTox
- Run local server for offline OpenTox predictions
- Include OpenTox modeling in wizards (e.g. ToxCreate)



UPPSALA  
UNIVERSITET

# Experiences from the Bioclipse- OpenTox integration

- Open standards and ontology/API greatly simplify integration
- Open and collaborative community
- Encouraging and constructive feedback



UPPSALA  
UNIVERSITET

# Bioclipse on desktop, web, and mobile



RAP Startup Page

http://pele.farmbio.uu.se/demo/rap

Search View

Name

Keywords

File Edit View Atom Bond Tools Templates Help

Structure	Name	Keywords	Link
	Oxeladin	cough suppressant	DB04022

substructure search

No operations to display at this time.





UPPSALA  
UNIVERSITET

# Acknowledgement: The Chemistry Development Kit (CDK)

- Most prominent open source Java library for cheminformatics, 10 year anniversary!
- Coordinated from Uppsala/Stockholm (Egon Willighagen) and EBI (Christoph Steinbeck)
- ~150 citations
- Widely Used in academia and pharma industry
  - AstraZeneca, Bayer, Merck-Serono, Sanofi Aventis, Eli Lilly, Novartis



UPPSALA  
UNIVERSITET

# Genetta

SOFT



- Providing commercial support around the topics of Bioclipse and statistical modeling.
- Research, implementations, education, branding/deployment
- [www.genettasoft.com](http://www.genettasoft.com)



UPPSALA  
UNIVERSITET

# Acknowledgements

## **Dept. Pharmaceutical Biosciences, Uppsala University, Sweden**

Prof. Jarl E. S. Wikberg

*Dr. Egon Willighagen*

*Dr. Martin Eklund*

Jonathan Alvarsson

Carl Mäsak

Eskil Anderssen

Annsöfi Andersson

Arvid Berg

Bjarni Juliusson

Samuel Lampa

## **Unilever Centre for Molecular Informatics, Univ. Cambridge, UK**

Prof. Peter Murray-Rust

Prof. Robert Glen

Dr. Sam Adams

## **Linnaeus Centre for Bioinformatics, Uppsala, Sweden**

Dr. Erik Bongcam-Rudloff

Sofia Burvall

Dr. Alvaro Martinez Barrio

## **European Bioinformatics Institute, Hinxton, UK**

Dr. Christoph Steinbeck

Dr. Gilleain Torrence

Dr. Mark Rijnbeek

Stefan Kuhn

## **Cologne University Bioinformatics Centre (CUBIC), Germany**

Dr. Tobias Helmus

Dr. Christian Hoppe

Miguel Rojas

Thomas Kuhn

## **Dept Clinical Pharmacology, Uppsala University, Sweden**

Prof. Rolf Larsson

Dr. Claes Andersson

Hanna Göransson

## **Ludwig-Maximilians-Universität, Munich, Germany**

Dr. Johannes Wagener

## **Global Safety Assessment, AstraZeneca, Möndal Sweden**

*Dr. Scott Boyer*

*Dr. Lars Carlsson*

*Dr. Ernst Ahlberg Helgee*

## **Indiana University, IN, USA**

Dr. Rajarshi Guha

## **OpenTox**

*Dr. Barry Hardy*

*Dr. Nina Jeliaskova*

## **Others**

Dr. Jerome Pansanel

... and all others for contributing code,  
feature requests, bug reports,  
proposals, and comments!



UPPSALA  
UNIVERSITET

**OpenTox**   
OpenTox 2011

- Thank you -

**Ola Spjuth**

*ola.spjuth@farmbio.uu.se*

Department of Pharmaceutical Biosciences

Faculty of Pharmacy, Uppsala University

