

OpenTox - An Open Source Predictive Toxicology Software Framework

239th ACS National Meeting, San Francisco

Barry Hardy, Douglas Connect
OpenTox Project Director

David Gallagher, CAChe Research

Steve Bowlus & Sunil Chawla
Seascape Learning

Overview

- Context and introduction to OpenTox
- An end-users view
- The OpenTox framework
- Summary and impact
- A continuing collaboration

Context - REACH



Context - REACH

European Chemical Bureau (ECB) study showed new REACH regulations will require an estimated 3.9 million **additional** test animals if no alternative methods were accepted

Same study pointed to possible reduction in animal testing by using existing experimental data in conjunction with *in silico* based methods

Largest number of test animals will be required for chronic and reproductive toxicity, mutagenicity, carcinogenicity endpoints because no alternative *in vitro* assays are currently available

Context - Challenges to *in silico* Applications

- Toxicity data collected in many different databases using different formats, frequently incompatible with QSAR and other *in silico* programs
- *In silico* models have been published in a variety of different formats ranging from simple regression based equations to full-fledged computer programs
- Application, interpretation, and development of *in silico* models is still difficult for most toxicological experts

Introduction - OpenTox Goals



Framework

- Toxicity data
- *In silico* models
- Validation support
- Interpretation aids

Unified Access

- Toxicologists
- *In silico* and *in vitro* R&D
- API for service development & integration

Interoperability

- Promote Standards
- Core Open Source Platform
- Support integration of multiple resources and licenses

User Requirements - Use Cases

- OpenTox needs to be very flexible to meet individual needs
- A use case driven development & testing approach assures that users' requirements of the system will be met
- The following end-user view of a common task uses the **ToxPredict** application:

Given a structure, predict the value of a toxicity endpoint

- A second demonstration application, **ToxCreate**, supports the case of creating a predictive toxicology model

Step 1: Search
 Select structure(s)

Step 2: Verify structure
 Verify structure

Step 3: Models
 Select prediction models

Step 4: Estimate
 Estimate

Step 5: Results
 Display results

NEXT

In the first step of your ToxPredict workflow you are invited to select the structure(s) for which you would like to apply some OpenTox model(s). The **Search**, **Draw** and **Upload** tabs below provide different ways for structure selection:

- **Search:** You can enter any free text, including but not limited to **CAS registry number**, **EINECS number**, **SMILES**, **InChI**, chemical name, or any other keyword; ToxPredict will automatically recognize the intended search type and return zero, one or more hits, depending on the search type, the information currently available in the OpenTox database and the selected preferred number of hits.
- **Draw:** You can draw a chemical structure in the built-in editor and perform exact structure or substructure search, or structural similarity based search; ToxPredict will return zero, one or more hits, depending on the search type, the information currently available in the OpenTox database and the selected preferred number of hits.
- **Upload:** You can upload your own dataset(s); ToxPredict accepts **SDF**, **MOL**, **CSV** and **RDF** input file formats, among many others.

Once you have entered your query string, chemical structure drawing or file name for upload, press **NEXT** to proceed to the following step of your ToxPredict workflow.

Search

Draw

Upload

Structure diagram ★

😊	CLR		DEL		D-R	QRY	+/-		UDC		JME	
↶	—	=	≡	~	△	□	⬡	⬢	⬣	⬤	⬥	⬦
C												
N												
O												
S												
F												

Search for
☒ Structure

Step 1: Search

Select structure(s)

Step 2: Verify structure

Verify structure

Step 3: Models

Select prediction models

Step 4: Estimate

Estimate

Step 5: Results

Display results

NEXT

Please select one or more models from the table below and press **NEXT** to proceed to the following step of your ToxPredict workflow. The selected model(s) would be applied to all structures that you have selected in the previous steps of your ToxPredict workflow.

Model	Endpoint	Descriptors	Training dataset	Algorithm
<input checked="" type="checkbox"/> MolecularWeight		-		MolecularWeight
<input type="checkbox"/> CPSA descriptor		-		CPSA descriptor
<input checked="" type="checkbox"/> ZagrebIndex		-		ZagrebIndex
<input type="checkbox"/> Predictive QSAR Model generated by the algorithm mlr		YES		Multiple Linear Regression Training Algorithm
<input type="checkbox"/> Predictive QSAR Model generated by the algorithm mlr		YES		Multiple Linear Regression Training Algorithm
<input type="checkbox"/> OpenTox model created with TUM's kNNregression model learning web service.	Acute toxicity to fish (lethality)	YES		http://opentox.informatik.tu- muenchen.de:8080/OpenTox- dev/algorithm/kNNregression
<input type="checkbox"/> ToxTree: Benigni/Bossa rules for				ToxTree: Benigni/Bossa rules for

Step 1: Search
Select structure(s)


Step 2: Verify structure
Verify structure

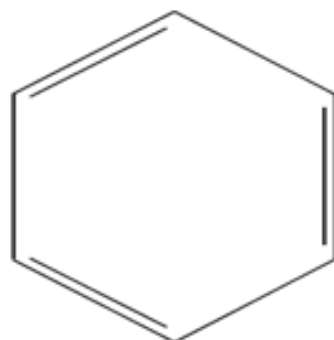
Step 3: Models
Select prediction models

Step 4: Estimate
Estimate

Step 5: Results
Display results

This page lists your ToxPredict workflow results for the structure(s) you have selected and the model prediction(s) you have chosen to run. You could also retrieve the ToxPredict report in various other formats, e.g. [SDF](#), [CML](#), [SMI](#), [PDF](#), [CSV](#), [ARFF](#), [RDF/XML](#) or [RDF/N3](#).

Download as 



CAS RN
EINECS
IUPAC name
Synonym

71-43-2
200-753-7
benzene
(6)annulene; benzine; Benzol; Benzolene;
bicarburet of hydrogen; carbon oil; Coal naphtha;
cyclohexatriene; mineral naphtha; motor benzol;
nitration benzene; Phene; Phenyl hydride;
pyrobenzol.

Synonym
Synonym
Synonym
Quality label

21742.0
Benzene
benzene
OK

MolecularWeight  **MolecularWeight**

MW

78.1112

OpenTox Framework - Definition

- OpenTox is a platform-independent collection of components that interact via well defined language-independent interfaces
- OpenTox is committed to Open Source approaches for core platform components and to integrating multiple licenses as required for any specific application and user context
- OpenTox is committed to the support and further development of Open Standards and Public Ontologies
- The preferred form of communication between components is through web services

OpenTox Framework - Standards

- Minimum information standards for biological experiments
 - Minimum Information for Biological and Biomedical Investigations (MIBBI), Functional Genomics Experiment (FuGE), Predictive Model Markup Language
- Toxicity Data
 - DSSTox, Pub Chem, OECD harmonized templates ...
- Validation
 - **Algorithm:** common best practices such as k-fold cross validation, leave-one-out, scrambling
 - **Model:** OECD Principles, standardized model and prediction formats
- Reports - driven by regulatory requirements
 - REACH guidance documents, Chemicals Safety Report

OpenTox Framework - Components and Interfaces

The initial specifications for the OpenTox Application Program Interfaces (APIs) have been defined

The objects specified are Endpoint, Structure, Structure Identifiers, Feature Definition, Feature, Feature Service, Reference, Algorithm, Algorithm Type, Model, Dataset, Validation Result, Applicability Domain, Feature Selection, and Reporting

The Representational State Transfer (REST) architecture is being used as the web service approach for the communication between components in a distributed system

	OECD Principle	OpenTox addresses by...
1	Defined Endpoint	providing a unified source of well defined and documented toxicity data
2	Unambiguous Algorithm	providing unified access to documented models and algorithms as well as to the source code of their implementation
3	Defined Applicability Domain	integrating tools for the determination of applicability domains and considering these during the validation of (Q)SAR models
4	Goodness-of-fit, robustness and predictivity	providing scientifically sound validation routines for the determination of these measures
5	Mechanistic interpretation (if possible)	providing tools for the prediction of toxicological mechanisms, for the web-mining for toxicological information, and data resources with references relevant to particular (Q)SARs and datasets

Ontology and Data - Endpoints

OpenTox toxicity data infrastructure

- The OpenTox toxicity data infrastructure includes the toxicological end points for which data are required under the REACH regulation
- In current toxicological testing, these endpoints are addressed by both *in vitro* and *in vivo* experiments carried out according to OECD guidelines

REACH toxicological endpoints

- Skin irritation
- Skin corrosion
- Eye irritation
- Dermal sensitisation
- Mutagenicity
- Acute oral toxicity
- Acute inhalative toxicity
- Acute dermal toxicity
- Toxicokinetics
- Repeated dose toxicity (28 days)
- Repeated dose toxicity (90 days)
- Reproductive toxicity screening
- Developmental toxicity
- Two-generation reproductive toxicity study
- Carcinogenicity study

Ontology and Data - Schema and Sources

- ToxML public schema initiated by Leadscope
 - supporting broadly encompassing and meaningful representations of toxicology experiments
 - indexing the data with the chemical structures
 - The ISSCAN carcinogenicity database is fully mapped
 - The ISS *in vivo* micronucleus, bacterial mutagenesis, and the RepDose database are fully mapped to ToxML XSD schema
- OECD harmonized templates
 - contains schemas for all the various endpoints of regulatory relevance
 - required for regulatory reporting
 - The ISSCAN carcinogenicity database is partially mapped
- Additional mapping exercises include those for aquatic toxicity (EPAFHM in DSSTox), repeated doses toxicity (REPDOSE), endocrine disruptors (NCTRER in DSSTox), and a second carcinogenicity database (CPDBAS in DSSTox)

Algorithms

- OpenTox algorithms offer support for common tasks, such as feature generation and selection, aggregation, and visualization
 - The open architecture is designed to allow an easy integration of external programs (open source and commercial)
 - A flexible plug-in architecture for applying, testing and validating algorithms interactively and systematically is used
- OpenTox is starting with the integration of chemoinformatics, statistical and data mining tools including functionality from other Open Source projects

Validation

An **objective validation framework** is paramount for the acceptance and the development of QSAR models.

- The risk assessor needs reliable validation results to assess the quality of predictions.

Model developers need this information:

- to **avoid** the overfitting of models
- to **compare new models** with benchmarked techniques
- to get ideas for the **improvement of algorithms** (eg. from the inspection of misclassified instances).

OpenTox actively supports the **OECD Principles for QSAR Validation** so as to provide easy-to-use validation tools for algorithm and model developers.

Conclusion - Potential OpenTox Impacts

Science

- common standards for data and model exchange
- increasing the reproducibility of QSAR models
- providing scientifically sound prediction and validation routines

Systems

- inclusion of international community of external developers and researchers
- improving interoperability

Target industries

- reduce the number of expensive efficacy and toxicity animal experiments
- reduce the costs for candidate development
- compounds with potential adverse effects will be removed earlier from the product pipeline
- speed up in development cycle

Potential OpenTox Impacts on REACH

ECB estimates that the initial implementation of REACH could result in an additional **3.9 million** animals being used



Chronic effects like reproductive and developmental toxicity, *in vivo* mutagenicity & carcinogenicity will require ~72% of the test animals (~2.8 million animals)



QSAR techniques estimated to reduce animal tests by 33-50%



OpenTox focuses initially on improved QSAR techniques for reproductive, developmental and repeated dose toxicity, and for *in vivo* mutagenicity and carcinogenicity endpoints



So OpenTox could play a major role in the reduction of 1.4 million animal experiments alone for REACH!

New OpenTox website with community, content management, and collaboration areas...

The screenshot shows the OpenTox website homepage. At the top is a purple header with the OpenTox logo on the left and navigation links (Site Map, Accessibility, Contact, Log In, Join) and a search bar on the right. Below the header is a light blue navigation bar with links: Home, Toxicity Prediction, OpenTox Blog, Development, and OpenToxipedia. A grey bar below that contains links: About, Reading Room, FP7, REACH, Mailinglist, and FAQ.

News

- OpenTox Workshop - Development of Predictive Toxicology Applications**
Mar 03, 2010
- Initial OpenTox Applications now available: ToxPredict and ToxCreate**
Feb 13, 2010
- Interoperability Requirements**
Feb 13, 2010
- OpenTox Collaboration Pool is**

You are here: [Home](#)

Welcome to the OpenTox Community site

Welcome to the OpenTox community site. The goal of OpenTox is to develop an interoperable predictive toxicology framework which may be used as an enabling platform for the creation of predictive toxicology applications.

OpenTox applications can combine multiple web services providing users access to distributed toxicological resources including data, computer models, validation and reporting. Applications are based on use cases that satisfy user needs in predictive toxicology.

You can register on the site using the **Join** function. You can also join the **OpenTox Interest Group** through a mailing list and digest. To try out OpenTox applications please visit the **Toxicity Prediction** area.

If interest in participating in OpenTox technical developments please visit the **Development** area.

OpenTox Work Activities

```
graph TD; A[Framework Design] --> B[Framework Implementation]; B --> A;
```

OpenTox was initiated as a collaborative project involving a combination of different enterprise, university and government research groups to design and build the initial OpenTox framework. Additionally numerous organizations with industry, regulatory or expert interests are active in providing guidance and direction. The goal is to expand OpenTox as a

OpenTox Advisory Board

- European Centre for the Validation of Alternative Methods
- European Chemicals Bureau
- U.S Environmental Protection Agency
- U.S. Food & Drug Administration
- Nestlé
- Roche
- AstraZeneca
- LHASA
- Leadscope
- University of North Carolina
- EC Environment Directorate General
- Organisation for Economic Cooperation & Development
- CADASTER
- Bayer Healthcare

Can you help?

For more information, visit the OpenTox website

www.OpenTox.org

or

<http://www.opentox.org/toxicity-prediction/userinput>

or contact:

barry.hardy -(at)- douglasconnect.com

OpenTox - An Open Source Predictive Toxicology Framework, 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).