



OpenTox Deliverable Report 6.2

# OpenTox Tutorials

Grant Agreement	Health-F5-2008-200787
Acronym	OpenTox
Name	An Open Source Predictive Toxicology Framework
Coordinator	Douglas Connect



Contract No.	Health-F5-2008-200787	
Document Type:	Deliverable Report	
WP/Task:	WP6	
Name	OpenTox Tutorials	
Document ID:	OpenTox WP6 D6.2 report	
Date:	Sept 30, 2010	
Status:	Final Version	
Organisation:	Douglas Connect	
Contributors	Roman Affentranger (Author)	DC
	Barry Hardy (Review & editing)	DC

Distribution:	Final version
---------------	---------------

Purpose of Document:	This document describes the initial OpenTox tutorials.
----------------------	--

Document History:	1 - RA (DC) authored
	2 - BH (DC) reviewed and edited final version

## Table of Contents

<b>Table of Contents</b> .....	<b>3</b>
<b>Table of Figures</b> .....	<b>5</b>
<b>Acknowledgements</b> .....	<b>7</b>
<b>Summary</b> .....	<b>8</b>
<b>1 Predict the Toxicity of a Compound</b> .....	<b>9</b>
1.1 Introduction .....	9
1.2 ToxPredict Tutorial A: Predict the Toxicity of a Chemical in the Database .....	9
1.3 ToxPredict Tutorial B: Predict the Toxicity of a New Chemical (not in the Database) .....	12
<b>2 Build a Predictive QSAR Model and Validate it</b> .....	<b>13</b>
2.1 Introduction .....	13
2.2 ToxCreate step 1: Create .....	13
2.3 ToxCreate Step 2: Inspect .....	14
2.4 ToxCreate Step 3: Predict .....	15
<b>3 Build a Model Based on <i>In Vitro</i> Data</b> .....	<b>16</b>
3.1 Introduction .....	16
3.2 Inspect US EPA's ToxCast Data .....	16
3.3 Select <i>in vitro</i> Assays based on Correlations to an <i>in vivo</i> Endpoint .....	16
3.4 Use the ToxCreate Web Application to Create a Model .....	17
3.4.1 Create the data file .....	17
3.4.2 Build the Model .....	18
3.4.3 Make a prediction .....	19
3.5 Create a model using the OpenTox API .....	20
3.5.1 Locate data on the ambit web service .....	20
3.5.2 Create a model on the command line using cURL .....	21
3.5.3 Make a prediction .....	21
<b>4 Validate your QSAR Model and Create a Report (for Developers)</b> .....	<b>23</b>
4.1 Introduction .....	23
4.2 Prerequisites .....	23
4.3 Validation Examples .....	23
4.3.1 Validate an algorithm on a dataset via training–test–split .....	23
4.3.2 Validate an algorithm on a dataset via bootstrapping .....	24
4.4 Validation Reports .....	24

4.4.1	Create validation report from validation.....	24
4.4.2	Create a QMRF Report .....	25
4.5	Further validation techniques.....	25
<b>5</b>	<b>Query and Access Toxicity Data .....</b>	<b>26</b>
5.1	Introduction .....	26
5.2	Inspect Data of the ISSMIC Dataset .....	26
5.3	Retrieve Data in Selected Formats .....	27
5.4	Find Similar Structures and Browse the Available Data for a Given Chemical.....	27
<b>6</b>	<b>Drug Discovery Predictive Toxicology Application I: Prioritizing compounds.....</b>	<b>28</b>
6.1	Introduction .....	28
6.2	Step 1: Predicting Oral Toxicity.....	28
6.3	Step 2: Analyse Cytotoxicities of the Cramer Class I compounds .....	30
6.4	Step 3: Predicting the Mutagenicity of the Selected Compounds.....	32
6.5	Step 4: Predicting Sites of Cytochrome P450 Metabolism .....	34
<b>7</b>	<b>Drug Discovery Predictive Toxicology Application II: Building a Model to Predict Kinase Inhibitor Activity .....</b>	<b>37</b>
7.1	Introduction.....	37
7.2	Activity B: Selecting a subset to create a model with ToxCreate.....	37
<b>8</b>	<b>Set up an OpenTox Data Service .....</b>	<b>40</b>
8.1	Introduction .....	40
8.2	Setting up a Windows System.....	41
8.2.1	Installing the Java Runtime Environment (JRE) and Java Development Kit (JDK) .....	41
8.2.2	Installing MySQL.....	41
8.2.3	Installing Tomcat .....	41
8.3	Setting Up a CentOS 5 Linux System .....	43
8.3.1	Installing MySQL.....	43
8.3.2	Installing the Java Runtime Environment and Development Kit.....	43
8.3.3	Installing Tomcat .....	43
8.4	Download and Deploy AMBIT 2.0 .....	48
<b>9</b>	<b>Conclusions .....</b>	<b>50</b>

## Table of Figures

<i>Figure 1 Starting page of ToxPredict.....</i>	<i>9</i>
<i>Figure 2 Structure entering by name, identifier, SMILES, etc. ....</i>	<i>10</i>
<i>Figure 3 The structures found for the search string are presented for verification.....</i>	<i>10</i>
<i>Figure 4 The list of endpoints currently available in ToxPredict. The user can select any number of endpoints. ....</i>	<i>10</i>
<i>Figure 5 The “Run” page indicates which computations are currently running, and which have completed already. ....</i>	<i>11</i>
<i>Figure 6 The results are displayed for each of the selected endpoints. ....</i>	<i>11</i>
<i>Figure 7 The structure upload tab of ToxPredict. ....</i>	<i>12</i>
<i>Figure 8 Starting page of ToxCreate ....</i>	<i>13</i>
<i>Figure 9 Specify a name for the model you are about to create. Unless you plan on choosing regression for the model building, i.e. for a classification model, a unit needs not be specified.....</i>	<i>13</i>
<i>Figure 10 How to format a ToxCreate training data set in Excel ....</i>	<i>14</i>
<i>Figure 11 How to format a ToxCreate training data set as CSV file ....</i>	<i>14</i>
<i>Figure 12 The “Inspect” tab of ToxCreate provides information about the created model.....</i>	<i>15</i>
<i>Figure 13 Structures can be drawn or searched for in ToxCreate’s Predict tab.....</i>	<i>15</i>
<i>Figure 14 Extensive details are provided for predictions.....</i>	<i>15</i>
<i>Figure 15 The ToxCast correlation spreadsheet.....</i>	<i>17</i>
<i>Figure 16 Select the data of your chosen assay in one of the files containing the original data ....</i>	<i>17</i>
<i>Figure 17 Build your data file by combining assay data (column B) with chemical structure data (SMILES, column A).....</i>	<i>18</i>
<i>Figure 18 Starting page of ToxCreate ....</i>	<i>18</i>
<i>Figure 19 Results page of ToxCreate ....</i>	<i>19</i>
<i>Figure 20 The “Predict” tab of ToxCreate allows to run predictions using the created model.....</i>	<i>19</i>
<i>Figure 21 ToxCreate provides detailed results on the predictions run with the created model ....</i>	<i>20</i>
<i>Figure 22 Find your chosen endpoint in the OpenTox data service.....</i>	<i>20</i>
<i>Figure 23 Searching for the ISSMIC dataset on <a href="http://apps.ideaconsult.net:8080/ambit2/dataset">http://apps.ideaconsult.net:8080/ambit2/dataset</a>.....</i>	<i>26</i>
<i>Figure 24 Download icons for the various available formats ....</i>	<i>27</i>
<i>Figure 25 Searching form similar structures by clicking on the first magnifying glass icon. ....</i>	<i>27</i>
<i>Figure 26 Browse available data on a given compound using the “All available feature values” link. ....</i>	<i>27</i>
<i>Figure 27 Launching a predictive model on a chosen dataset URL from within the AMBIT data service.....</i>	<i>28</i>
<i>Figure 28 The results of the predictive model are delivered as a URL.....</i>	<i>29</i>
<i>Figure 29 Distribution of the Cramer classes in the TCAMS dataset, depicted using the OpenTox chart generation service.....</i>	<i>29</i>
<i>Figure 30 Browsing the results.....</i>	<i>30</i>
<i>Figure 31 Retrieving the feature URL.....</i>	<i>31</i>

*Figure 32 A column is added to the compounds with the feature\_uris[] directive ..... 31*

*Figure 33 Another column is added to the compounds with the feature\_uris[] directive ..... 32*

*Figure 34 Selecting the Benigni/Bossa rulebase for mutagenicity and carcinogenicity ..... 33*

*Figure 35 Search results enhanced with columns for carcinogenicity predictions ..... 34*

*Figure 36 Running SmartCYP on the malaria box ..... 35*

*Figure 37 Example SmartCYP output ..... 35*

*Figure 38 SmartCYP color code ..... 36*

*Figure 39 Example SOME output ..... 36*

*Figure 40 SOME color code ..... 36*

*Figure 41 The list of antimalarial datasets on <http://pirin.uni-plovdiv.bg:8080/malaria/dataset> ..... 37*

*Figure 42 Search results for “Ser/Thr protein kinase” on the TCAMS antimalarial dataset ..... 38*

*Figure 43 SourceForge page of AMBIT 2.0 ..... 40*

*Figure 44 The Tomcat icon in the Windows7 taskbar ..... 42*

*Figure 45 Configuring the Tomcat startup type ..... 42*

*Figure 46: Tomcat “home” page ..... 47*

*Figure 47: Tomcat “manager” page ..... 48*

*Figure 48: The AMBIT 2.0 page ..... 49*

## Acknowledgements

### Research Funding

OpenTox – An Open Source Predictive Toxicology Framework, [www.opentox.org](http://www.opentox.org), is funded under the EU Seventh Framework Program: HEALTH-2007-1.3-3 Promotion, development, validation, acceptance and implementation of QSARs (Quantitative Structure-Activity Relationships) for toxicology, Project Reference Number Health-F5-2008-200787 (2008-2011).

### Project Partners

Douglas Connect (DC), In Silico Toxicology (IST), Ideaconsult (IDEA), Istituto Superiore di Sanita' (ISS), Technical University of Munich (TUM), Albert Ludwigs University Freiburg (ALU), National Technical University of Athens (NTUA), David Gallagher (DG), Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences (IBMC), Seascope Learning (SL), Jawaharlal Nehru University (JNU), Fraunhofer Institute for Toxicology & Experimental Medicine (ITEM).

### Advisory Board

European Centre for the Validation of Alternative Methods, European Joint Research Centre, U.S Environmental Protection Agency, U.S. Food & Drug Administration, Nestlé, Roche, AstraZeneca, Lhasa, Leadscope, University of North Carolina, Pharmatropé, Bioclipse, EC Environment Directorate General, Organisation for Economic Co-operation & Development, CADASTER, Bayer Healthcare.

### Correspondence

Dr. Barry Hardy, OpenTox Project Coordinator and Director, Community of Practice & Research Activities, Douglas Connect, Baermeggenweg 14, 4314 Zeiningen, Switzerland

Email: [barry.hardy -\(at\)- douglasconnect.com](mailto:barry.hardy-(at)-douglasconnect.com)

## Summary

This document represents a collection of tutorial materials on several OpenTox topics including walk-throughs of the two end user prototype applications ToxPredict and ToxCreate, illustration of the use of validation and reporting services applied to predictive toxicology models, the application of OpenTox facilities in a drug discovery workflow, and detailed instructions on how to get a system set up to host an OpenTox data service.

The tutorial example on the prototype OpenTox application ToxPredict ([www.toxpredict.org](http://www.toxpredict.org)) accepts chemical structures and names as input from the user and generates toxicity reports based on various pre-calibrated toxicity models and existing toxicity data.

In the ToxCreate ([www.toxcreate.org](http://www.toxcreate.org)) tutorial, the user provides a dataset of chemical structures and target variable data. ToxCreate subsequently builds and validates a Quantitative Structure-Activity Relationship (QSAR) predictive toxicology model. The user receives a reporting on details of model results and model predictions which they may examine, and including using the model for new predictions.

In the *in vitro* data model building tutorial, a predictive model is built based on *in vitro* data using OpenTox web services. Several models can be built and inspected based on application to the US EPA ToxCast dataset.

The tutorial on web validation and reporting web services, which are also behind the end user applications ToxPredict and ToxCreate, shows how cURL calls can be used to validate a predictive model or an algorithm. A number of different validation methods are used, including K-fold split, training-test-split and bootstrapping. Furthermore, QMRF reports are generated and visualized using the QMRF Editor web start application.

The objective of the ISSMIC data analysis tutorial is to illustrate searching facilities and data visualization tools in the OpenTox framework, specifically in the context of *in vivo* micronucleus mutagenicity assays contained within ISSMIC, a curated database, containing critically-selected information on chemical compounds tested with the assay.

A tutorial example of a predictive toxicology application in drug discovery is provided using the data on anti-malarial compounds made available at the ChEMBL Neglected Tropical Disease (NTD) archive ([www.ebi.ac.uk/chemblntd/](http://www.ebi.ac.uk/chemblntd/)). The anti-malarial compounds are prioritized based on a strongly conservative model for predicting oral toxicity. Experimentally-determined cytotoxicities against human cells of the compounds predicted to be safe are further examined, and their mutagenicities predicted. Sites of cytochrome P450 metabolism are predicted for selected compounds with no mutagenicity alerts, low human cytotoxicity, but high anti-malarial activity.

A tutorial is provided to guide the user through the setup of an OpenTox data service based on the download of the AMBIT software and its subsequent installation either on Windows or Linux.

All tutorials and their updates are made available online under [www.opentox.org/tutorials](http://www.opentox.org/tutorials).



## 1 Predict the Toxicity of a Compound

### 1.1 Introduction

The objective of these tutorials is to demonstrate the prototype OpenTox application ToxPredict ([www.toxpredict.org](http://www.toxpredict.org)) that accepts chemical structures and names as input and automatically generates a toxicity report based on various precalibrated toxicity models.



Figure 1 Starting page of ToxPredict

ToxPredict provides a web-based interface for predicting the toxicity of individual chemicals. Users can either search for a compound in the OpenTox prototype database, which currently includes quality-labelled data for 163,122 chemicals (including all chemicals currently registered with REACH) grouped in 2,409 datasets, or can upload their own chemical structure in SDF format. ToxPredict runs the selected calculations automatically using a collection of distributed computational services. Its initial version currently includes eighteen validated models addressing 15 different endpoints, and is extensible.

ToxPredict has been designed to be easy-to-use and self-explanatory. Therefore, the need for extensive tutorials should not be great, and the examples presented here are kept brief.

### 1.2 ToxPredict Tutorial A: Predict the Toxicity of a Chemical in the Database

ToxPredict searches its database for chemical structures entered as chemical names, SMILES strings, CAS numbers or via an integrated 2D chemical structure drawing editor. The following example illustrates using the chemical name.

1. Using a web browser, navigate to the starting page of [www.toxpredict.org](http://www.toxpredict.org) (see Figure 1)
2. Click "NEXT"
3. Type a chemical name in the text box, e.g. "benzene" (or CAS number, SMILES string, ...) (see Figure 2)

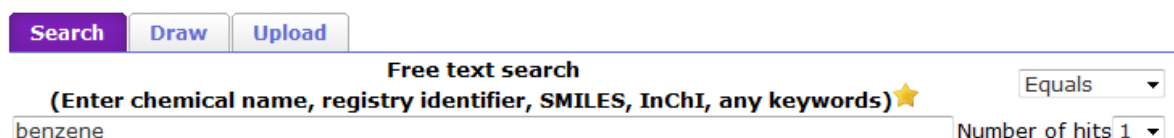
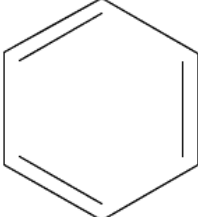


Figure 2 Structure entering by name, identifier, SMILES, etc.

- Click "NEXT" (to see if the search was successful. If not, go back to step 3) (see Figure 3)

« Page 1 Records per page 1 » [Structure\(s\) & Experimental Data](#) [SDF](#) [CSV](#) [PDF](#)

1.



**CASRN 71-43-2**  
**Synonym(s)** benzene,(6)annulene; benzine; Benzol; Benzolene; bicarburet of hydrogen; carbon oil; Coal naphtha; cyclohexatriene; mineral naphtha; motor benzol; nitration benzene; Phene; Phenyl hydride; pyrobenzol.,Benzene,BENZENE,C6H6  
**EINECS** 200-753-7  
**IUPAC name** benzene  
**InChIKey\_std** UHOVQNZJYSORNB-UHFFFAOYSA-N  
**InChI\_std** InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H  
**REACHRegistrationDate** 30.11.2010  
**SMILES** c1ccccc1,C1=CC=CC=C1

Figure 3 The structures found for the search string are presented for verification

- Click "NEXT" again (to view the list of endpoints), given the search in step 4 was successful
- Click to select any of the 18 boxes in the left column (multiple selections are possible) (see Figure 4)

Select
+

Welcome, [guest](#)  
[Admin](#)  
[Help](#)

**ToxPredict**

OpenTox demo application

1. Select structure(s)

2. Verify structure(s)

3. Select model(s)



4. Run prediction(s)

5. Display result(s)

NEXT

Model	Endpoint	Algorithm	Validation
<input checked="" type="checkbox"/> MolecularWeight		MolecularWeight	
<input checked="" type="checkbox"/> ToxTree: Verhaar scheme for predicting toxicity mode of action	Acute toxicity to fish (lethality)	ToxTree: Verhaar scheme for predicting toxicity mode of action	
<input checked="" type="checkbox"/> ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	Carcinogenicity	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	
<input checked="" type="checkbox"/> pKa	Dissociation constant (pKa)	pKa	
<input checked="" type="checkbox"/> ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	
<input checked="" type="checkbox"/> ToxTree: Michael acceptors	Endpoints	ToxTree: Michael acceptors	
<input checked="" type="checkbox"/> ToxTree: Eye irritation	Eye irritation/corrosion	ToxTree: Eye irritation	
<input checked="" type="checkbox"/> Caco-2 Cell Permeability <a href="http://www.ncbi.nlm.nih.gov/pubmed/16959190">http://www.ncbi.nlm.nih.gov/pubmed/16959190</a>	Gastrointestinal absorption	Regression: Linear regression	Model validation report
<input checked="" type="checkbox"/> OpenTox model created with TUM's PLRegression model learning web service.	Gastrointestinal absorption	<a href="http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/PLRegression">http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/PLRegression</a>	
<input checked="" type="checkbox"/> OpenTox model created with TUM's kNNregression model learning web service.	Gastrointestinal absorption	<a href="http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression">http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression</a>	
<input checked="" type="checkbox"/> <a href="http://opentox.ntua.gr:3000/model/l/679a80e6-b2d9-45c1-ba42-4489e85b5898">http://opentox.ntua.gr:3000/model/l/679a80e6-b2d9-45c1-ba42-4489e85b5898</a>	Gastrointestinal absorption	Multiple Linear Regression Training Algorithm	
<input checked="" type="checkbox"/> Lipinski Rule of Five	Human health effects	Lipinski Rule of Five	
<input checked="" type="checkbox"/> ToxTree: Cramer rules	Human health effects	ToxTree: Cramer rules	
<input checked="" type="checkbox"/> XLogP	Octanol-water partition coefficient (Kow)	XLogP	
<input checked="" type="checkbox"/> START biodegradation and persistence plug-in	Persistence: Biodegradation	START biodegradation and persistence plug-in	
<input checked="" type="checkbox"/> SmartCYP: Cytochrome P450-Mediated Drug Metabolism	Protein-binding	SmartCYP: Cytochrome P450-Mediated Drug Metabolism	
<input checked="" type="checkbox"/> ToxTree: Skin irritation	Skin irritation /corrosion	ToxTree: Skin irritation	
<input checked="" type="checkbox"/> ToxTree: Skin sensitisation alerts (M. Cronin)	Skin sensitisation	ToxTree: Skin sensitisation alerts (M. Cronin)	

Figure 4 The list of endpoints currently available in ToxPredict. The user can select any number of endpoints.

- Click "NEXT" to start the calculations. Calculations that are currently running are indicated with , those that have finished already with  (see also Figure 5)

Loading...

## ToxPredict

OpenTox demo application

Welcome, [guest](#) [Admin](#) [Help](#)

1. Select structure(s)
2. Verify structure(s)
3. Select model(s)
4. Run prediction(s)
5. Display result(s)

**NEXT**

Model	Endpoint	Algorithm	Validation	
<input checked="" type="checkbox"/> ToxTree: Verhaar scheme for predicting toxicity mode of action	Acute toxicity to fish (lethality)	ToxTree: Verhaar scheme for predicting toxicity mode of action		Completed ✓
<input checked="" type="checkbox"/> XLogP	Octanol-water partition coefficient (Kow)	XLogP		Processing
<input checked="" type="checkbox"/> Caco-2 Cell Permeability <a href="http://www.ncbi.nlm.nih.gov/pubmed/16959190">http://www.ncbi.nlm.nih.gov/pubmed/16959190</a>	Gastrointestinal absorption	Regression: Linear regression	Model validation report	Completed ✓
<input checked="" type="checkbox"/> ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	Carcinogenicity	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity		Completed ✓
<input checked="" type="checkbox"/> ToxTree: Michael acceptors	Endpoints	ToxTree: Michael acceptors		Completed ✓
<input checked="" type="checkbox"/> ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents		Completed ✓
<input checked="" type="checkbox"/> ToxTree: Skin irritation	Skin irritation /corrosion	ToxTree: Skin irritation		Processing
<input checked="" type="checkbox"/> ToxTree: Eye irritation	Eye irritation/corrosion	ToxTree: Eye irritation		Completed ✓
<input checked="" type="checkbox"/> ToxTree: Cramer rules	Human health effects	ToxTree: Cramer rules		Processing
<input checked="" type="checkbox"/> pKa	Dissociation constant (pKa)	pKa		Completed ✓

Figure 5 The "Run" page indicates which computations are currently running, and which have completed already.

8. Click "NEXT" again (to view the results) (see Figure 6)

Display

## ToxPredict

OpenTox demo application

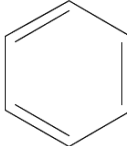
Welcome, [guest](#) [Admin](#) [Help](#)

1. Select structure(s)
2. Verify structure(s)
3. Select model(s)
4. Run prediction(s)
5. Display result(s)

« Page 1 Records per page 1 »

Structure(s) & Model predictions & Experimental Data SDF CSV PDF

1.



**CASRN 71-43-2**

**Synonym(s)** benzene,(6)annulene; benzine; Benzol; Benzolene; bicarburet of hydrogen; carbon oil; Coal naphtha; cyclohexatriene; mineral naphtha; motor benzol; nitration benzene; Phene; Phenyl hydride; pyrobenzol

**EINECS** 200-753-7

**IUPAC name** benzene

**InChIKey\_std** UH0VQNZJYSORNB-UHFFFAOYSA-N

**InChI\_std** InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H

**REACHRegistrationDate** 30.11.2010

**SMILES** c1ccccc1,C1=CC=CC=C1

benzene, (6)annulene; benzine; Benzol; Benzolene; bicarburet of hydrogen; carbon oil; Coal naphtha; cyclohexatriene; mineral naphtha; motor benzol; nitration benzene; Phene; Phenyl hydride; pyrobenzol, Benzene, BENZENE, C6H6

**OpenTox model created with TUM's PLSregression model learning web service.**

Prediction feature for <http://apps.ideaconsult.net:8080/ambit2/feature/22200>  
 endpoint prediction -4.496099948883057

**LipinskiFailures** **Lipinski Rule of Five**

LipinskiFailures 0.0

**Acute\_toxicity\_to\_fish\_lethality** **ToxTree: Verhaar scheme for predicting toxicity mode of action**  
 Verhaar scheme Class 1 (narcosis or baseline toxicity)

**Carcinogenicity** **ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity**

Structural Alert for genotoxic carcinogenicity NO

Structural Alert for nongenotoxic carcinogenicity NO

Potential S. typhimurium TA100 mutagen based on QSAR NO

Unlikely to be a S. typhimurium TA100 mutagen based on QSAR NO

Potential carcinogen based on QSAR NO

Unlikely to be a carcinogen based on QSAR NO

For a better assessment a QSAR calculation could be applied. NO

Negative for genotoxic carcinogenicity YES

Negative for nongenotoxic carcinogenicity YES

Structural Alert for genotoxic carcinogenicity#explanation

QA1.Acyl halides **No**

QA2.Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid **No**

QA3.N-methylol derivatives **No**

QA4.Monohaloalkene **No**

QA5.S or N mustard **No**

QA6.Propiolactones and propiolactones **No**

QA7.Epoxides and aziridines **No**

QA8.Aliphatic halogens **No**

Done

Figure 6 The results are displayed for each of the selected endpoints.

### 1.3 ToxPredict Tutorial B: Predict the Toxicity of a New Chemical (not in the Database)

To predict the toxicity of a new chemical that is not in the current OpenTox data infrastructure, you will need to upload the structure file as an SDF ("SD File") as follows:

1. Using a web browser, navigate to [www.toxpredict.org](http://www.toxpredict.org) (see Figure 1)
2. Click "NEXT"
3. Click the "Upload" tab (Figure 7)

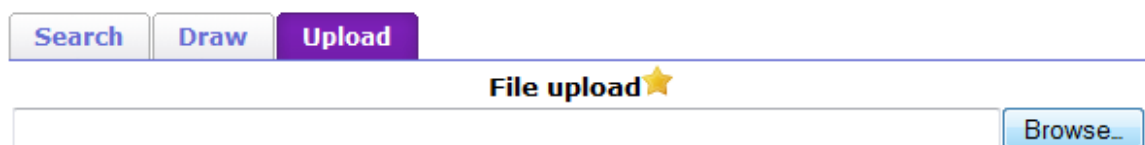


Figure 7 The structure upload tab of ToxPredict.

4. Click "Browse..."
5. Navigate to and select your SD file
6. Click "Open"
7. Click "NEXT" (to validate your input structure, a 2D drawing of the structure is presented, see Figure 3)
8. Click "NEXT" again (to view the list of endpoints, see Figure 4)
9. Click to select any of the 18 boxes in the left column (multiple selections are possible, see Figure 4)
10. Click "NEXT" to start the calculations (see Figure 5)
11. Click "NEXT" again (to view the results, see Figure 6)

## 2 Build a Predictive QSAR Model and Validate it

### 2.1 Introduction

The objective of this tutorial is to demonstrate the prototype OpenTox application ToxCreate ([www.toxcreate.org](http://www.toxcreate.org)). Based on a provided dataset, ToxCreate builds and validates a Quantitative Structure–Activity Relationship (QSAR) model.

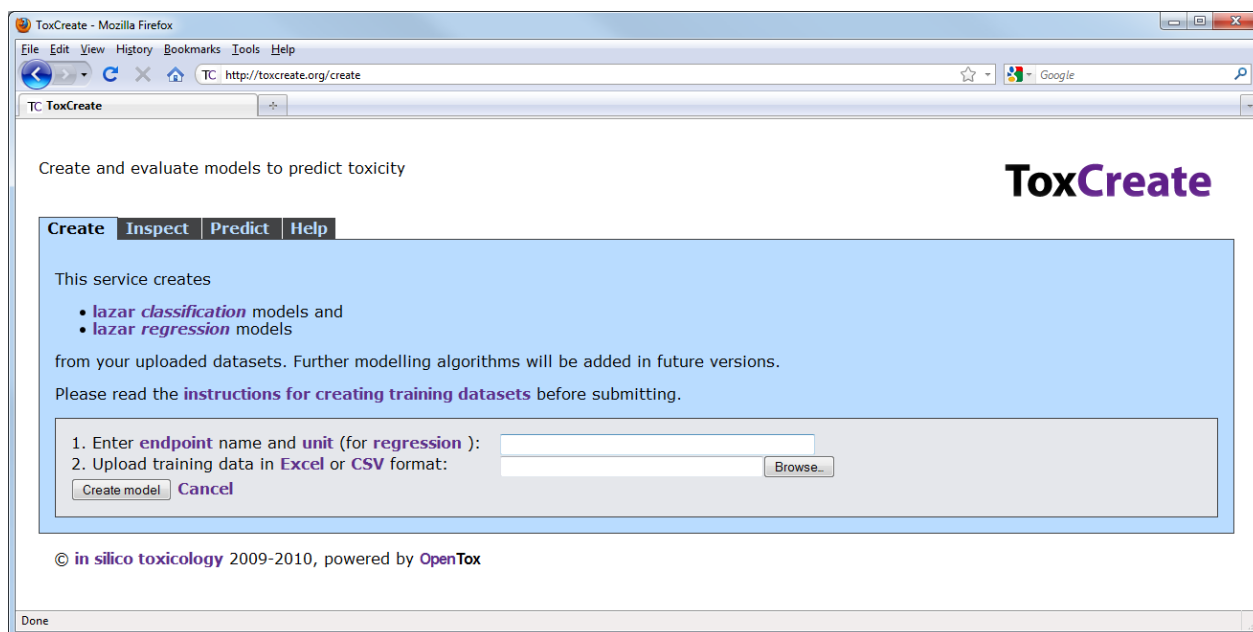


Figure 8 Starting page of ToxCreate

Similar to ToxPredict, ToxCreate has been designed to be easy-to-use and self-explanatory. Therefore, the need for extensive tutorials should now be necessary, and the example presented here is kept fairly brief. The requirements for this tutorial are: a web browser with Java script activated, Java installed and a working internet connection.

### 2.2 ToxCreate step 1: Create

1. Navigate to the ToxCreate starting page at [www.toxcreate.org](http://www.toxcreate.org) (see Figure 8) and follow the instructions on the web page. Click on bold and purple text to get further information on any topic.
2. Specify a name for the endpoint for which you plan to create a QSAR model. Specify the unit of your input data if you want to use regression for model building. (Figure 9)

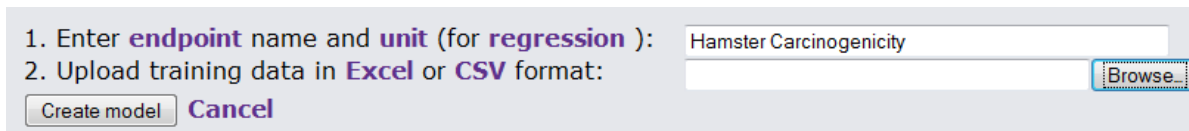


Figure 9 Specify a name for the model you are about to create. Unless you plan on choosing regression for the model building, a unit need not be specified, e.g. for a classification model,

3. Upload training data in Excel or CSV format (make sure you prepare your training data file according to the rules given at [www.toxcreate.org/help/](http://www.toxcreate.org/help/) (see also Figure 10 and Figure 11).

	A	B
1	<chem>CC(=O)Nc1ccc(O)cc1</chem>	1
2	<chem>O=c1[nH]cnc2[nH]ncc12</chem>	1
3	<chem>CCCCNc1cc(cc(c1O)c2ccccc2)S(=O)(=O)N)C(=O)O</chem>	1
4	<chem>CC(C)(C)NCC(O)COc1cccc2NC(=O)CCc12</chem>	1
5	<chem>CN(C)CCCC1(OCc2cc(C#N)ccc21)c3ccc(F)cc3</chem>	1
6	<chem>CCC(CC)CCN1C(=O)CN=C(C2CCCC2F)c3cc(Cl)ccc13</chem>	0
7	<chem>CCN(CC)CC(=O)Nc1c(C)cccc1C</chem>	0
8	<chem>CC(C)(C)NCC(O)COc1cccc2CC(O)C(O)Cc12</chem>	0
9	<chem>CN1CCCC1c2ccnc2</chem>	0

Figure 10 How to format a ToxCreat training data set in Excel

```

CC(=O)Nc1ccc(O)cc1, 1
O=c1[nH]cnc2[nH]ncc12, 1
CCCCNc1cc(cc(c1O)c2ccccc2)S(=O)(=O)N)C(=O)O, 1
CC(C)(C)NCC(O)COc1cccc2NC(=O)CCc12, 1
CN(C)CCCC1(OCc2cc(C#N)ccc21)c3ccc(F)cc3, 1
CCC(CC)CCN1C(=O)CN=C(C2CCCC2F)c3cc(Cl)ccc13, 0
CCN(CC)CC(=O)Nc1c(C)cccc1C, 0
CC(C)(C)NCC(O)COc1cccc2CC(O)C(O)Cc12, 0
CN1CCCC1c2ccnc2, 0
    
```

Figure 11 How to format a ToxCreat training data set as CSV file

- Click the "Create Model" button to start the model building and validation calculations

## 2.3 ToxCreat Step 2: Inspect

On this page you get a brief summary of all models with validation results (see Figure 12). Find your model by name and click on bold and purple links to view or download more detailed information (e.g. the feature dataset as an XML file, or a detailed validation report as a PDF document).

Create
Inspect
Predict
Help

This service is for testing purposes only - once a week all models will be deleted. Please send bug reports and feature requests to our [issue tracker](#).

Get an overview about ToxCreat models. This page is refreshed every 15 seconds to update the model status.

### dv\_gui\_multi\_cell

**Status:** Completed ( [delete](#) )

**Started:** 09/14/2010 - 04:08:57PM

**Training compounds:** 976

**Warnings:** [show](#)

**Algorithm:** lazar

**Type:** classification

**Descriptors:** [Fminer backbone refinement classes](#)

**Training dataset:** [Excel sheet](#), [YAML \(experts\)](#)

**Feature dataset:** [YAML \(experts, dataset too large for Excel\)](#)

**Model:** [YAML \(experts, models cannot be represented in Excel\)](#)

**Validation:** [show](#)

**Detailed report:** [show](#)

**Number of predictions:** 914

**Correct predictions:** 65.00 %

**Weighted area under ROC:** 0.627

**Specificity:** 0.833

**Sensitivity:** 0.500

**Confusion Matrix:**

		Measured	
		active	inactive
Predicted	active	245	71
	inactive	245	353

Figure 12 The “Inspect” tab of ToxCreat provides information about the created model

## 2.4 ToxCreat Step 3: Predict

1. Click on the “Predict” tab
2. On the “Predict” the user can draw a structure, or enter a name (or InChI, SMILES, CAS number) (see Figure 13).

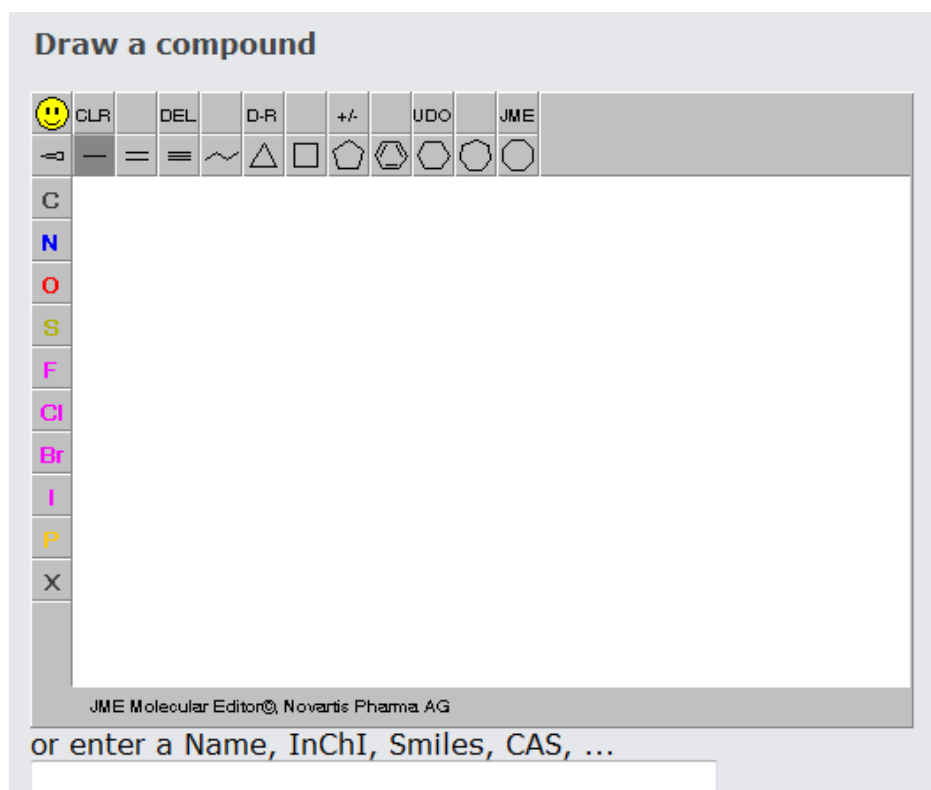


Figure 13 Structures can be drawn or searched for in ToxCreat’s Predict tab.

3. Select one or more prediction models (find the one just created by its name), and click on “Predict”
4. Check the results and click on “Details” for prediction details such as similar compounds, relevant substructures, etc. (see Figure 14).

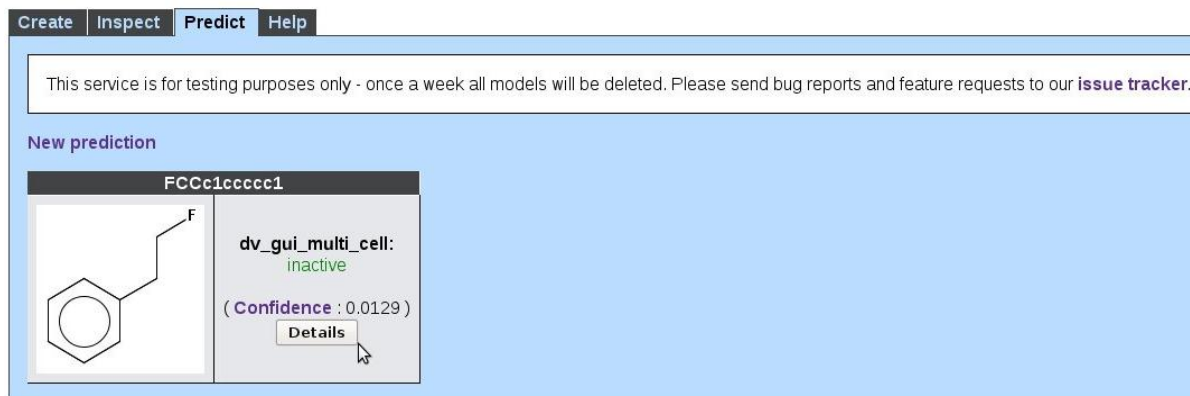


Figure 14 Extensive details are provided for predictions

5. Click on bold and purple topics to get further information.

### 3 Build a Model Based on *In Vitro* Data

#### 3.1 Introduction

The objective of this tutorial is to illustrate searching facilities, data representation, and the API of the OpenTox framework, specifically in the context of using *in vitro* data for building predictive toxicity models. In the tutorial, a predictive model is built based on *in vitro* data with OpenTox web services using a graphical user interface (GUI) as well as *via* the command line. Several models can be built and inspected.

For this tutorial the following software is needed:

- Microsoft Office (preferably version 2003 or later) is needed, or a similar office suite
- A web browser (e.g. Mozilla Firefox, [www.mozilla.com/firefox](http://www.mozilla.com/firefox))
- An important command-line tool used in this tutorial is called "cURL" ([curl.haxx.se](http://curl.haxx.se)). On many linux systems, cURL can be easily installed from a package repository using a standard package manager. It can also be downloaded from [curl.haxx.se/download.html](http://curl.haxx.se/download.html). Under Windows, there are two options for using cURL: 1) installing cURL natively, preferable using this version: [www.gknw.net/mirror/curl/win32/curl-7.21.1-devel-mingw32.zip](http://www.gknw.net/mirror/curl/win32/curl-7.21.1-devel-mingw32.zip), or 2) installing the VMWare Player ([www.vmware.com/products/player/](http://www.vmware.com/products/player/)) and running a small Linux environment ([www.maunz.de/opentox/dsl-4.1.zip](http://www.maunz.de/opentox/dsl-4.1.zip)) under Windows (after installing VMWare Player and unpacking the dsl-4.1.zip file, just double-click the dsl-4.1.vmx file).

In addition, the ToxCast Data Utilities need to be downloaded from [www.maunz.de/opentox/ToxCastDataUtils.zip](http://www.maunz.de/opentox/ToxCastDataUtils.zip). The archive contains data in a convenient Excel sheet form. The .zip archive needs to be expanded. It contains the files referenced in this tutorial.

#### 3.2 Inspect US EPA's ToxCast Data

Open the file "ToxCastDataUtils/10 Tong Liver Toxicity TDAS.pdf". The presentation addresses several aspects of the liver toxicity data in ToxCast ([www.epa.gov/ncct/ToxCast](http://www.epa.gov/ncct/ToxCast)). Take 15 minutes time to inspect the characteristics of the *in vitro* assay data.

#### 3.3 Select *in vitro* Assays based on Correlations to an *in vivo* Endpoint

Open the file "ToxCastDataUtils/ToxCastCorrelations\_20100708.xlsx". This Excel file analyses all correlations between *in vivo* endpoints and *in vitro* assays present in the ToxCast dataset (phase 1), for qualitative (nominal) as well as quantitative (continuous) values (see Figure 15).

1. Under **Assay\_2**, select a liver toxicity endpoint in Rat or Mouse, e.g. *CHR\_Rat\_LiverProliferativeLesions*.
2. Under **Assay\_1**, select an assay among the top correlated ones (red), e.g. *ATG\_PPARg\_trans*.

Note: You can find the original data in the files

- ToxCastDataUtils/ToxCastAssays\_20100708.xlsx (Assays, LOEL values)
- ToxCastDataUtils/ToxCastAssays\_discrete\_20100708.xlsx (Assays, discretized active/inactive)
- ToxCastDataUtils/ToxCastEndpoints\_20100708.xlsx (Endpoints, LOEL values)
- ToxCastDataUtils/ToxCastEndpoints\_discrete\_20100708.xlsx (Endpoints, discretized active/inactive)



A18		ATG_PPARG_TRANS													
A		B													
1	Assay_1	Assay_2	tp	fp	fn	tn	sens	spec	ba	accurac	relative	odds.ra	ppv	npv	p.valu
18	ATG_PPARG_TRANS	CHR_Rat_LiverProliferativeLesions	45	79	16	108	0,738	0,578	0,658	0,617	2,81	3,84	0,363	0,871	2,84
19	PS_Gene_PPARG	CHR_Rat_LiverProliferativeLesions	45	79	16	108	0,738	0,578	0,658	0,617	2,81	3,84	0,363	0,871	2,84
37	PS_KEGG_Thyroid_cancer	CHR_Rat_LiverProliferativeLesions	49	97	12	90	0,803	0,481	0,642	0,56	2,85	3,79	0,336	0,882	8,20
74	PS_Gene_PLAT	CHR_Rat_LiverProliferativeLesions	13	8	48	179	0,213	0,957	0,585	0,774	2,93	6,06	0,619	0,789	0,00
440	CLM_OxidativeStress_24hr	CHR_Rat_LiverProliferativeLesions	11	8	50	179	0,18	0,957	0,589	0,766	2,65	4,92	0,579	0,782	0,00
649	NVS_ADME_hCVP3A4	CHR_Rat_LiverProliferativeLesions	8	4	53	183	0,131	0,979	0,555	0,77	2,97	6,9	0,667	0,775	0,00
729	BSK_SM3C_Thrombomodulin_up	CHR_Rat_LiverProliferativeLesions	13	12	48	175	0,213	0,936	0,574	0,758	2,42	3,95	0,52	0,785	0,00
806	PS_PathwayCommons_0_NCI_NATURE_BMP_recept	CHR_Rat_LiverProliferativeLesions	34	63	27	124	0,557	0,663	0,61	0,637	1,96	2,48	0,351	0,821	0,00
982	BSK_BE3C_IL1a_down	CHR_Rat_LiverProliferativeLesions	9	6	52	181	0,148	0,968	0,558	0,766	2,69	5,22	0,6	0,777	0,00
1117	CLM_Hepat_Steatosis_24hr	CHR_Rat_LiverProliferativeLesions	19	25	42	162	0,311	0,866	0,589	0,73	2,1	2,93	0,42	0,794	0,00
1194	BSK_SM3C_MCP1_up	CHR_Rat_LiverProliferativeLesions	12	11	49	176	0,197	0,941	0,569	0,758	2,4	3,92	0,522	0,782	0,00
1195	CLM_Hepat_DNADamage_24hr	CHR_Rat_LiverProliferativeLesions	12	11	49	176	0,197	0,941	0,569	0,758	2,4	3,92	0,522	0,782	0,00
1224	NVS_NR_hPPARG	CHR_Rat_LiverProliferativeLesions	5	1	56	186	0,082	0,995	0,538	0,77	3,6	16,6	0,833	0,769	0,00
1257	BSK_BE3C_tPA_up	CHR_Rat_LiverProliferativeLesions	8	5	53	182	0,131	0,973	0,552	0,766	2,73	5,49	0,615	0,774	0,00
1508	CLZD_HMGCS2_48	CHR_Rat_LiverProliferativeLesions	9	7	52	180	0,148	0,963	0,555	0,762	2,51	4,45	0,562	0,776	0,00
1581	PS_Ingeniuty_ERKMAPK_Signaling	CHR_Rat_LiverProliferativeLesions	49	114	12	73	0,803	0,39	0,597	0,492	2,13	2,61	0,301	0,859	0,00
2809	PS_PathwayCommons_0_NCI_NATURE_IL1_mediati	CHR_Rat_LiverProliferativeLesions	33	66	28	121	0,541	0,647	0,594	0,621	1,77	2,16	0,333	0,812	0,00
2858	PS_PathwayCommons_CELL_MAP_AndrogenRecep	CHR_Rat_LiverProliferativeLesions	44	99	17	88	0,721	0,471	0,596	0,532	1,9	2,3	0,308	0,838	0,00
2859	PS_Ingeniuty_0_SAPKJK_Signaling	CHR_Rat_LiverProliferativeLesions	5	2	56	185	0,082	0,99	0,536	0,766	3,07	8,26	0,714	0,768	0,00
3052	PS_PathwayCommons_0_NCI_NATURE_Glypican_1	CHR_Rat_LiverProliferativeLesions	36	74	25	113	0,59	0,604	0,597	0,601	1,81	2,2	0,327	0,819	0,00
3053	PS_PathwayCommons_0_NCI_NATURE_Glypican_p	CHR_Rat_LiverProliferativeLesions	36	74	25	113	0,59	0,604	0,597	0,601	1,81	2,2	0,327	0,819	0,00
3054	PS_PathwayCommons_0_NCI_NATURE_Regulation	CHR_Rat_LiverProliferativeLesions	36	74	25	113	0,59	0,604	0,597	0,601	1,81	2,2	0,327	0,819	0,00
3055	PS_PathwayCommons_0_NCI_NATURE_Regulation	CHR_Rat_LiverProliferativeLesions	36	74	25	113	0,59	0,604	0,597	0,601	1,81	2,2	0,327	0,819	0,00
3056	PS_PathwayCommons_0_NCI_NATURE_TGF_beta	CHR_Rat_LiverProliferativeLesions	36	74	25	113	0,59	0,604	0,597	0,601	1,81	2,2	0,327	0,819	0,00
3191	PS_PathwayCommons_0_NCI_NATURE_IFN_gamma	CHR_Rat_LiverProliferativeLesions	37	78	24	109	0,607	0,583	0,595	0,589	1,78	2,15	0,322	0,82	0,00
3792	PS_PathwayCommons_0_NCI_NATURE_p38_MAPK	CHR_Rat_LiverProliferativeLesions	31	61	30	126	0,508	0,674	0,591	0,633	1,75	2,13	0,337	0,808	0,00
3793	PS_PathwayCommons_0_NCI_NATURE_Regulation	CHR_Rat_LiverProliferativeLesions	31	61	30	126	0,508	0,674	0,591	0,633	1,75	2,13	0,337	0,808	0,00
3794	PS_PathwayCommons_0_NCI_NATURE_TNF_recept	CHR_Rat_LiverProliferativeLesions	31	61	30	126	0,508	0,674	0,591	0,633	1,75	2,13	0,337	0,808	0,00
3800	cyclohexene_oxime	CHR_Rat_LiverProliferativeLesions	3	0	58	187	0,0492	1	0,525	0,766	4,22	0	1	0,763	0,00
3801	Alcohol_alkenyl_cyclic_alkyl	CHR_Rat_LiverProliferativeLesions	3	0	58	187	0,0492	1	0,525	0,766	4,22	0	1	0,763	0,00
4335	PS_Ingeniuty_0_NRF2_mediati_Oxidative_Stress	CHR_Rat_LiverProliferativeLesions	6	4	55	183	0,0984	0,979	0,538	0,762	2,6	4,99	0,6	0,769	0,00
4380	PS_PathwayCommons_0_NCI_NATURE_Endothelin	CHR_Rat_LiverProliferativeLesions	34	70	27	117	0,557	0,626	0,592	0,609	1,74	2,1	0,327	0,812	0,00

Figure 15 Correlation spreadsheet analysis of ToxCast dataset

### 3.4 Use the ToxCreat Web Application to Create a Model

#### 3.4.1 Create the data file

Fill the data into an Excel Sheet: open an original data file, such as

ToxCastDataUtils/ToxCastAssays\_discrete\_20100708.xlsx. Select the column with your assay. (see Figure 16).

Figure 16 Select the data of your chosen assay in one of the files containing the original data

Create a new Excel file and insert the column into the first sheet, second column. Remove the first row

(header). Open ToxCastDataUtils/ToxCast\_Chemicals\_20100708.xlsx and select column E (SMILES codes). Copy it to the first column of the first sheet of your new workbook (Figure 17).



Get an overview about ToxCreate models. This page is refreshed every 15 seconds to update the model status.

### ATG\_PPARG\_trans

**Status:** Completed ( [delete](#) )  
**Started:** 09/13/2010 - 10:20:28AM  
**Training compounds:** 320  
**Warnings:** [show](#)  
**Algorithm:** [lazar](#)  
**Type:** [classification](#)  
**Descriptors:** [Fminer backbone refinement classes](#)  
**Training dataset:** [Excel sheet](#) , [RDF/XML \(experts\)](#) , [YAML \(experts\)](#)  
**Feature dataset:** [RDF/XML](#) , [YAML \(experts, dataset too large for Excel\)](#)  
**Model:** [RDF/XML](#) , [YAML \(experts, models cannot be represented in Excel\)](#)

**Validation:**  
**Detailed report:** [show](#)  
**Number of predictions:** 288  
**Correct predictions:** 65.00 %  
**Weighted area under ROC:** 0.664  
**Specificity:** 0.675  
**Sensitivity:** 0.635  
**Confusion Matrix:**

		Measured	
		active	inactive
Predicted	active	87	49
	inactive	50	102

Figure 19 Results page of ToxCreate

The model is validated by 10-fold crossvalidation. Click on “Show” next to “Detailed report” for more detailed information provided by the OpenTox reporting service (see Figure 19). When you are done inspecting the detailed report, select the “Predict” tab.

### 3.4.3 Make a prediction

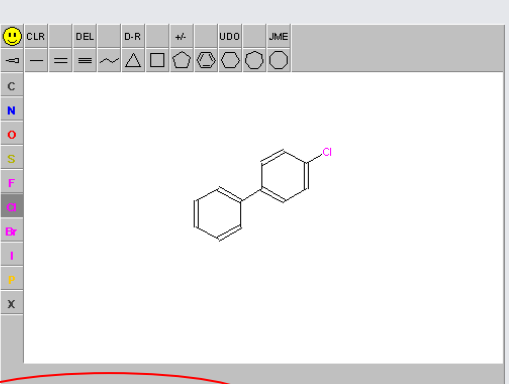
Draw a chemical or identify it in the text box. Hit predict.

Create Inspect **Predict** Help

This service is for testing purposes only - once a week all models will be deleted. Please send bug reports and feature requests to our [issue tracker](#).

Use this service to obtain predictions from OpenTox models.

Draw a compound



or enter a Name, InChI, Smiles, CAS, ...  
 Choose one or more prediction models  
 ATG\_PPARG\_trans  
 Hamster

Figure 20 The “Predict” tab of ToxCreate allows to run predictions using the created model

On the next page, view the prediction results and related details on confidence (not a probability), nearest neighbours and relevant substructures (see Figure 21).

Create Inspect **Predict** Help

This service is for testing purposes only - once a week all models will be deleted. Please send bug reports and feature requests to our [issue tracker](#).

New prediction

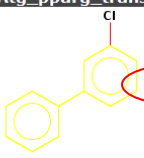
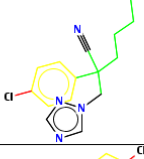
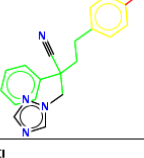

Atg_pparg_trans	Prediction	Confidence	Supporting information
 inactive		0.0415	Names and synonyms Significant fragments
Neighbors (1-5/16) next	Measured activity	Similarity	Supporting information
 inactive		0.517	Names and synonyms Significant fragments
 inactive		0.483	Names and synonyms Significant fragments
			

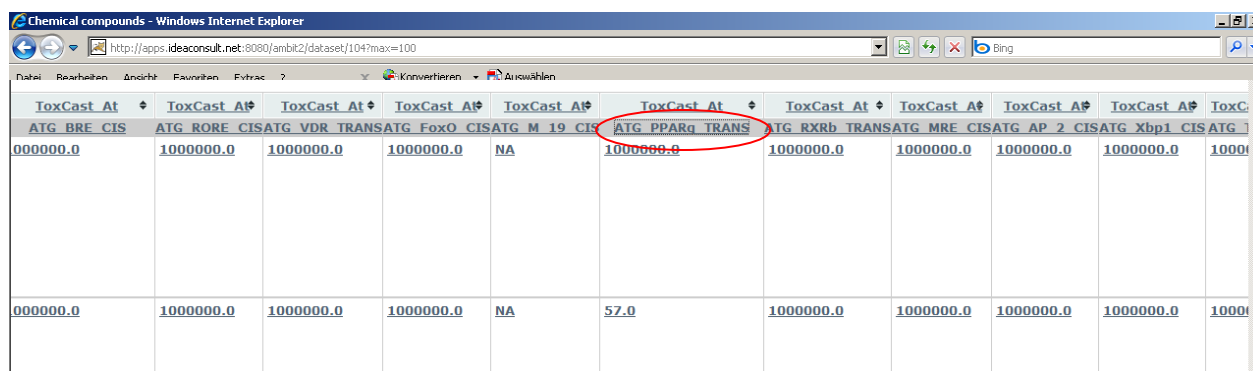
Figure 21 ToxCreat provides detailed results on the predictions run with the created model

### 3.5 Create a model using the OpenTox API

#### 3.5.1 Locate data on the ambit web service

The ToxCast data can be found online at [apps.ideaconsult.net:8080/ambit2/dataset/](http://apps.ideaconsult.net:8080/ambit2/dataset/).

- Search for the datasets analysed in step 2 (see section 3.2). Assay providers are spelled out here and are easily recognizable: ATG = Attagene, BSK = Bioseek, CLM = Cellumen, ... For example, the dataset by ATG is available at [apps.ideaconsult.net:8080/ambit2/dataset/104](http://apps.ideaconsult.net:8080/ambit2/dataset/104)
- Find the endpoint, e.g. *ATG\_PPARG\_trans* for ATG, in the columns. In OpenTox terms, it is represented as a feature. Point with the mouse to the columns header: The address is [apps.ideaconsult.net:8080/ambit2/feature/22333](http://apps.ideaconsult.net:8080/ambit2/feature/22333)



ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At	ToxCast At
ATG BRE CIS	ATG RORE CISATG	VDR	TRANSATG	FoxO M 19 CIS	ATG_PPARG_TRANS	ATG RXRb	TRANSATG	MRE	CISATG AP 2	CISATG Xbp1
000000.0	1000000.0	1000000.0	1000000.0	NA	1000000.0	1000000.0	1000000.0	1000000.0	1000000.0	1000000.0
000000.0	1000000.0	1000000.0	1000000.0	NA	57.0	1000000.0	1000000.0	1000000.0	1000000.0	1000000.0

Figure 22 Find your chosen endpoint in the OpenTox data service

### 3.5.2 Create a model on the command line using cURL

Open a command shell (under Windows, you will find it under “All Programs → Accessories → Command Prompt”). We will now learn a  $k$ -nearest neighbour model for the dataset from the above process. First, we will gather some information about the model (if you are using the Windows Command Prompt, replace “curl” in the following command lines by the path to the cURL executable on your system, e.g.

C:\Users\Username\Programs\curl-7.21.0-win64\curl.exe):

```
curl http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression
```

The output is rdf/xml, the OpenTox standard data format. This rich representation contains links to formal logical concepts and definitions, making the OpenTox framework explicit to machines. As a consequence, this enables logic inference and reasoning.

For example, the description tells you that this model is an exposure of WEKA’s knn-model. Following that, it lists the parameters that the model can be configured with, e.g. whether crossvalidation should be used to select parameter  $k$ , dataset uri, nearest neighbour search algorithm, etc.

We will now feed it the data we just selected. Execute the following code (replace with your own data):

```
curl -X POST
-d "dataset_uri=http://apps.ideaconsult.net:8080/ambit2/dataset/104"
-d "prediction_feature=http://apps.ideaconsult.net:8080/ambit2/feature/22333"
http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression
```

Since model generation can take some time, this will return a URI for the running task. Use it to obtain the model URI, once it has finished (just keep trying).

```
curl -i -H "accept: text/uri-list"
http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/task/<task-id>
```

Once you have the model URI, locate the compound we had in the GUI example. To do so, direct your browser to <http://apps.ideaconsult.net:8080/ambit2/query/similarity?search=c1cccc1Oc2cccc2&threshold=0.9> and search for your structure (here: “c1c2ccc(c1cccc1)cc2”, which you can find at <http://apps.ideaconsult.net:8080/ambit2/compound/16479>). If your structure is not already there select the most similar one.

### 3.5.3 Make a prediction

Use the model URI to predict the compound:

```
curl -i -X POST
-d "dataset_uri=http://apps.ideaconsult.net:8080/ambit2/compound/16479"
http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/model/TUMOpenToxModel_kNN_98
```

Once again, you will be returned a task uri. Keep querying it to receive a dataset URI:

```
curl -i -H "accept: text/uri-list"
http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/task/<task-id>
```

This will return a dataset URI:

```
curl -i -H "accept: text/plain" http://apps.ideaconsult.net:8080/ambit2/dataset/2678
```

Note that the predicted value is 1000000.0, which corresponds to *inactive* (as in the GUI example).

The above code gave you plain text information for purposes of this tutorial. You can also try the OpenTox standard format `rdf/xml`:

```
curl -i -H "accept: rdf/xml" http://apps.ideaconsult.net:8080/ambit2/dataset/2678
```

## 4 Validate your QSAR Model and Create a Report (for Developers)

### 4.1 Introduction

The objective of this tutorial is to demonstrate how to handle the validation and reporting services for predictive models using cURL calls. The web validation and reporting web services that are also behind the end user applications ToxPredict and ToxCreate are contacted using cURL calls to validate a model or an algorithm. A number of different validation methods are used, such as K-fold split, training-test-split or bootstrapping. Furthermore, QMRF reports are generated and visualized using the QMRF Editor web start application. This tutorial is mostly aimed at developers who want to get a picture of how these services work in detail.

### 4.2 Prerequisites

It is probably of advantage to have the web page with the OpenTox API definitions for the validation services (<http://www.opentox.org/data/documents/development/validation/Validation/>) open in a web browser.

An important command-line tool used in this tutorial is called "cURL" (<http://curl.haxx.se>). On many linux systems, cURL can be easily installed from a package repository using a standard package manager. It can also be downloaded from <http://curl.haxx.se/download.html>. Under Windows, there are two options for using cURL: 1) installing cURL natively, preferable using this version: <http://www.gknw.net/mirror/curl/win32/curl-7.21.1-devel-mingw32.zip>, or 2) installing the VMWare Player (<http://www.vmware.com/products/player/>) and running a small Linux environment (<http://www.maunz.de/opentox/dsl-4.1.zip>) under Windows (after installing VMWare Player and unpacking the dsl-4.1.zip file, just double-click the dsl-4.1.vmx file).

Also needed is Java 6, with web start enabled:

(<http://www.oracle.com/technetwork/java/javase/downloads/index.html>).

### 4.3 Validation Examples

First we want to list all available validations. To do that, it is necessary, to execute the following command in a terminal window (under Windows, you will find it under "All Programs -> Accessories -> Command Prompt"; replace "curl" in the following command lines by the path to the cURL executable on your system, e.g.

C:\Users\Username\Programs\curl-7.21.0-win64\curl.exe):

```
curl http://opentox.informatik.uni-freiburg.de/validation
```

#### 4.3.1 Validate an algorithm on a dataset via training-test-split

This will create a new validation object. A model is constructed by splitting a dataset into two parts: one for learning a model and one for testing, i.e. predicting and estimating the performance of the constructed model. Splitting the dataset is done in random fashion. One can also define the ratio for splitting into training and test, the default is 67% training and 33% test.

```
curl -X POST
-d algorithm_uri="http://opentox.informatik.uni-freiburg.de/algorithm/lazar"
-d dataset_uri="http://opentox.informatik.uni-freiburg.de/dataset/1"
-d prediction_feature=
"http://localhost/toxmodel/feature%23Hamster%2520Carcinogenicity%2520(DSSTOX/CPDB)"
-d algorithm_params=
"feature_generation_uri=http://opentox.informatik.uni-freiburg.de/algorithm/fminer"
-d split_ratio=0.9 -d random_seed=2
http://opentox.informatik.uni-freiburg.de/validation/training_test_split
```



Validating algorithms can be a time consuming task. Therefore the result of the above cURL call is a task URI. To query the status of the task URI, enter the following command in the terminal (where the term <TASK-ID> should be replaced with the correct task ID).

```
curl http://opentox.informatik.uni-freiburg.de/task/<TASK-ID>
```

As soon as the task is completed, your validation is available. The validation URI can be found in the resultURI property of the task:

```
---
:uri: http://opentox.informatik.uni-freiburg.de/task/<id>
:hasStatus: Completed
:resultURI: http://opentox.informatik.uni-freiburg.de/validation/<VALIDATION-ID>
[...]
```

Use cURL to get a closer look at your validation result:

```
curl http://opentox.informatik.uni-freiburg.de/validation/<VALIDATION-ID>
```

Just like the task result, the validation result is formatted in YAML, a markup language that is human readable. You could look at the statistics such as area-under-roc or confusion matrix values.

### 4.3.2 Validate an algorithm on a dataset via bootstrapping

Bootstrapping is a machine learning technique that splits a dataset into training and test set via "sampling with replacement".

```
curl -X POST
-d algorithm_uri="http://opentox.informatik.uni-freiburg.de/algorithm/lazar"
-d dataset_uri="http://opentox.informatik.uni-freiburg.de/dataset/1"
-d prediction_feature=
"http://localhost/toxmodel/feature%23Hamster%2520Carcinogenicity%2520(DSSTOX/CPDB)"
-d algorithm_params=
"feature_generation_uri=http://opentox.informatik.uni-freiburg.de/algorithm/fminer"
-d random_seed=2
http://opentox.informatik.uni-freiburg.de/validation/bootstrapping
```

Again, this cURL call returns a task. As soon as the bootstrapping validation is finished, you validation is provided as before.

## 4.4 Validation Reports

Validation reports present validation results in a nice human readable format. This cURL call gives you a list of available validation reports:

```
curl http://opentox.informatik.uni-freiburg.de/validation/report/validation
```

### 4.4.1 Create validation report from validation

This cURL call will create a report for the validation that you just performed before. Choose which validation you like (training-test split or bootstrapping).

```
curl -X POST
-d validation_uris=
"http://opentox.informatik.uni-freiburg.de/validation/<VALIDATION-ID>"
http://opentox.informatik.uni-freiburg.de/validation/report/validation
```

A report is created that is wrapped in a task URI as above.

You can access your report in YAML-format with the following cURL call (this time you have to specify YAML as requested format manually, as the default report format is 'text/html')



```
curl -H "accept:application/x-yaml"
http://opentox.informatik.uni-freiburg.de/validation/report/validation/<REPORT-ID>
```

You can also view this report via a web browser, where connected information for this validation object is available. You should initiate a web browser, open a new tab and simply enter

<http://opentox.informatik.uni-freiburg.de/validation/report/validation/<REPORT-ID>>

in the address line of the browser.

#### 4.4.2 Create a QMRF Report

QMRF (QSAR Model Reporting Format) is a harmonized template supported by the European Commission Joint Research Centre (EC JRC) for summarizing and reporting key information on QSAR models.

A QMRF is created for a particular QSAR model. To this end, you can build a model on the complete dataset we have been using so far with the following cURL call:

```
curl -d dataset_uri="http://opentox.informatik.uni-freiburg.de/dataset/1"
-d prediction_feature=
"http://localhost/toxmodel/feature%23Hamster%2520Carcinogenicity%2520(DSSTOX/CPDB)"
-d feature_generation_uri="http://opentox.informatik.uni-freiburg.de/algorithm/fminer"
http://opentox.informatik.uni-freiburg.de/algorithm/lazar
```

Use the new model to build a QMRF report via:

<<< [http://opentox.informatik.uni-freiburg.de/validation/report/validation/id\\_i](http://opentox.informatik.uni-freiburg.de/validation/report/validation/id_i)

```
curl -X POST
-d model_uri=http://opentox.informatik.uni-freiburg.de/model/<MODEL-ID>
http://opentox.informatik.uni-freiburg.de/validation/reach_report/QMRF
```

This report can be accessed via cURL:

```
curl http://opentox.informatik.uni-freiburg.de/validation/reach_report/QMRF/<REPORT-ID>
```

Alternatively, use the QMRF editor to edit this report by visiting the address with your browser:

[http://opentox.informatik.uni-freiburg.de/validation/reach\\_report/QMRF/<REPORT-ID>/editor](http://opentox.informatik.uni-freiburg.de/validation/reach_report/QMRF/<REPORT-ID>/editor)

#### 4.5 Further validation techniques

For additional technical description and further examples including:

- how to validate a model on a test dataset
- how to validate an algorithm on a training and test dataset
- how to create a validation object by comparing feature values
- how to validate an algorithm on a dataset via k-fold cross-validation

please consult the examples web page located at:

<http://opentox.informatik.uni-freiburg.de/validation/examples>

## 5 Query and Access Toxicity Data

### 5.1 Introduction

The objective of this tutorial is to illustrate searching facilities and data visualization tools in the OpenTox framework, specifically in the context of *in vivo* micronucleus mutagenicity assay in rodent's data of the ISSMIC database. ISSMIC is a curated database, containing critically-selected information on chemical compounds tested with the *in vivo* micronucleus mutagenicity assay in rodents. *In vivo* mutagenicity testing is ranked three (well ahead of rodent carcinogenicity) as an animal consuming experimentation. Results in bone marrow cells, peripheral blood cells and splenocytes for male/female rat/mouse are reported. The data are collected from publicly available datasets (Toxnet, NTP) and from the Leadscope FDA CRADA Toxicity Database. ISSMIC provides both biological calls and chemical structures, and is the basis for establishing sound read-across and QSAR risk assessment. The ISSMIC data is available from OpenTox data services.

### 5.2 Inspect Data of the ISSMIC Dataset

1. Open the following URL in your web browser: <http://apps.ideaconsult.net:8080/ambit2/dataset>
2. Enter "ISSMIC\_v2a\_151\_2Apr09" in the search box and press the "Search" button.

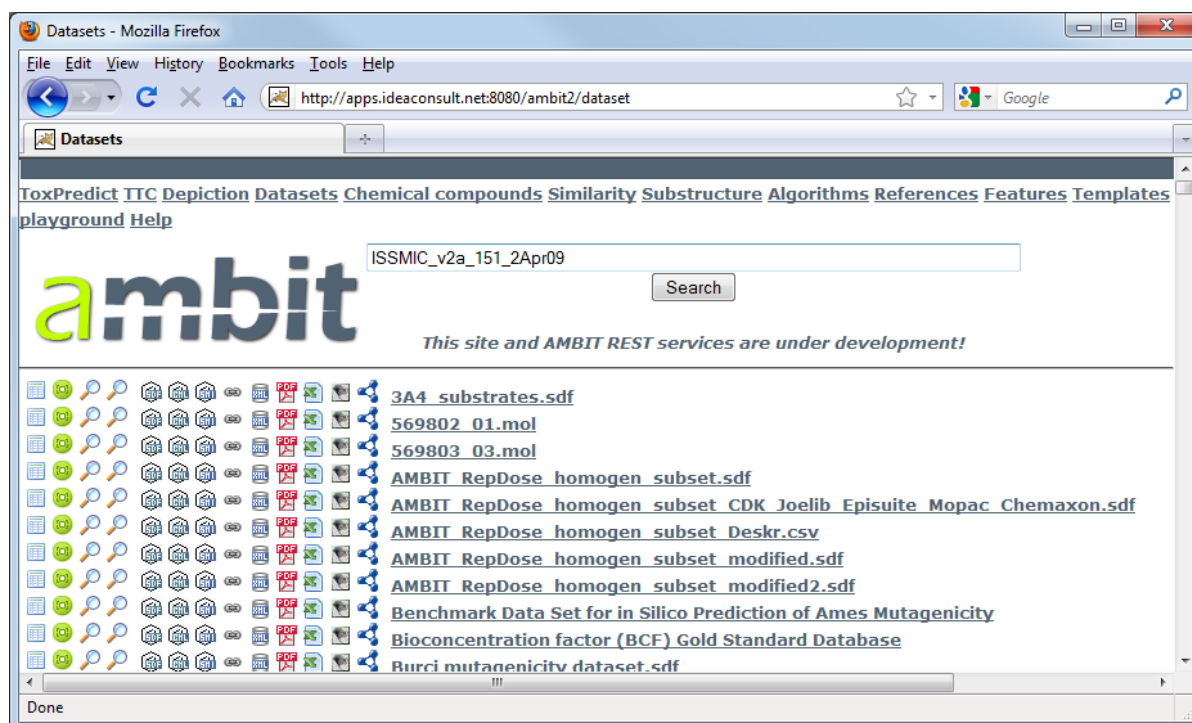


Figure 23 Searching for the ISSMIC dataset on <http://apps.ideaconsult.net:8080/ambit2/dataset>

3. Click on the returned URL and take your time to browse the information available in the ISSMIC dataset.
4. Click on a given property value of interest to retrieve a list of chemicals with similar property values

### 5.3 Retrieve Data in Selected Formats

Download and inspect a selected subset (or the entire dataset) in SDF, CML, SMILES, URI list, XML, PDF, CSV, plain text, ARFF and/or RDF format. These downloads are accessible through the following icons:



Figure 24 Download icons for the various available formats

### 5.4 Find Similar Structures and Browse the Available Data for a Given Chemical

Select a chemical of interest and perform a search for similar compounds (e.g. Tanimoto > 0.85) in the entire OpenTox database by clicking on the top magnifying glass icon on the left of the 2D drawing of the chemical (see Figure 25).

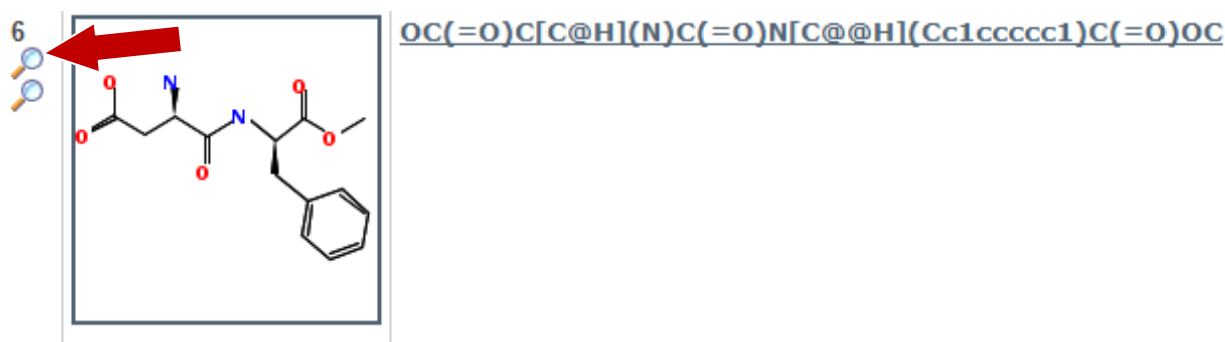
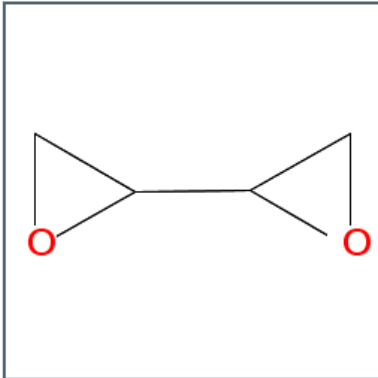


Figure 25 Searching for similar structures in the OpenTox database (by clicking on the first magnifying glass icon).

Alternatively, you can search the OpenTox database using your selected chemical as a substructure by clicking on the second magnifying glass icon.

Browse available data for a selected compound by clicking on its 2D drawing, and on the following page on the “All available feature values” link.



[CAS RN](#)  
[EINECS](#)  
[Chemical name\(s\)](#)  
[All available feature values](#)  
[Feature values by groups](#)  
[Feature values by dataset](#)  
[Features](#)  
[Model predictions](#)

Figure 26 Browse available data on a given compound using the “All available feature values” link.

## 6 Drug Discovery Predictive Toxicology Application I: Prioritizing compounds

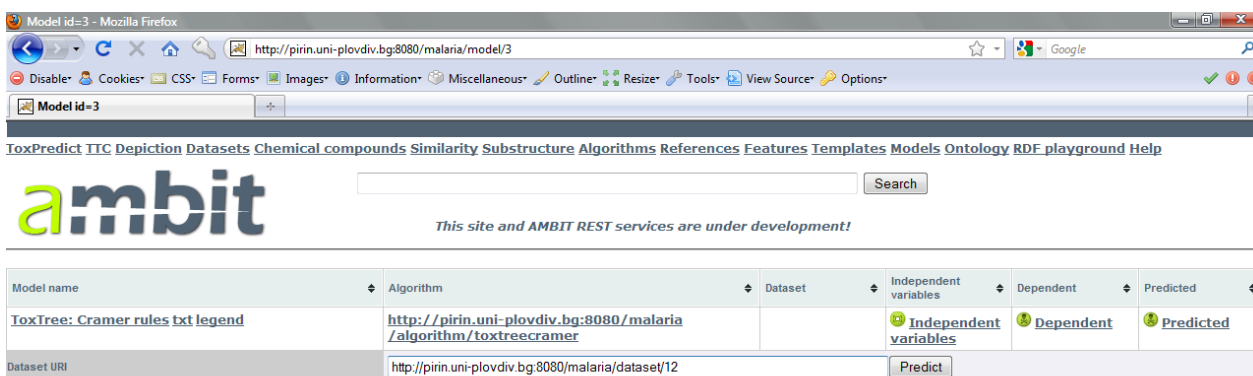
### 6.1 Introduction

An example of a predictive toxicology application in drug discovery is provided here using the data on antimalarial compounds made available at the ChEMBL Neglected Tropical Disease (NTD) archive ([www.ebi.ac.uk/chemblntd/](http://www.ebi.ac.uk/chemblntd/)). In this tutorial example, the antimalarial compounds are prioritized based on a strongly conservative model for predicting oral toxicity. Experimentally-determined cytotoxicities against human cells of the compounds predicted to be safe are further examined, and their mutagenicities predicted. Sites of cytochrome P450 metabolism are predicted for selected compounds with no mutagenicity alerts, low human cytotoxicity, but high anti-malarial activity.

### 6.2 Step 1: Predicting Oral Toxicity

Go to the list of antimalarial datasets at <http://pirin.uni-plovdiv.bg:8080/malaria/dataset>. We'll first predict oral toxicity for the TCAMS dataset. Start by clicking on the TCAMS dataset link. The URL in the browser should read <http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12>. You can browse the compounds (See Figure 27.)

In a new tab of your browser, go to the list of OpenTox models at <http://pirin.uni-plovdiv.bg:8080/malaria/model> (or follow the "Models" link at the top of the page listing the datasets). To predict oral toxicity we will use the "Toxtree Cramer rules" model. Clicking on the Cramer rules link will open its page. OpenTox models accept dataset URLs as input (instead of file names). Enter (or paste) the TCAMS URL ("<http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12>") into the text box. Click "Predict"



Model name	Algorithm	Dataset	Independent variables	Dependent	Predicted
<a href="#">ToxTree: Cramer rules txt legend</a>	<a href="http://pirin.uni-plovdiv.bg:8080/malaria/algorithm/toxtreecramer">http://pirin.uni-plovdiv.bg:8080/malaria/algorithm/toxtreecramer</a>		Independent variables	Dependent	Predicted
Dataset URI	<a href="http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12">http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12</a>		<input type="button" value="Predict"/>		

Figure 27 Launching a predictive model on a chosen dataset URL from within the AMBIT data service

which will launch calculations. You might click on the links to find out if the calculations are completed. When completed, clicking on the link will lead to a dataset with the results (Figure 28).

Tasks: [Running](#) [Cancelled](#) [Completed](#) [Error](#)

Start time	Elapsed time,ms	Task	Name	Status
Mon Sep 13 22:35:59 EEST 2010		<a href="#">11c84b75-90ce-4ba4-a035-56e2ab8d78bb</a>	<a href="#">Apply_Model_ToxTree: Cramer rules to http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12</a>	Running

Figure 28 The results of the predictive model are delivered as a URL

The Cramer rules model is an implementation of Cramer et al., *Estimation of Toxic Hazard – A Decision Tree Approach*, J Cosmet Toxicol, Vol. 16, pp. 255–276, Pergamon Press, 1978. It comprises 33 structural rules and places evaluated compounds into one of three classes:

- Class I substances are simple chemical structures with efficient modes of metabolism suggesting a low order of oral toxicity;
- Class III substances are those that permit no strong initial presumption of safety, or may even suggest significant toxicity or have reactive functional groups; and finally,
- Class II are intermediate. This model is very conservative and places most of the compounds in Class III.

During this exercise, we'll look for compounds of low toxicity (Class I) and high antimalarial activity. There are a small number of Class I compounds, the distribution can be seen via the OpenTox chart generation service

[http://pirin.uni-plovdiv.bg:8080/malaria/chart/pie?dataset\\_uri=http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/212](http://pirin.uni-plovdiv.bg:8080/malaria/chart/pie?dataset_uri=http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/212)

### toxTree.tree.cramer.CramerRules

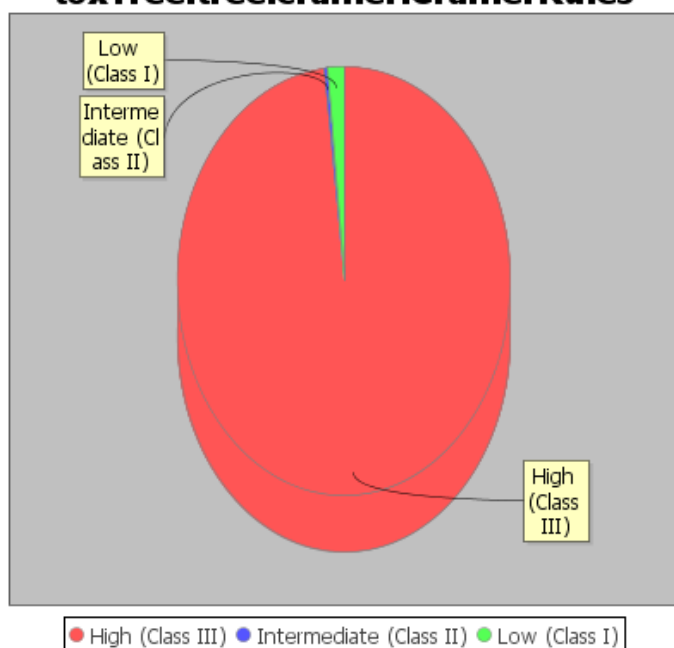
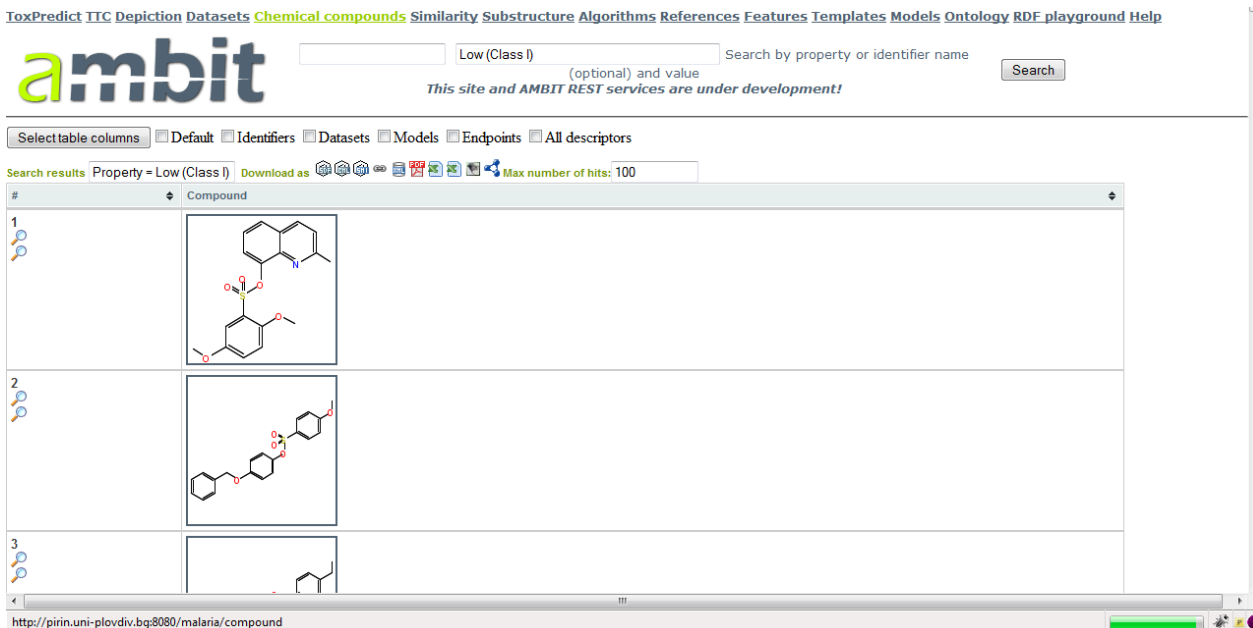


Figure 29 Distribution of the Cramer classes in the TCAMS dataset, depicted using the OpenTox chart generation service

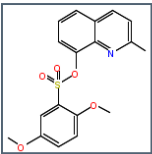
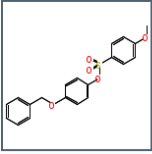

To filter for Class I compounds, click the “Chemical compounds” links on the top, and enter “Low (Class I)” in the search box. This results in the following web address:

<http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29>

(which could be also used directly, instead of typing the search query in the text box). The results can be browsed as below (Figure 30).



The screenshot shows the OpenTox web interface. At the top, there are navigation links: ToxPredict, TTC, Depiction, Datasets, Chemical compounds, Similarity, Substructure, Algorithms, References, Features, Templates, Models, Ontology, RDF, playground, Help. The main search area has a search bar containing 'Low (Class I)' and a 'Search' button. Below the search bar, there are tabs for 'Select table columns' (Default, Identifiers, Datasets, Models, Endpoints, All descriptors). The search results section shows 'Property = Low (Class I)' and 'Max number of hits: 100'. A table displays three chemical structures:

#	Compound
1	
2	
3	

The browser address bar shows the URL: <http://pirin.uni-plovdiv.bg:8080/malaria/compound>

Figure 30 Browsing the results

### 6.3 Step 2: Analyse Cytotoxicities of the Cramer Class I compounds

From the previous step we ended up with a list of compounds considered safe according to Cramer rules. However, we would like to have some more information other than just the chemical structures. For example, we would like to know the antimalarial activity of these compounds.

To add such a column, we need to edit the URL by adding an entry denoting the antimalarial activity given in the TCAMS Dataset. All data columns in OpenTox have their unique URL, and in this example, the URL of the data indicating the percentage inhibition of the growth of the *P. falciparum* strain 3D7 (column "Percentage\_Inhibition\_3D7" in the TCAMS dataset) is

<http://pirin.uni-plovdiv.bg:8080/malaria/feature/190>

Commercial Supplier Reference	Chemical cluster Nr	EXT CMPD NUMBERXC50 MOD 3D7	P. falciparum locus	Percentage inhibition 3D7	Percentage inhibition 3D7 PFLD
	1185.0	TCMDC-131240		100.0	-8.0
	544.0	TCMDC-131241		99.0	0.0
	544.0	TCMDC-131242		94.0	-2.0

<http://pirin.uni-plovdiv.bg:8080/malaria/feature/190>

Figure 31 Retrieving the feature URL

To add this column to our filtered list of compounds considered safe according to Cramer rules (Cramer class I), we simply add a feature\_uris[] parameter to the URL of our filtered list:

[http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190](http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190)

Copy this address into the web browser. There will be a small number of nonempty entries in the second column (see Figure 32).

ToxPredict TIC Depiction Datasets Chemical compounds Similarity Substructure Algorithms References Features Templates Models Ontology RDF playground Help

**ambit**  Search by property or identifier name (optional) and value

This site and AMBIT REST services are under development!

Selectable columns:  Default  Identifiers  Datasets  Models  Endpoints  All descriptors  <http://pirin.uni-plovdiv.bg:8080/malaria/feature/190>

Search results: Property = Low (Class I) Download as        Max number of hits: 100

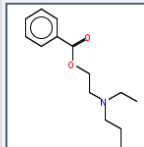
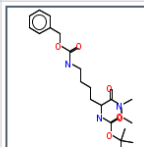

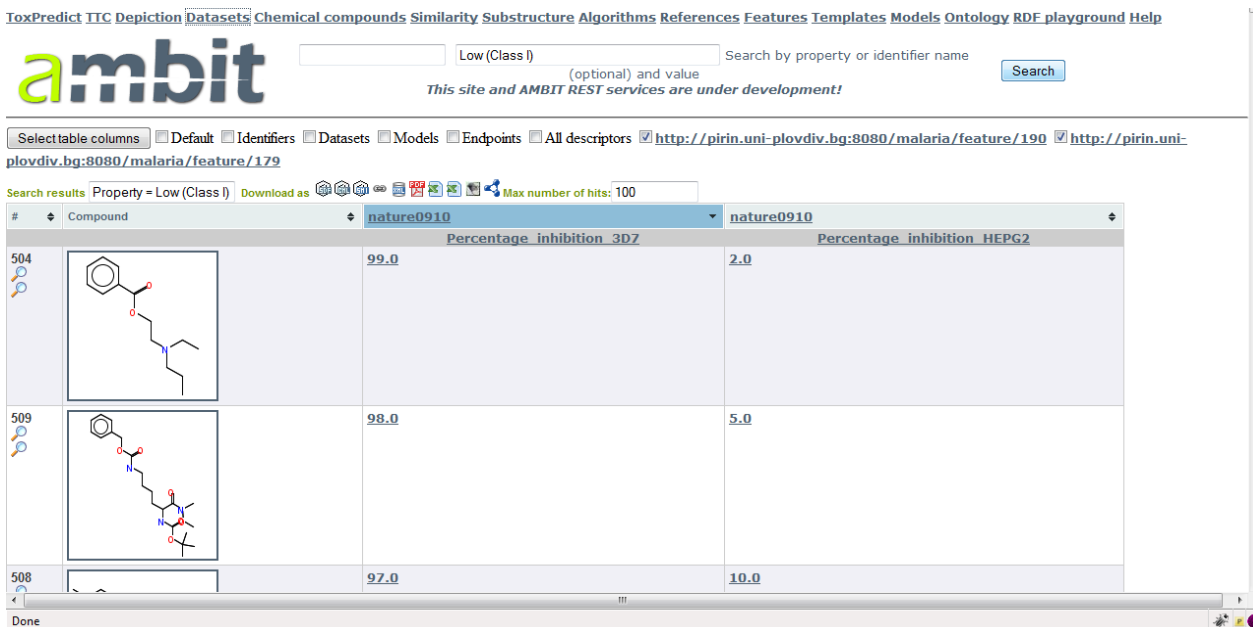
#	Compound	nature0910	Percentage inhibition 3D7
504		99.0	
509		98.0	
508		97.0	

Figure 32 A column is added to the compounds with the feature\_uris[] directive

We're not only interested in the antimalarial activity, but would also like to take into account the experimentally determined human cytotoxicity. To do so, we add a second data column to our filtered list, now with human cytotoxicity data from the TCAMS dataset ([Percentage inhibition\\_HEPG2](#)).

<http://pirin.uni-plovdiv.bg:8080/malaria/feature/179> ). The combination of the two features – antimalarial activity and human cytotoxicity – will result in the following URL (see Figure 33):

[http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/179](http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/179)



The screenshot shows the AMBIT web interface. At the top, there is a search bar with the text 'Low (Class I)' and a 'Search' button. Below the search bar, there are several checkboxes for filtering options: 'Default', 'Identifiers', 'Datasets', 'Models', 'Endpoints', 'All descriptors', and two checked checkboxes for specific URLs. The main area displays search results for 'Property = Low (Class I)'. The results are shown in a table with columns for 'Compound', 'nature0910', 'Percentage inhibition 3D7', and 'Percentage inhibition HEPG2'. Three compounds are listed with their respective chemical structures and inhibition percentages.

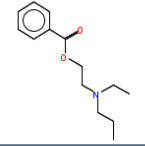
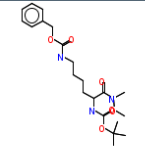

#	Compound	nature0910	Percentage inhibition 3D7	Percentage inhibition HEPG2
504		nature0910	99.0	2.0
509		nature0910	98.0	5.0
508		nature0910	97.0	10.0

Figure 33 Another column is added to the compounds with the feature\_uris[] directive

## 6.4 Step 3: Predicting the Mutagenicity of the Selected Compounds

To add a further criterion to be used when selecting our drug candidate, we predict the compounds' mutagenicities. To do so, we'll use the Toxtree Benigni/Bossa rules for mutagenicity and carcinogenicity (Benigni et al., *Mechanistic QSAR of aromatic amines: new models for discriminating between mutagens and nonmutagens, and validation of models for carcinogens*, Environ Mol. Mutag. **48**:754–771 (2007).). The URL of this model is <http://pirin.uni-plovdiv.bg:8080/malaria/model/12>.

Analogously as you have done for the Cramer rules, follow the URL of the Benigni/Bossa model (<http://pirin.uni-plovdiv.bg:8080/malaria/model/12>), type or paste the URL or the TCAMS dataset (<http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12>) in the text box and click "Predict". Alternatively, the URL of the filtered list could be entered here, as well.

OpenTox models store the prediction results again under data columns with unique URL. These are available via <http://host/model/{id}/predicted> , which in our example corresponds to

<http://pirin.uni-plovdiv.bg:8080/malaria/model/12/predicted>



## Features

Find	Name	Units	Same as	Origin (Dataset, Model or Algorithm)	Nominal values
	Structural Alert for genotoxic carcinogenicity		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Structural Alert for nongenotoxic carcinogenicity		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Potential <i>S. typhimurium</i> TA100 mutagen based on QSAR		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Unlikely to be a <i>S. typhimurium</i> TA100 mutagen based on QSAR		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Potential carcinogen based on QSAR		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Unlikely to be a carcinogen based on QSAR		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	For a better assessment a QSAR calculation could be applied.		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Negative for genotoxic carcinogenicity		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Negative for nongenotoxic carcinogenicity		<a href="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">http://www.opentox.org/echaEndpoints.owl#Carcinogenicity</a>	Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)	NO
	Error when applying the decision tree		<a href="http://www.opentox.org">http://www.opentox.org</a>	Benigni / Bossa rulebase (for mutagenicity and	NO

<http://pirin.uni-plovdiv.bg:8080/malaria/feature/259>

Figure 34 Selecting the Benigni/Bossa rulebase for mutagenicity and carcinogenicity


The Toxtree mutagenicity and carcinogenicity model predicts whether there are structural alerts for genotoxic or nongenotoxic carcinogenicity, and also uses a linear discriminant model for specific classes of compounds.

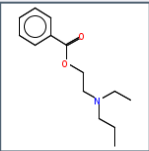
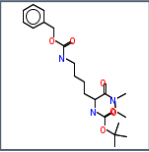
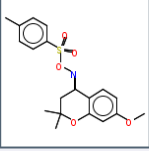
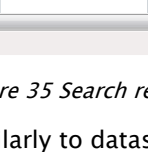
For our purpose, we select the columns “[Structural Alert for genotoxic carcinogenicity](http://pirin.uni-plovdiv.bg:8080/malaria/feature/258)“ (<http://pirin.uni-plovdiv.bg:8080/malaria/feature/258>) and “[Structural Alert for nongenotoxic carcinogenicity](http://pirin.uni-plovdiv.bg:8080/malaria/feature/259)“ (<http://pirin.uni-plovdiv.bg:8080/malaria/feature/259>). As before, we add data columns for these structural alerts to our Cramer–class filtered list of compounds, again using the `feature_uris[]` method. The resulting URL is:

[http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+\(Class+I\)&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/179&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/258&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/259](http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+(Class+I)&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/179&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/258&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/259)

The resulting table (as well as any other) can be sorted according to the values in any column by clicking on the column header.

In the following examples, we’ll consider the first compound in the image below as our antimalarial drug candidate. It is a Cramer class I compound that inhibits growth of *P. falciparum* 3D7 by 99% at the concentration tested (2µM), has a very low human cytotoxicity and no structural alerts for carcinogenicity. (You may choose a different compound).

Search results Property = Low (Class I) Download as  Max number of hits: 100

#	Compound	Percentage inhibition 3D7	Percentage inhibition HEPG2	Benigni / Structural Alert for genotoxic carcinogenicity	Benigni / Structural Alert for nongenotoxic carcinogenicity
504		99.0	2.0	NO	NO
509		98.0	5.0	YES	NO
508		97.0	10.0	NO	NO
512		96.0	30.0	NO	NO

Done

Figure 35 Search results enhanced with columns for carcinogenicity predictions

Similarly to datasets and models, each compound in OpenTox services also has its unique URL. You can find the URL of a compound by clicking on its 2D structure, and stripping off the “?media=text/html” part at the end of the URL this brings you to.

The URL of the compound selected above is

<http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441>.

## 6.5 Step 4: Predicting Sites of Cytochrome P450 Metabolism

The URL of our drug candidate will be used to submit this compound to two models predicting cytochrome P450 sites of metabolism, namely

SmartCYP<sup>2</sup> (<http://pirin.uni-plovdiv.bg:8080/malaria/model/10>) and

SOME<sup>3</sup> (<http://pirin.uni-plovdiv.bg:8080/malaria/model/21> )

Model prediction is done analogously to the two models used in this exercise. Go to

<http://pirin.uni-plovdiv.bg:8080/malaria/model/10> and copy the compound URL into the text box.

<sup>2</sup> Rydberg P. et al. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. ACS Medicinal Chemistry Letters, 1(3), 96–100 (2010)

<sup>3</sup> Zheng M. et al. Site of metabolism prediction for six biotransformations mediated by cytochromes P450. Bioinformatics 25(10): 1251–1258 (2010)

This site and AMBIT REST services are under development!

Model name	Algorithm	Dataset	Independent variables	Dependent	Predicted
SmartCYP: Cytochrome P450-Mediated Drug Metabolism txt legend	<a href="http://pirin.uni-plovdiv.bg:8080/malaria/algorithm/toxtreesmartcyp">http://pirin.uni-plovdiv.bg:8080/malaria/algorithm/toxtreesmartcyp</a>		Independent variables	Dependent	Predicted
Dataset URI	<a href="http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441">http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441</a>		<input type="button" value="Predict"/>		

Figure 36 Running SmartCYP on the malaria box

When completed, the results will be available at

[http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441?feature\\_uris\[\]=http%3A%2F%2Fpirin.uni-plovdiv.bg%3A8080%2Fmalaria%2Fmodel%2F10%2Fpredicted](http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441?feature_uris[]=http%3A%2F%2Fpirin.uni-plovdiv.bg%3A8080%2Fmalaria%2Fmodel%2F10%2Fpredicted)

and will consist of information on which atoms are of rank 1, 2, 3 or lower. Higher rank means a more labile site. This information will be best viewed graphically, which could be achieved by the following URL

[http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441?model\\_uri=http://pirin.uni-plovdiv.bg:8080/malaria/model/10](http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441?model_uri=http://pirin.uni-plovdiv.bg:8080/malaria/model/10) (see Figure 37).

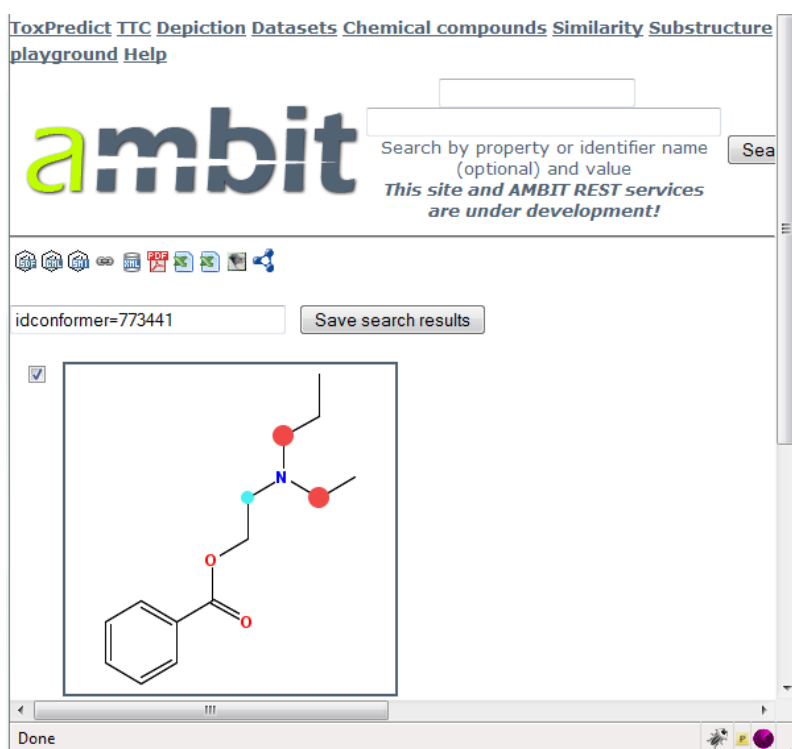


Figure 37 Example SmartCYP output

The colour code for the result can be found by clicking on the “legend” link on the model page.



Figure 38 SmartCYP color code

Similarly, the SOME model predictions are visualized via

[http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441?model\\_uri=http://pirin.uni-plovdiv.bg:8080/malaria/model/21](http://pirin.uni-plovdiv.bg:8080/malaria/compound/458166/conformer/773441?model_uri=http://pirin.uni-plovdiv.bg:8080/malaria/model/21)

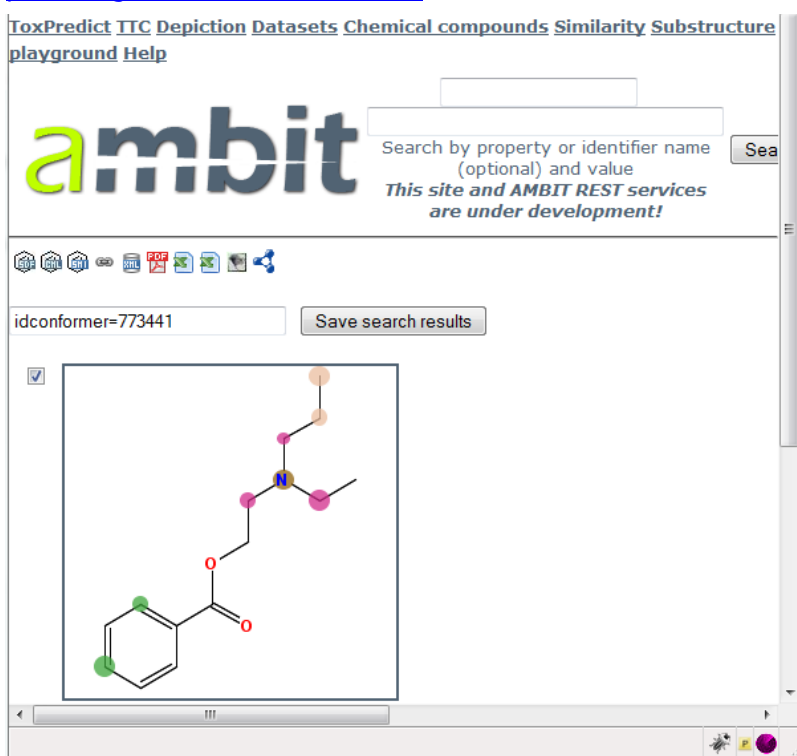


Figure 39 Example SOME output

And the color code is

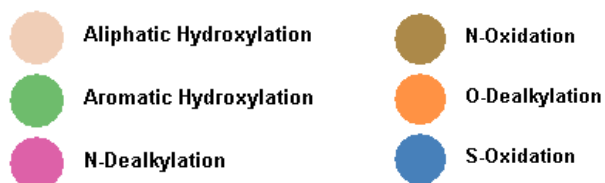


Figure 40 SOME color code

The color-coding of the metabolic sites according to the type of metabolic reaction taking place allows the user - with a little knowledge in organic chemistry - to work out the metabolites of the compound.

To obtain the predictions of the sites of metabolism for the entire dataset, one can use the following URL:

[http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/179&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/258&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/259&model\\_uri=http://pirin.uni-plovdiv.bg:8080/malaria/model/21&w=250&h=250](http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=&search=Low+%28Class+I%29&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/190&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/179&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/258&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/259&model_uri=http://pirin.uni-plovdiv.bg:8080/malaria/model/21&w=250&h=250)

## 7 Drug Discovery Predictive Toxicology Application II: Building a Model to Predict Kinase Inhibitor Activity

### 7.1 Introduction

Using the data on antimalarial compounds made available at the ChEMBL Neglected Tropical Disease (NTD) archive (<http://www.ebi.ac.uk/chemblntd/>), in this exercise subsets of the antimalarials are extracted to be used in a model building exercise via the OpenTox prototype application ToxCreat. 857 of the 13'519 compounds contained in the TCAMS dataset are annotated with a protein (class) target. Of these 857 compounds, 233 are annotated as Ser/Thr kinase inhibitors. In this exercise we'll use this information to create a dataset that can be used to build a model that predicts whether or not a given compound is likely to be a kinase inhibitor. The dataset for the model building therefore needs to consist of two columns: the SMILES string of the compound and its classification (Ser/Thr kinase inhibitor = 1, otherwise 0).

### 7.2 Activity B: Selecting a subset to create a model with ToxCreat

To create the dataset required for model building go to <http://pirin.uni-plovdiv.bg:8080/malaria/dataset>

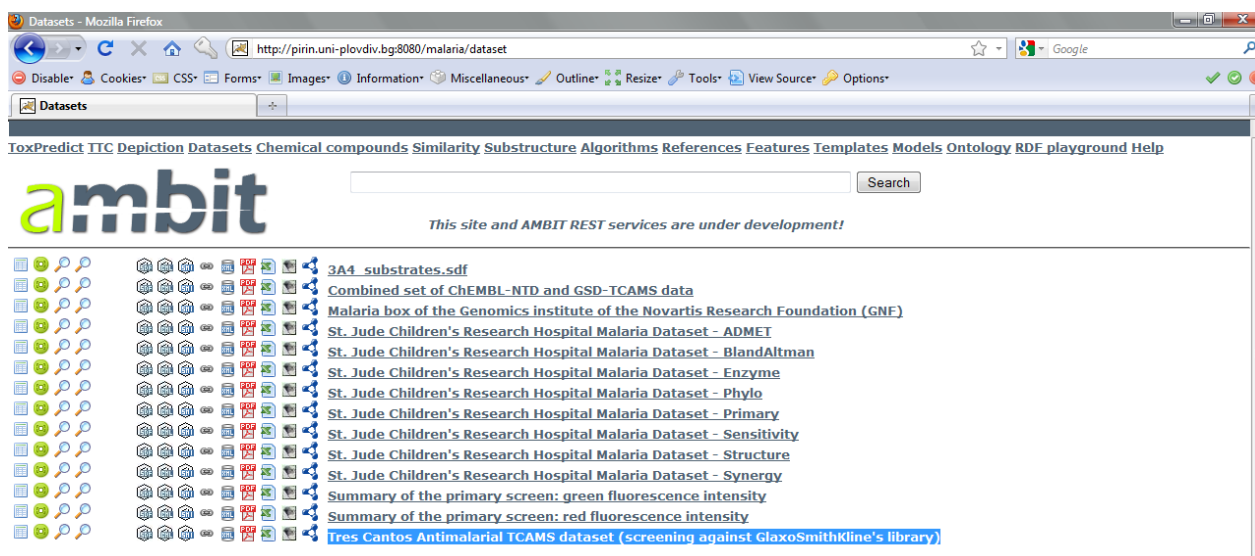
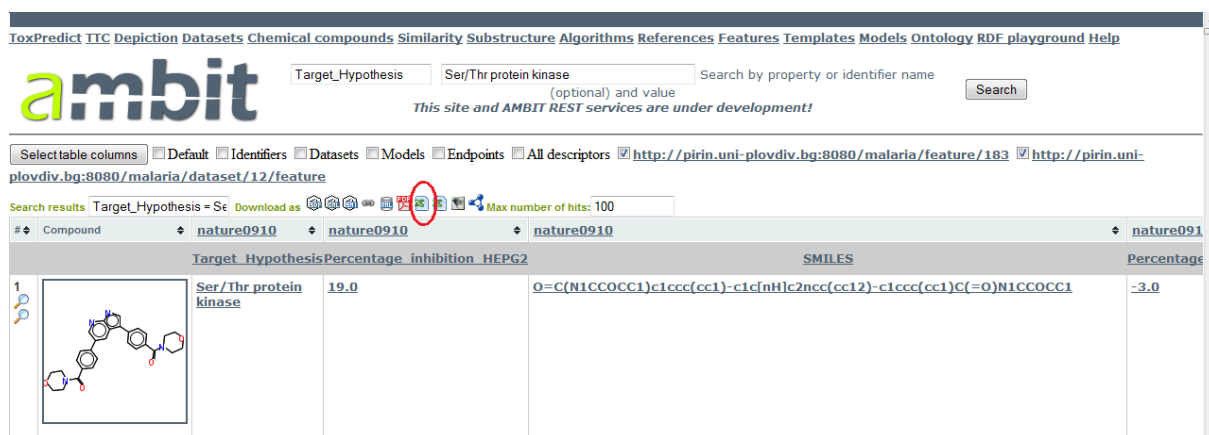


Figure 41 The list of antimalarial datasets on <http://pirin.uni-plovdiv.bg:8080/malaria/dataset>

Click on [“Tres Cantos Antimalarial TCAMS dataset \(screening against GlaxoSmithKline's library\)”](#)

Browse the dataset and find the column “Target hypothesis”. You will note that most entries are empty (only ~6% of the compounds have a target hypothesis annotated). In the 100 compounds displayed by default when following the link to the TCAMS data, you will only find one entry with value [“Adrenergic receptor antagonist”](#). You could click on the link, which would filter out only compounds with this potential target.

For our purpose, we want the list of compounds annotated to be kinase inhibitors. You could try to increase the number of displayed compounds until you find one, or you could enter “Ser/Thr protein kinase” in the searching text box at the top of the page and click the “Search” button. The results will be displayed as below (see Figure 42).



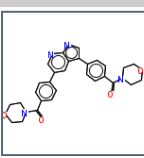
#	Compound	Target_Hypothesis	Percentage_inhibition_HEPG2	SMILES	Percentage
1		Ser/Thr protein kinase	19.0	<chem>O=C(N1CCOCC1)c1ccc(cc1)-c1c[nH]c2ncc(cc12)-c1ccc(cc1)C(=O)N1CCOCC1</chem>	-3.0

Figure 42 Search results for “Ser/Thr protein kinase” on the TCAMS antimalarial dataset

To build a model, it is not enough to have a list of Ser/Thr kinase inhibitors. We also need some “negatives”. Although strictly speaking we don’t have any true negatives, we will use the compounds that do have a target hypothesis annotation – but one that is not “Ser/Thr kinase” – as negatives. So, we extract the whole list of compounds with non-empty target hypothesis, and replace “Ser/Thr kinase” with a “1”, and all the other target hypotheses with “0”.

To extract the list of compounds with non-empty target hypotheses, use the following URL:

[http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=Target\\_Hypothesis&search=+&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/183&feature\\_uris\[\]=http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12/feature&max=1000&condition=!%3D](http://pirin.uni-plovdiv.bg:8080/malaria/compound?type=smiles&property=Target_Hypothesis&search=+&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/feature/183&feature_uris[]=http://pirin.uni-plovdiv.bg:8080/malaria/dataset/12/feature&max=1000&condition=!%3D)

This operation is not (yet) possible via the “Search” text field (it does not allow negation, e.g. something like Target\_Hypothesis !=”), but only via the URL: briefly, the search for non-empty Target Hypothesis is done in the above URL, first with **&search=+** (the “+” stands for empty) – thus searching for all the empties – and then negating the search by **&condition=!%3D** (%3D stands for the “=” sign, thus !%3D stands for !=, or “not equal”).

When following the above URL you’ll get a table with compounds that have a non-empty Target\_Hypothesis. The next step will be to export data. Click on the left one of the two little Excel icons (when moving the mouse pointer on top of it, a small text box “text/csv” should appear) to save the selected data as CSV.

For the model building, we will use the OpenTox application ToxCreate ([www.toxcreate.org](http://www.toxcreate.org)). Thus, first we need to format the data as explained at [www.toxcreate.org/help](http://www.toxcreate.org/help). That is, we leave only the SMILES column and the Target\_Hypothesis column.

Now you should have the Target\_Hypothesis in column 1 (or A), and the SMILES in column 2 (or B). If you are using Excel, go to the cell C2. Type

=IF(A2="Ser/Thr protein kinase"; 1; 0)

and hit “Enter”. Again click on cell C2 to activate it. Now double-click on the little black square at the bottom-right corner of the cell’s border to fill the column with this formula.

Now, copy the whole column C, and paste it (at the same place) using Excel’s “Paste Special” function, pasting only the values. Once that’s done, delete column A (holding the text entries for the Target\_Hypothesis). Delete as well row 1 and save the resulting table as text CSV file to TCAMS-kinase\_full.csv.

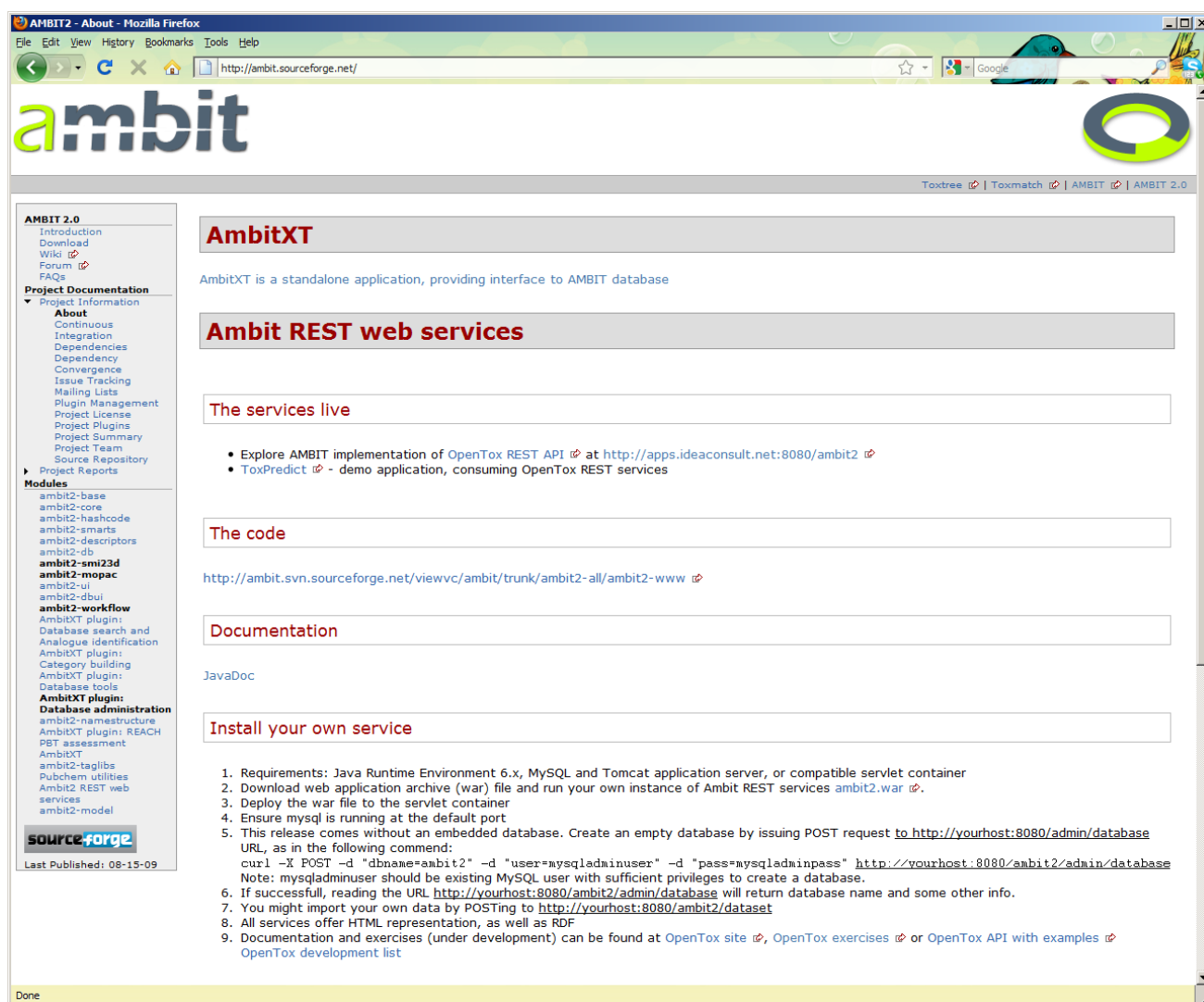
In your web browser, navigate to [www.toxcreate.org](http://www.toxcreate.org). Read the instructions, and try to create a model using your dataset. As ToxCreate is currently a prototype, there are still some limitations. You might get an error in the model building, in which case you could try to reduce the number of compounds used to build the model

to about 600. Just delete some rows until that table contains 600 rows or less. Save the resulting table to TCAMS-kinase-subset.csv.

## 8 Set up an OpenTox Data Service

### 8.1 Introduction

IDEA has developed a standalone application, AMBIT 2.0, that implements the OpenTox REST API. The application is made available as a web archive, to be downloaded at [ambit.sourceforge.net](http://ambit.sourceforge.net). AMBIT 2.0 allows for the easy setup of one's own OpenTox data service. Instructions are given at [ambit.sourceforge.net](http://ambit.sourceforge.net) on how to deploy the web archive and how to set up the data service:



The screenshot shows the SourceForge page for AMBIT 2.0. The browser window title is "AMBIT 2.0 - About - Mozilla Firefox" and the address bar shows "http://ambit.sourceforge.net/". The page features a navigation menu on the left with categories like "AMBIT 2.0", "Project Documentation", "Modules", and "sourceforge". The main content area has several sections: "AmbitXT" (describing it as a standalone application), "Ambit REST web services", "The services live" (with links to the OpenTox REST API and ToxPredict demo), "The code" (with a link to the SVN repository), "Documentation" (with a link to JavaDoc), and "Install your own service" (with a numbered list of requirements and steps). The "Install your own service" section includes requirements for Java Runtime Environment 6.x, MySQL, and Tomcat, and provides a curl command to create a database.

Figure 43 Source Forge page of AMBIT 2.0

However, getting point 1 of the instructions (“Requirements: Java Runtime Environment 6.x, MySQL and Tomcat application server, or compatible servlet container”) satisfied can pose some problems under certain Linux distributions. The following sections describe how to get set up both on Windows and on a CentOS Linux system.



## 8.2 Setting up a Windows System

Getting the software set up that is necessary to install the AMBIT 2.0 application is fairly straightforward under Windows. The example given here is for a Windows 7 Home Premium 64bit system.

### 8.2.1 Installing the Java Runtime Environment (JRE) and Java Development Kit (JDK)

The 32-bit version of JRE 6 Update 21 was downloaded from [www.java.com](http://www.java.com) located at <http://javadl.sun.com/webapps/download/AutoDL?BundleId=41723> and installed. Most likely, JRE will already be installed on a Windows system, and the installer will inform you about that.

The 64-bit version of JRE 6 Update 21 was downloaded from <http://www.oracle.com/technetwork/java/javase/downloads/index.html>. The direct download link is [this](#) (link is too long to be displayed).

JDK 6 Update 21 was downloaded from [www.oracle.com](http://www.oracle.com) located at <http://www.oracle.com/technetwork/java/javase/downloads/index.html>; the direct download link to the Windows 64-bit software is [this](#) (link too long to display). The installer will again inform if JDK 6u21 is already installed on your system, in which case the installation can be aborted.

### 8.2.2 Installing MySQL

The current version of MySQL is version 5.1.50. Unfortunately, the 5.1.\* versions of MySQL do not work properly anymore with the current version of the AMBIT application as of September 2010. Therefore, if MySQL is not installed yet, download version 5.0.\* from <http://downloads.mysql.com/archives.php?p=mysql-5.0&o=-win> (e.g., version 5.0.91 as Windows 64-bit installer from <http://downloads.mysql.com/archives/mysql-5.0/mysql-essential-5.0.91-winx64.msi>).

For this tutorial, the Setup Type “Typical” was chosen, and “Configure the MySQL Server now” was selected.

A “Standard Configuration” was chosen for the MySQL Server, the “Install As Windows Service” and the “Launch the MySQL Server automatically” checkboxes were left checked on, and the “Include Bin Directory in Windows PATH” checkbox was checked.

Security Settings were modified in the sense that a root password was defined, and root access from remote machines was enabled. No anonymous account was created.

### 8.2.3 Installing Tomcat

Apache Tomcat 6.0.29 was downloaded as a service installer file from <http://apache.mirror.testserver.li//tomcat/tomcat-6/v6.0.29/bin/apache-tomcat-6.0.29.exe>. A “Normal” type of install was selected. The connector port was left at its default value of 8080. A Tomcat Administrator Login was defined. The path to the installed 64-bit JRE was given as C:\Program Files\Java\jdk1.6.0\_21\jre.

After installing Tomcat, the Release Notes explain that Tomcat 6.0 no longer needs to have the complete JDK, and that a JRE is now sufficient. So, probably it wasn’t necessary to install JDK to deploy AMBIT 2.0, but no tests were made if that’s true.

If everything went alright during the installation of Tomcat 6.0, the service should be running now, and navigating to <http://localhost:8080> should display the common Tomcat starting page (see Figure 46). If this page cannot be found, some settings might still be wrong. There should be a small icon for Tomcat in the taskbar of Windows 7 (Figure 44). Right-clicking on this icon allows starting the service, selecting “Configure...” allows configuring an automatic start of the service (Figure 45). Try clicking the “Start” button under the “General” tab, to see if you can start the service manually. If not, take a look at the entries in the “Java” tab, in

particular the one under “Java Virtual Machine”. If you’re running a 64-bit machine, you should point Tomcat to a 64-bit Java Virtual Machine. Revise your settings, and try if you can start the service.

Once Tomcat is started, continue with paragraph 8.4 “Download and Deploy AMBIT 2.0”.

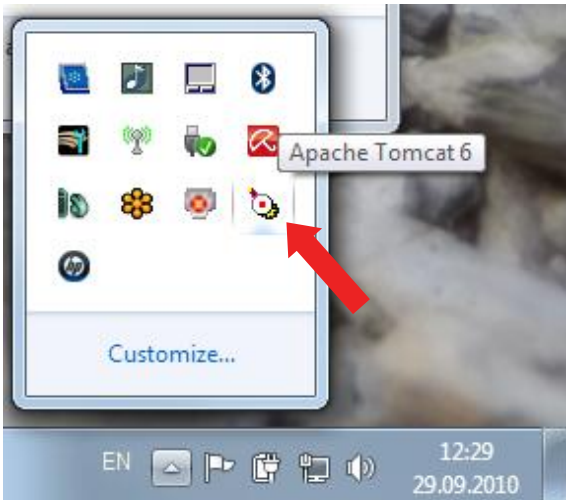


Figure 44 The Tomcat icon in the Windows7 taskbar

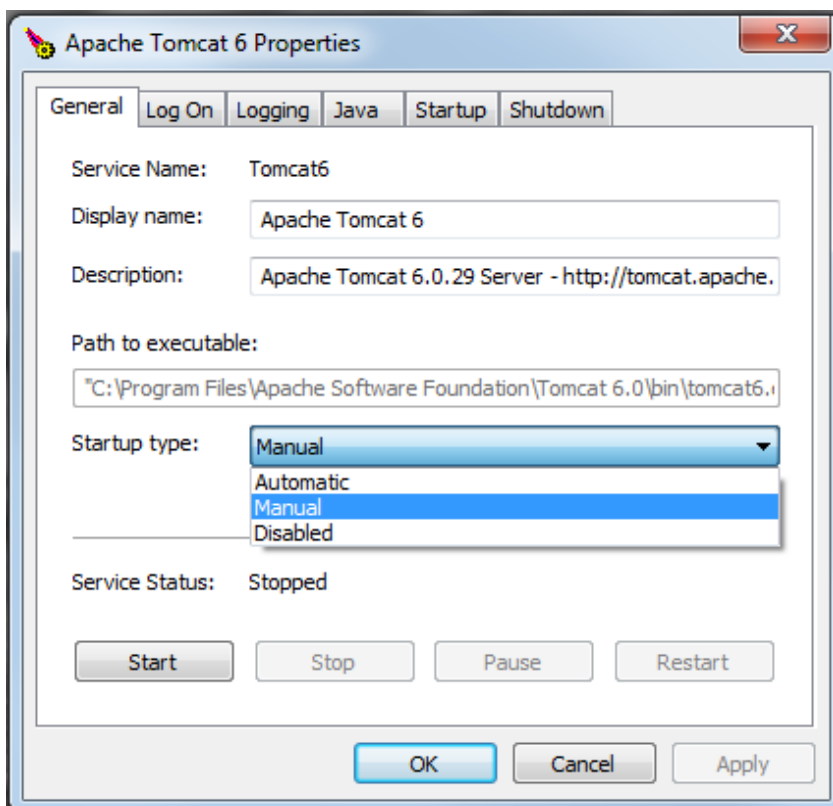


Figure 45 Configuring the Tomcat startup type

## 8.3 Setting Up a CentOS 5 Linux System

Reminding ourselves of the requirements to deploy the AMBIT 2.0 web archive, “Requirements: Java Runtime Environment 6.x, MySQL and Tomcat application server, or compatible servlet container”, the following sections explain how to set up each one of these software. We assume a more or less “virgin” CentOS x86\_64 system, that is, neither of these software is installed already. We also assume that root privileges are available.

### 8.3.1 Installing MySQL

One of the more easy parts is the installation of MySQL. It turns out that a MySQL installation using the YUM package manager is sufficient. Given that package repositories for YUM have been configured, the MySQL installation (as root) is as simple as

```
# yum install mysql.x86_64 mysql-server.x86_64
```

To start the MySQL daemon, run

```
# /etc/init.d/mysqld start
```

To have mysqld start automatically when the system is booted, go to the directory /etc/init.d and run

```
# chkconfig --add mysqld
```

Follow the MySQL manual to create a root password for MySQL. The current version of MySQL distributed with CentOS 5 is MySQL 5.1 (September 2010). Find the manual in A4 format at <http://downloads.mysql.com/docs/refman-5.1-en.a4.pdf>, and in US Letter format at <http://downloads.mysql.com/docs/refman-5.1-en.pdf>. In particular, follow the instructions (and links) in section 2.13 “Post Installation Setup and Testing” on page 133 of the A4 manual and on page 143 of the US Letter manual.

### 8.3.2 Installing the Java Runtime Environment and Development Kit

Download the Java Runtime Environment (JRE) from

<http://javadl.sun.com/webapps/download/AutoDL?BundleId=40911>. In this case, it is the x64\_64 linux package `jre-6u21-linux-x64.bin`. Save it e.g. to /usr/java/ (create the directory if necessary). Also download the Java SE Development Kit (JDK) from <http://www.oracle.com/technetwork/java/javase/downloads/jdk6-jsp-136632.html>. The direct download link for the Linux x86\_64 file is [here](#) (URL is too long to display). Also save the JDK (`jdk-6u21-linux-x64.bin`) to /usr/java/.

Install the two packages as simply as follows:

```
# cd /usr/java
# sh jre-6u21-linux-x64.bin
# sh jdk-6u21-linux-x64.bin
```

### 8.3.3 Installing Tomcat

#### 8.3.3.1 Download, Unpack and Prepare Tomcat

Download `apache-tomcat` (<http://apache.mirror.testserver.li//tomcat/tomcat-6/v6.0.29/bin/apache-tomcat-6.0.29.tar.gz>) and `apache-ant` (<http://apache.mirror.testserver.li//ant/binaries/apache-ant-1.8.1-bin.tar.gz>).

Should you have problems downloading, find the downloads from <http://apache.org> and change the mirror.

Save the two archives wherever you want; here we assume they’re saved to /usr/share.

Unzip the archives to /usr/share:

```
# tar -xzf apache-tomcat-6.0.29.tar.gz
# tar -xzf apache-ant-1.8.1-bin.tar.gz
```

Create a symbolic link for ant to work:

```
# ln -s /usr/share/apache-ant-1.8.1/bin/ant /usr/bin
```

In case ant is already present in /usr/bin, it is probably from a previous package installation. Try to remove ant with

```
# yum remove ant
```

and set the symbolic link again.

Next, we need to tell Tomcat where to find the Java Runtime Environment. This can be done by setting the appropriate variable in the Tomcat run script catalina.sh.

```
# cd /usr/share/apache-tomcat-6.0.29/bin
```

Open catalina.sh with your favourite text editor

```
# emacs catalina.sh
```

and add the line `JAVA_HOME=/usr/java/jdk1.6.0_21` right after the large comment block at the top of catalina.sh.

---

```
...
# $Id: catalina.sh 947714 2010-05-24 16:57:18Z markt $
# -----
JAVA_HOME=/usr/java/jdk1.6.0_21
# OS specific support. $var _must_ be set to either true or false.
cygwin=false
...

```

---

### 8.3.3.2 Start Up Tomcat

To test if the current set up is working, run

```
# /usr/share/apache-tomcat-6.0.29/bin/startup.sh
```

Take a look at the log file catalina.out (/usr/share/apache-tomcat-6.0.29/logs/catalina.out), and check for potential error messages. Even if there are no errors, most likely you'll find an INFO stating

```
INFO: The APR based Apache Tomcat Native library which allows optimal performance in
production environments was not found on the java.library.path:
/usr/java/jre1.6.0_21/lib/amd64/server:/usr/java/jre1.6.0_21/lib/amd64:/usr/java/jre1.6.
0_21/..lib/amd64:/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib

```

This message can safely be ignored. If ever needed, though, one could install APR following <http://tomcat.apache.org/tomcat-6.0-doc/apr.html>.

To make starting up tomcat easier, create a symbolic link for startup.sh to /usr/bin/tomcat:

```
# ln -s /usr/share/apache-tomcat-6.0.29/bin/startup.sh /usr/bin/tomcat
```

### 8.3.3.3 Creating a Startup Script to run Tomcat as user tomcat

To be able to start Tomcat as user tomcat upon system start, we need to compile a small program that comes with the tomcat archive. In older versions of Tomcat this used to be the package `jsvc.tar.gz`, located in the /bin directory of the Tomcat installation.

In recent versions (see [https://issues.apache.org/bugzilla/show\\_bug.cgi?id=49585](https://issues.apache.org/bugzilla/show_bug.cgi?id=49585)), the program archive is `commons-daemon-native.tar.gz`, located still in the /bin directory of the Tomcat installation. To install the program (assuming gcc is installed. Otherwise, run `yum install gcc -y`) follow these steps:

```
# cd /usr/share/apache-tomcat-6.0.29/bin/
# tar -xzf commons-daemon-1.0.2-native.tar.gz
```

```
# cd commons-daemon-1.0.2-native-src/unix
# export CFLAGS=-m64
# export LDFLAGS=-m64
# ./configure --with-java=/usr/java/jdk1.6.0_21/
# make
```

#### Check whether you have set alternatives for java:

```
# ls -l /etc/alternatives/java
lrwxrwxrwx 1 root root 35 Aug 31 01:50 /etc/alternatives/java -> /usr/lib/jvm/jre-1.6.0-sun/bin/java
```

#### Remove the alternative if anything is found

```
# alternatives -remove java /usr/lib/jvm/jre-1.6.0-sun/bin/java
```

#### Check again

```
# ls -l /etc/alternatives/java
lrwxrwxrwx 1 root root 35 Aug 31 20:59 /etc/alternatives/java -> /usr/lib/jvm/jre-1.4.2-gcj/bin/java
```

#### Again, remove the alternative if anything is found

```
# alternatives -remove java /usr/lib/jvm/jre-1.4.2-gcj/bin/java
```

#### Repeat this procedure until no alternatives are set anymore. Then, install the following alternative:

```
# alternatives --install /etc/alternatives/java java /usr/java/jdk1.6.0_21/bin/java 90
```

#### Now, add a new user, tomcat, to the system:

```
# useradd -d /usr/share/apache-tomcat-6.0.29/ tomcat
```

#### Find the tomcat startup script to run tomcat as user tomcat:

```
# cd /usr/share/apache-tomcat-6.0.29/bin/commons-daemon-1.0.2-native-src/unix/native
```

#### Even though we installed Tomcat 6, the startup script is still called Tomcat5.sh. Copy Tomcat5.sh to Tomcat6.sh and open it with your favorite text editor:

```
# cp Tomcat5.sh Tomcat6.sh
# emacs Tomcat6.sh
```

#### In the script Tomcat6.sh change the lines

```
JAVA_HOME=/home2/java/j2sdk1.4.2_03
CATALINA_HOME=/home/tomcat5/tomcat5/jakarta-tomcat-5/build
DAEMON_HOME=/home/jfclere/daemon
TOMCAT_USER=tomcat5
```

#### to

```
JAVA_HOME=/usr/java/jdk1.6.0_21
CATALINA_HOME=/usr/share/apache-tomcat-6.0.29
DAEMON_HOME=/usr/share/apache-tomcat-6.0.29/bin
TOMCAT_USER=tomcat
```

#### In addition, find any occurrence (there are two: one for starting the daemon, one for stopping it) of

```
$DAEMON_HOME/src/native/unix/jsvc \
```

#### and change it to

```
$DAEMON_HOME/commons-daemon-1.0.2-native-src/unix/jsvc \
```

With these changes, I still had some problems. It turned out that the variable `CATALINA_HOME` in `Tomcat6.sh` was used, although the script says it is only used for multi-instances of Tomcat and I was running only one instance of it. So, in `Tomcat6.sh` I also changed the line

```
CATALINA_BASE=/home/tomcat5/tomcat5/jakarta-tomcat-5/build
```

to

```
CATALINA_BASE=/usr/share/apache-tomcat-6.0.29
```

If you do intend to run multiple instances of Tomcat, please make sure that this (having `CATALINA_HOME` and `CATALINA_BASE` equal) doesn't cause any problems, otherwise adapt your set up.

Now, test to see if the startup script works:

```
# ./Tomcat6.sh start
# pgrep -u tomcat -l
31594 jsvc
```

The process number will likely be different on your system, but you should see something similar.

#### 8.3.3.4 Having the Tomcat Daemon Start Automatically

To have the Tomcat daemon start automatically upon system start, copy the `Tomcat6.sh` script to `/etc/init.d`:

```
# cp Tomcat6.sh /etc/init.d/tomcat6
# chmod +x /etc/init.d/tomcat6
```

Go to `/etc/init.d`, open `tomcat6` with your favourite text editor, and add the following lines right after `#!/bin/sh` (note that in the following two lines the “#” is not the root command prompt, but the shell-script comment symbol):

```
# chkconfig: 234 20 80
# description: Small shell script to start/stop Tomcat using jsvc.
```

You should already be in `/etc/init.d` now. Issue the following two commands:

```
# chkconfig --add tomcat6
# chkconfig --list tocat6
tomcat6          0:off  1:off  2:on   3:on   4:on   5:off  6:off
```

This tells you that `tomcat6` is started in runlevels 2, 3 and 4.

Because we installed Tomcat as root, we need to give the user “tomcat” permission/ownership for the whole Tomcat installation directory:

```
# cd /usr/share/apache-tomcat-6.0.29/
# chown -R tomcat *
# chgrp -R tomcat *
```

Now, try to reboot your system, and see if Tomcat starts properly. Before you do so, empty the log file to find possible errors more easily:

```
# mv /usr/share/apache-tomcat-6.0.29/logs/catalina.out /usr/share/apache-tomcat-6.0.29/logs/catalina.out.1
```

You might get errors related to the boot sequence, that is, tomcat is started too early in the boot process. In that case, you can try to move tomcat to be started later. To do so, change the line `# chkconfig 234 20 80` in `/etc/init.d/tomcat6` to `# chkconfig 2345 96 80`, followed by

```
# chkconfig --del tomcat6
# chkconfig --add tomcat6
```

### 8.3.3.5 Adding Tomcat Users

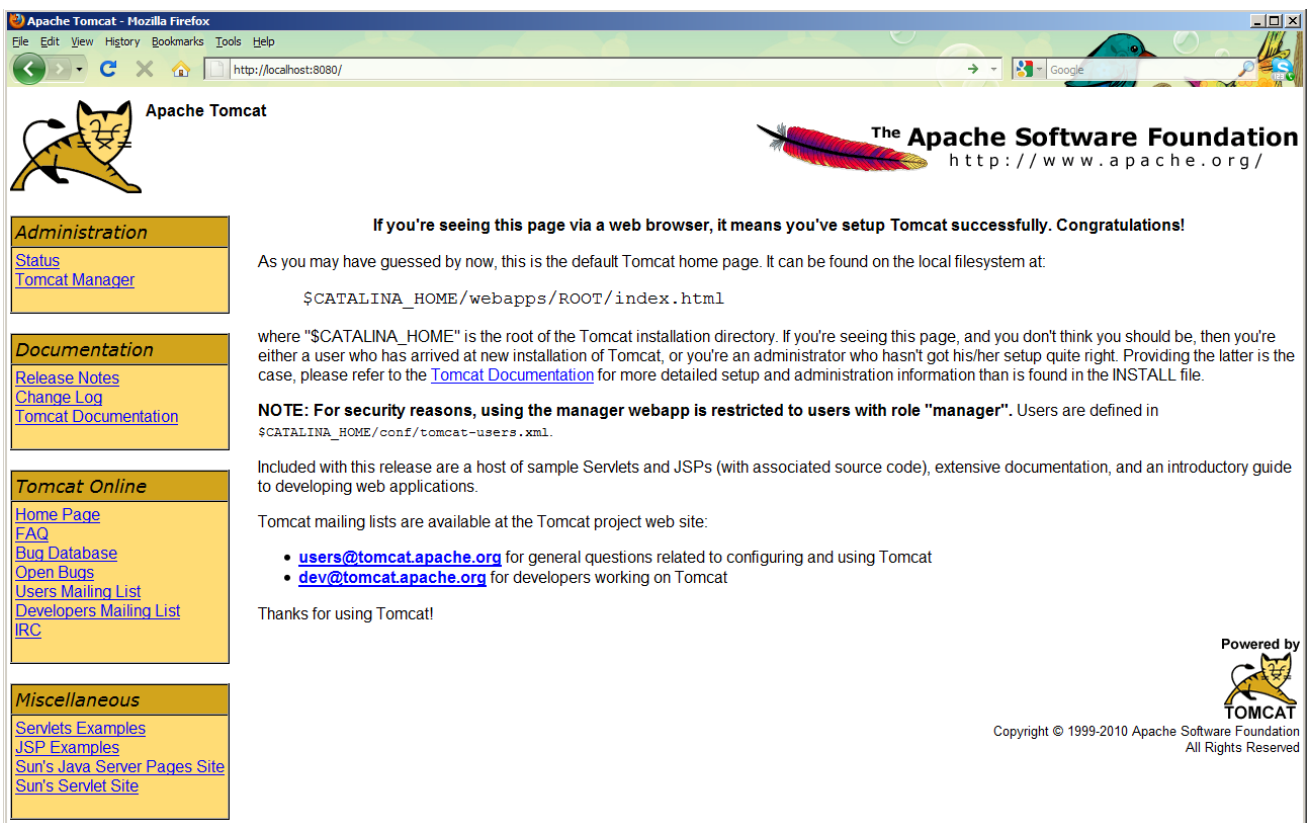
The next step is to add Tomcat users, in particular a manager user. Under the Tomcat configuration directory (`/usr/share/apache-tomcat-6.0.29/conf`), there is a template file to define users, `tomcat-users.xml`. I couldn't get this file to work, so I created one of my own with the following content:

```
# cd /usr/share/apache-tomcat-6.0.29/conf
# cat tomcat-users.xml
<?xml version='1.0' encoding='utf-8'?>
<tomcat-users>
  <role rolename="standard"/>
  <role rolename="manager"/>
  <user username="roman" password="*****" roles="standard,manager"/>
</tomcat-users>
```

Note that I replaced the password with stars above. In reality, the password is stored in the file in clear text. Once you have the users defined, reboot your system, or restart tomcat:

```
# /etc/init.d/tomcat6 stop
# /etc/inid.d/tomcat6 start
```

Open a web browser, and navigate to <http://localhost:8080/>. You should see the following screen (Figure 46):



**Administration**

- [Status](#)
- [Tomcat Manager](#)

**Documentation**

- [Release Notes](#)
- [Change Log](#)
- [Tomcat Documentation](#)

**Tomcat Online**

- [Home Page](#)
- [FAQ](#)
- [Bug Database](#)
- [Open Bugs](#)
- [Users Mailing List](#)
- [Developers Mailing List](#)
- [IRC](#)

**Miscellaneous**

- [Servlets Examples](#)
- [JSP Examples](#)
- [Sun's Java Server Pages Site](#)
- [Sun's Servlet Site](#)

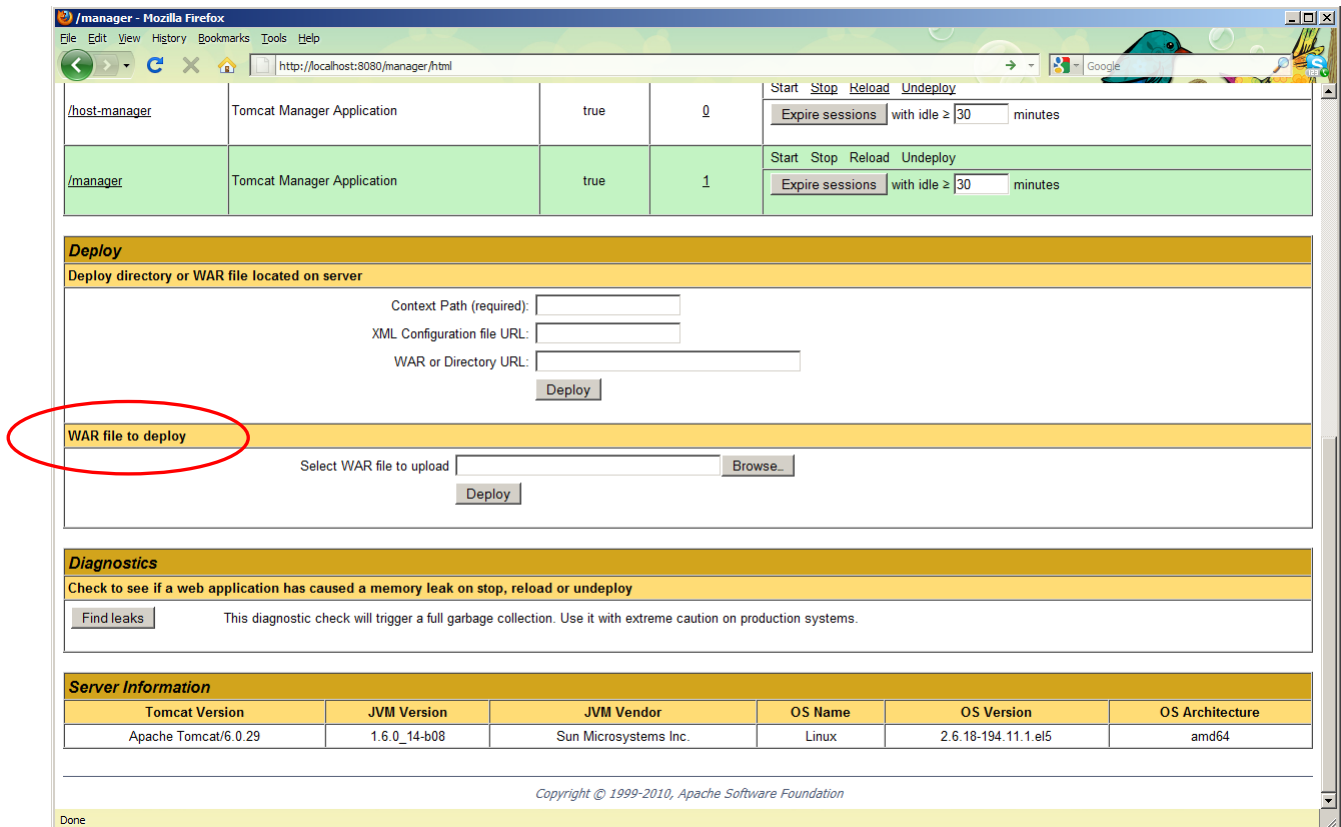
**Powered by TOMCAT**

Copyright © 1999-2010 Apache Software Foundation  
All Rights Reserved

Figure 46: Tomcat "home" page

## 8.4 Download and Deploy AMBIT 2.0

Download the AMBIT 2.0 application (<http://www.ideaconsult.net/downloads/ambit2/ambit2.war>). Save the file `ambit2.war` anywhere you like. With your web browser, navigate to <http://localhost:8080>, and click on “Tomcat Manager” in the Administration box at the top-left of the screen (Figure 46). You’ll be prompted to enter the user name and password of the Tomcat manager/administrator you have set up. On the manager page, scroll to the bottom and find the box entitled “WAR file to deploy” (Figure 47).



The screenshot shows the Tomcat Manager interface. At the top, there is a table with columns for host-manager, manager, and application details. Below this is the 'Deploy' section, which includes a 'WAR file to deploy' section. This section has a 'Select WAR file to upload' field with a 'Browse...' button and a 'Deploy' button. The 'WAR file to deploy' section is circled in red. Below the 'Deploy' section is the 'Diagnostics' section with a 'Find leaks' button. At the bottom is the 'Server Information' section with a table of system details.

Tomcat Version	JVM Version	JVM Vendor	OS Name	OS Version	OS Architecture
Apache Tomcat/6.0.29	1.6.0_14-b08	Sun Microsystems Inc.	Linux	2.6.18-194.11.1.el5	amd64

Figure 47: Tomcat “manager” page

Under “WAR file to deploy”, click “Browse...”, find `ambit2.war` and click “Deploy”. You should now have successfully installed the AMBIT 2.0 implementation of the OpenTox REST API. If you next navigate to <http://localhost:8080/ambit2> you should see the screen corresponding to the one shown in Figure 48. As explained in the installation instructions for AMBIT 2.0 (<http://ambit.sourceforge.net/>), this release (September 2010) comes without an embedded database.

Thus, create an empty database using cURL (<http://curl.haxx.se>). On many linux systems, cURL can be easily installed from a package repository using a standard package manager. It can also be downloaded from <http://curl.haxx.se/download.html>. Under Windows, there are two options for using cURL: 1) installing cURL natively, preferable using this version: <http://www.gknw.net/mirror/curl/win32/curl-7.21.1-devel-mingw32.zip>, or 2) installing the VMWare Player (<http://www.vmware.com/products/player/>) and running a small Linux environment (<http://www.maunz.de/opentox/dsl-4.1.zip>) under Windows (after installing VMWare Player and unpacking the `dsl-4.1.zip` file, just double-click the `dsl-4.1.vmx` file).

Under Linux, after installing cURL simply type the following command as root in a console:

```
# curl -X POST -d "dbname=ambit2" -d "user=mysqladminuser" -d "pass=mysqladminpass"
http://localhost:8080/ambit2/admin/database
```



(replace “*mysqladminuser*” and “*mysqladminpass*” with the username and password of a MySQL user that has sufficient privileges to create a database, see paragraph 8.3.1).

Under Windows, if you have downloaded the cURL .zip file and have extracted it e.g. to C:\Users\Username\PROGRAMS\curl\curl-7.21.0-win64, open a command prompt (cmd.exe, usually under All Programs -> Accessories -> Command prompt), and type the following command (without the “>” sign):

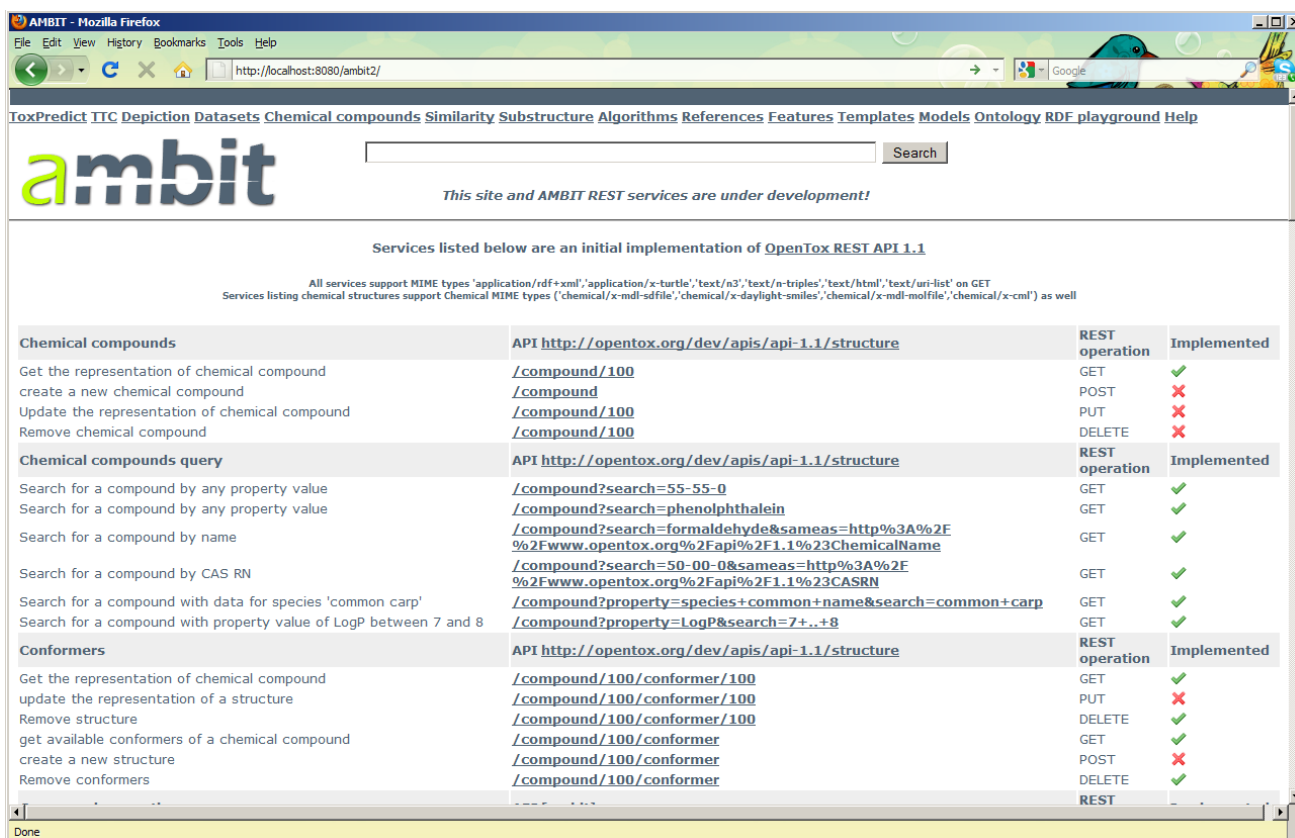
```
> C:\Users\Username\PROGRAMS\curl\curl-7.21.0-win64\curl -X POST -d "dbname=ambit2"
-d "user=mysqladminuser" -d "pass=mysqladminpass"
http://localhost:8080/ambit2/admin/database
```

(replace “*mysqladminuser*” and “*mysqladminpass*” with the username and password of a MySQL user that has sufficient privileges to create a database, see paragraph 8.2.2).

If your POST is successful, navigating to <http://localhost:8080/ambit2/admin/database> should yield something like:

```
Comparing 127.0.0.1 127.0.0.1 : true

ambit2
Version: 3.0
Created: 2010
Note: AMBIT2 schema
```



Services listed below are an initial implementation of [OpenTox REST API 1.1](#)

All services support MIME types 'application/rdf+xml', 'application/x-turtle', 'text/n3', 'text/n-triples', 'text/html', 'text/uri-list' on GET  
 Services listing chemical structures support Chemical MIME types ('chemical/x-mdl-sdfile', 'chemical/x-daylight-smiles', 'chemical/x-mdl-molfile', 'chemical/x-cml') as well

	API <a href="http://opentox.org/dev/apis/api-1.1/structure">http://opentox.org/dev/apis/api-1.1/structure</a>	REST operation	Implemented
<b>Chemical compounds</b>			
Get the representation of chemical compound	<a href="#">/compound/100</a>	GET	✓
create a new chemical compound	<a href="#">/compound</a>	POST	✗
Update the representation of chemical compound	<a href="#">/compound/100</a>	PUT	✗
Remove chemical compound	<a href="#">/compound/100</a>	DELETE	✗
<b>Chemical compounds query</b>	API <a href="http://opentox.org/dev/apis/api-1.1/structure">http://opentox.org/dev/apis/api-1.1/structure</a>	REST operation	Implemented
Search for a compound by any property value	<a href="#">/compound?search=55-55-0</a>	GET	✓
Search for a compound by any property value	<a href="#">/compound?search=phenolphthalein</a>	GET	✓
Search for a compound by name	<a href="#">/compound?search=formaldehyde&amp;sameas=http%3A%2F%2Fwww.opentox.org%2Fapi%2F1.1%23ChemicalName</a>	GET	✓
Search for a compound by CAS RN	<a href="#">/compound?search=50-00-0&amp;sameas=http%3A%2F%2Fwww.opentox.org%2Fapi%2F1.1%23CASRN</a>	GET	✓
Search for a compound with data for species 'common carp'	<a href="#">/compound?property=species+common+name&amp;search=common+carp</a>	GET	✓
Search for a compound with property value of LogP between 7 and 8	<a href="#">/compound?property=LogP&amp;search=7+..+8</a>	GET	✓
<b>Conformers</b>	API <a href="http://opentox.org/dev/apis/api-1.1/structure">http://opentox.org/dev/apis/api-1.1/structure</a>	REST operation	Implemented
Get the representation of chemical compound	<a href="#">/compound/100/conformer/100</a>	GET	✓
update the representation of a structure	<a href="#">/compound/100/conformer/100</a>	PUT	✗
Remove structure	<a href="#">/compound/100/conformer/100</a>	DELETE	✓
get available conformers of a chemical compound	<a href="#">/compound/100/conformer</a>	GET	✓
create a new structure	<a href="#">/compound/100/conformer</a>	POST	✗
Remove conformers	<a href="#">/compound/100/conformer</a>	DELETE	✓
		REST	

Figure 48: The AMBIT 2.0 page

## 9 Conclusions

Initial OpenTox applications can now be used to carry out several core fundamental activities in predictive toxicology. Tutorials have been developed to provide guidance to the user in using these services for model building, predicting toxicities and establishing their own data resources. Additionally tutorials for more complex applications such as examining the US EPA ToxCast dataset and evaluating the toxicity-related properties of anti-malarial inhibitor compounds have been provided. We expect that expansion of these tutorial materials as further applications are developed will provide a useful knowledge resource to the community.

All tutorials and their updates are made available online under [www.opentox.org/tutorials](http://www.opentox.org/tutorials)