OpenTox Applications

Dr. Nina Jeliazkova

SMi ADMET Conference

7,8 July 2010

London, UK

OpenTox

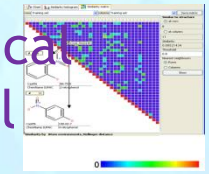
ToxPredict

SEVENTH FRAMEWORK PROGRAMME

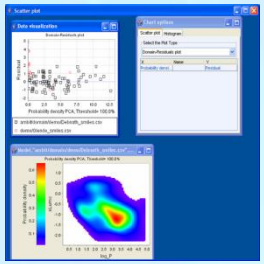
Introduction

- Develops and maintains several well-known open source software products, in particular
  - Toxtree (used in the application of the threshold of toxicological concern)
  - Toxmatch (for encoding and applying chemical similarity indices)
  - Ambit (a QSAR decision support system, including generic database management, structure conversions and searching, as well as applicability domain assessment).
- These tools have been carefully designed and developed in close co-operation both with academia, regulatory bodies and industry.

- Toxtree 2.1.0 - estimates toxic hazard by applying a decision tree approach
- Toxmatch 1.06 – A chemical similarity evaluation tool



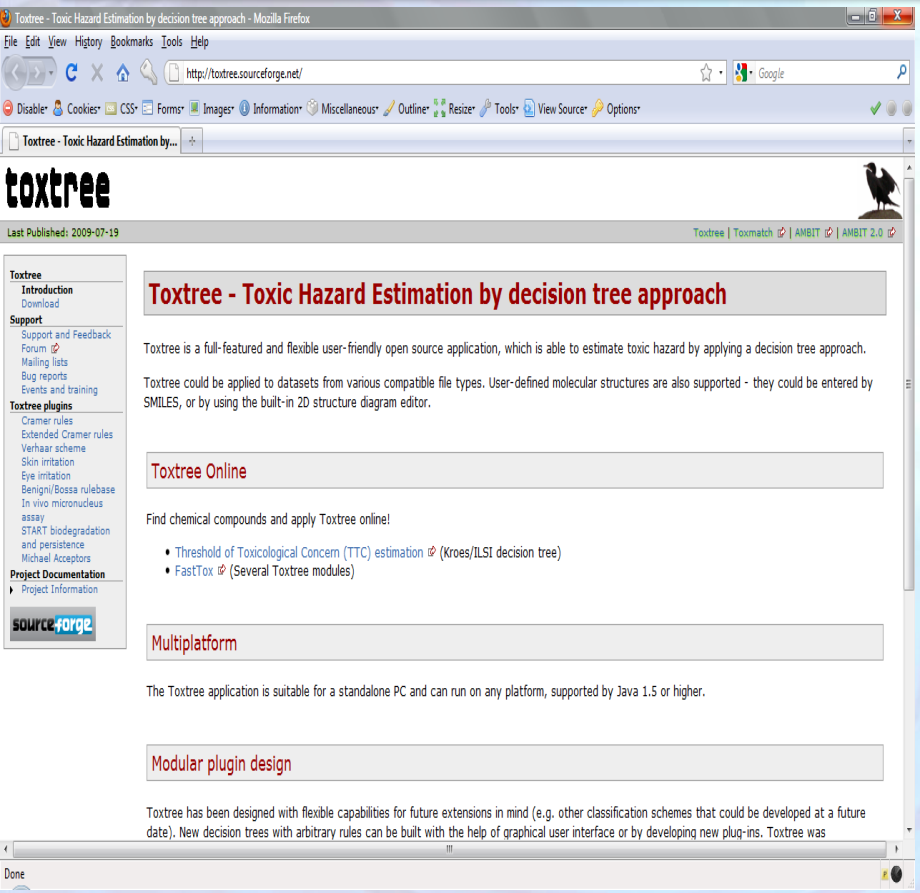
- Ambit Discovery
- Ambit Database Tools

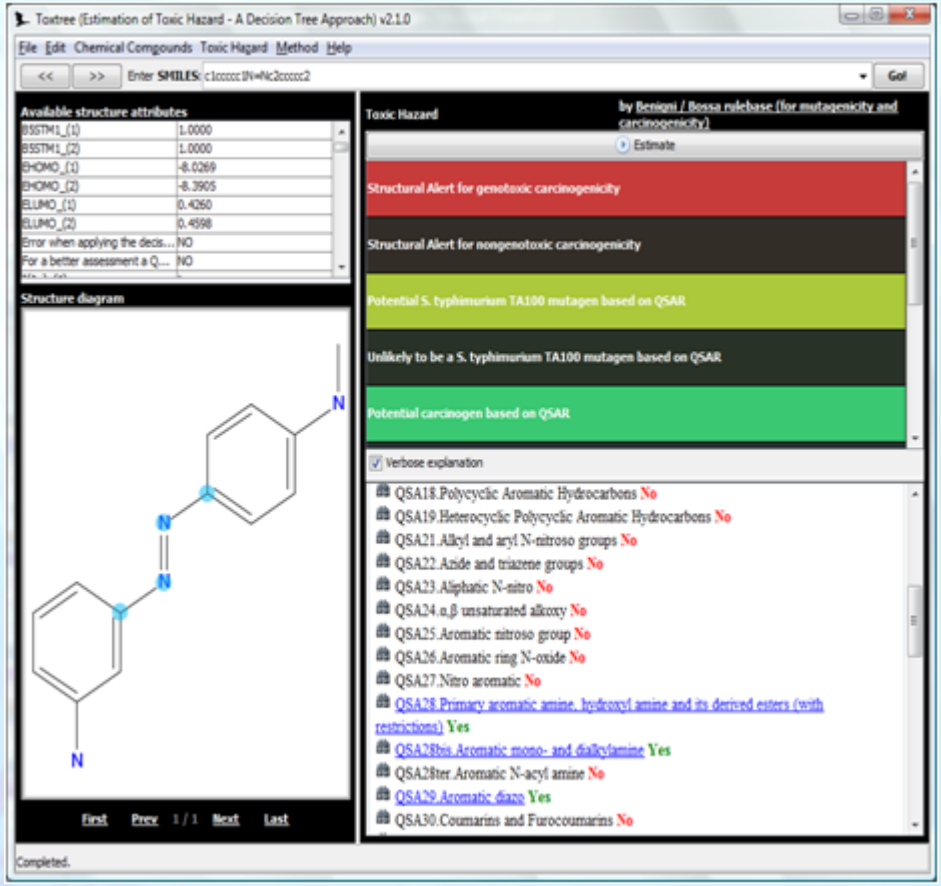


- QMRF repository
- Ambit XT
- Partner in OpenTox FP7 project
- Partner in CADASTER FP7 project

Toxtree 2.1.0

http://toxtree.sourceforge.net





Toxtree was commissioned by JRC Computational Toxicology. Plugins are currently being developed by IdeaConsult and several other organisations.

OpenTox

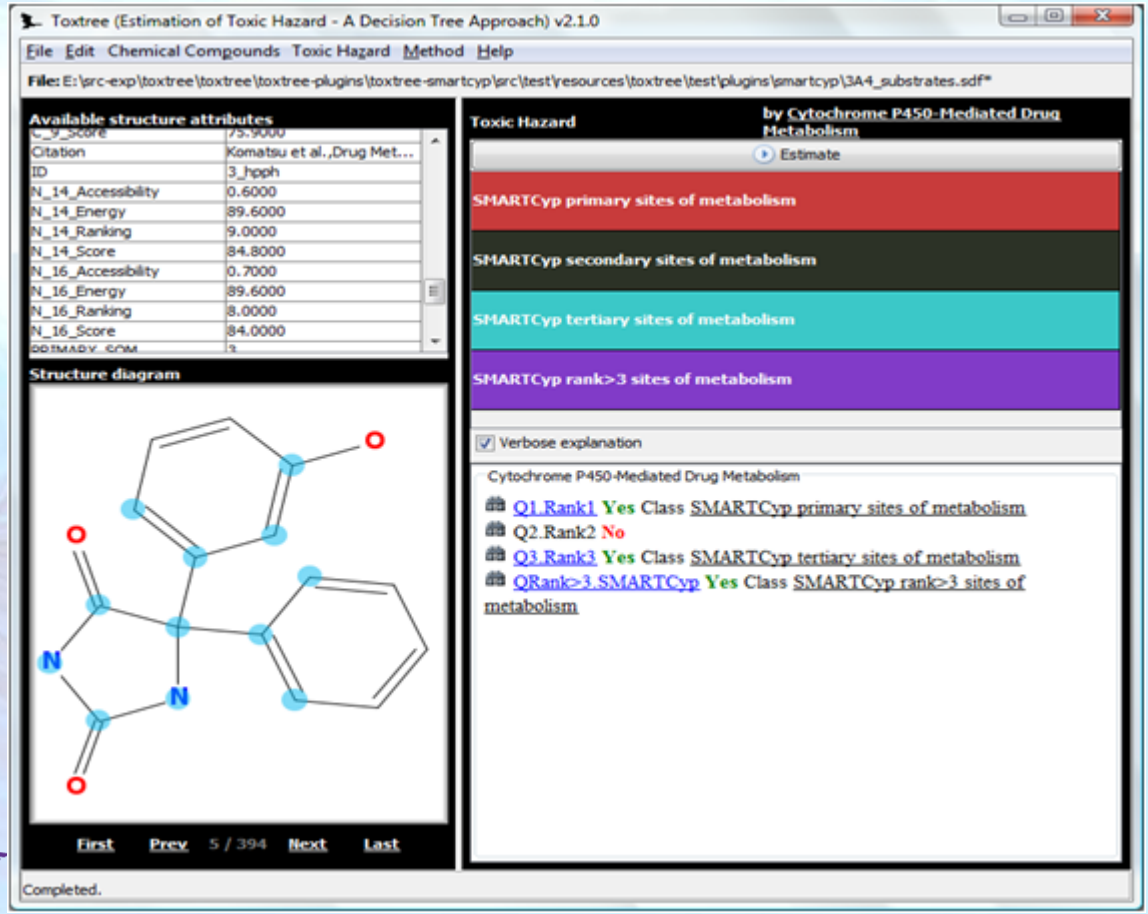
Toxtree 2.1.0 modules:
http://toxtree.sourceforge.net

1. Cramer rules
2. Verhaar scheme for predicting toxicity mode of actions
3. Skin irritation and corrosion potential estimation
4. Eye irritation and corrosion potential estimation
5. Benigni-Bossa rulebase for carcinogenicity and mutagenicity prediction
6. Cramer rules with extensions
7. Structure Alerts for the in vivo micronucleus assay in rodents
8. START (Structural Alerts for Reactivity in Toxtree) biodegradation and persistence plug-in
9. Michael Acceptors identification
10. **Skin sensitisation alerts (new)**
11. **Kroes TTC decision tree (new)**
12. **SMARTCyp - Cytochrome P450-Mediated Drug Metabolism (new)**

OpenTox

Toxtree 2.1.0 : 3 new modules
SMARTCyp - Cytochrome P450-Mediated Drug Metabolism

Patrik Rydberg, David E. Gloriam, Jed Zaretzki, Curt Breneman, Lars Olsen, SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism, ACS Med. Chem. Lett., 2010, 1 (3), pp 96–100



Why integration framework for predictive toxicology?

- What if we would like to:
  - Compare multiple models
  - Reproduce a model from the literature
  - Merge data from different sources (files, databases)
  - Find all models available for certain endpoint
  - More …

OpenTox

Why integration framework for predictive toxicology?

- Challenges:
  - Chemical structures
    - Might be ambiguous
    - Might be error prone or time consuming to reproduce from publications
  - Data
    - **Multiple formats**,
    - **Implicit semantics**, often buried in human readable documentation only
  - Models
    - **Tens of thousands** available, in software or in publications
    - Multiple software solutions, mostly incompatible
    - Predictions **reproducibility** is time consuming and often hard to achieve
    - Automatic **comparison of prediction results** difficult

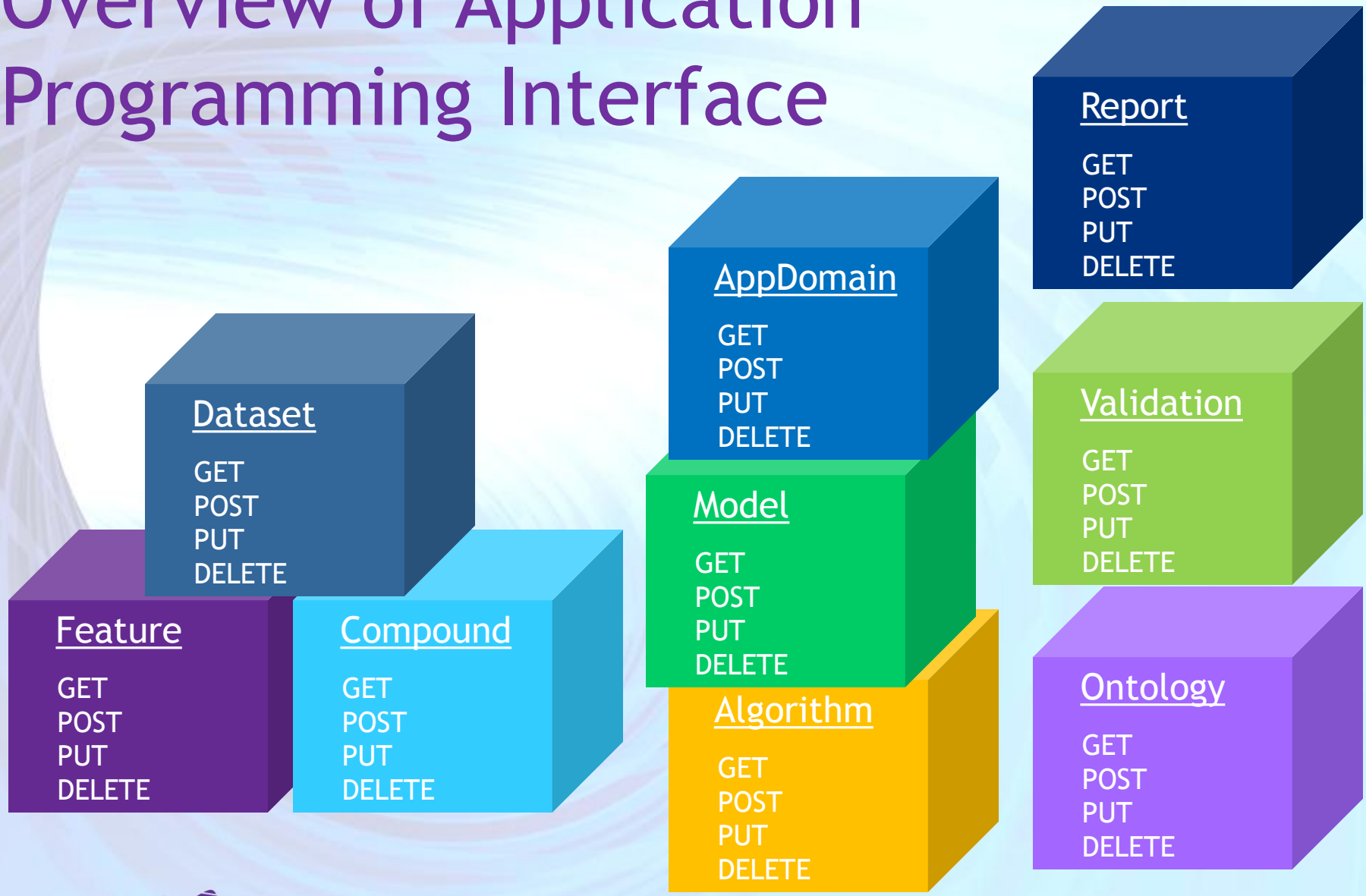
OpenTox Framework approach

User Requirements		Software Requirements
Umambiguous data	⇒	*formal way of representing information about **data***
Unambiguous access	⇒	*well-defined interfaces*
Transparency of computational tools	⇒	*formal way of representing information about **methods**, well-defined interfaces*
Variety of user groups	⇒	*simplicity and modularity of design*
Need to integrate various resources (e.g., databases, prediction methods, models, …) to make meaningful predictions	⇒	*distributed architecture, interoperability*
Need to integrate biological information	⇒	*again, modularity of design, extensibility*

OpenTox framework

- Distributed Web Services for predictive toxicology
- Several types of Web Services (using REST Web service technology)
- Service types corresponds to following building blocks :
  - Chemical compound;
  - Dataset of chemical compounds and their properties (calculated or measured)
  - Algorithm (descriptor calculation, regression, classification, structural alerts, QC, etc.)
  - Predictive model
  - Report, validation, applicability domain, etc.
- Every object (compound, dataset, algorithm, model, etc.) has an unique web address (e.g. http://myhost.com/model/bestpredictivemodel)
- These objects can be created, read, deleted, and updated
- Every object has RDF (W3C Resource Description Framework) representation, defined in OpenTox ontology (opentox.owl)

Overview of Application Programming Interface

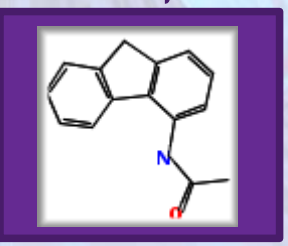


OpenTox

OpenTox datasets: Unified access to data

Everything described by W3C RDF (Resource Description framework)

Compound/ Data	http://myhost.com/feature/21580	http://myhost.com/feature/21589	http://myhost.com/feature/21573	http://myhost.com/feature/21576	http://myhost.com/feature/21588	http://myhost.com/feature/21858	http://myhost.com/feature/22114
http://myhost.com/compound/413	N,N-dimethyl-4-aminoazobenzene	CN(C1=CC=C(C=C1)N=N/C2=CC=CC=C2)C	3	3.31	225.3	YES	3.123
http://myhost.com/compound/44497	4-acetamidofluorene	O=C(Nc3c2c1ccccc1Cc2ccc3)C	1	NP	223.28	YES	2.085
...	...	...	...	...	...		



Dataset

GET

All columns have explicit and machine readable pointers to originating algorithms, models or data

Feature

GET
POST
PUT
DELETE

Compound

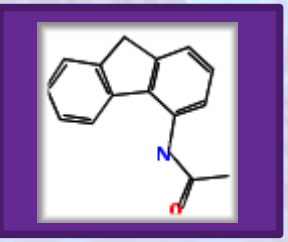
GET
POST
PUT
DELETE

OpenTox

OpenTox datasets: Unified access to data

Everything described by W3C RDF (Resource Description framework)

Compound/ Data	http://myhost.com/feature/21580	http://myhost.com/feature/21589	http://myhost.com/feature/21573	http://myhost.com/feature/21576	http://myhost.com/feature/21588	http://myhost.com/feature/21858	http://myhost.com/feature/22114
http://myhost.com/compound/413	N,N-dimethyl-4-aminoazobenzene	CN(C1=CC=C(C=C1)N=N/C2=CC=CC=C2)C	3	3.31	225.3	YES	3.123
http://myhost.com/compound/44497	4-acetamidofluorene	O=C(Nc3cc1Cc2cc					
...	...	...					



http://myhost.com/feature/21573
af:21573
a ot:Feature , ot:NumericFeature , ot:NominalFeature ;
dc:creator
"http://www.epa.gov/NCCT/dsstox/sdf_isscan_external.html" ;
dc:title "Canc" ;
ot:hasSource "ISSCAN_v3a_1153_19Sept08.1222179139.sdf" ;
= otee:Carcinogenicity .

GET
POST
PUT
DELETE

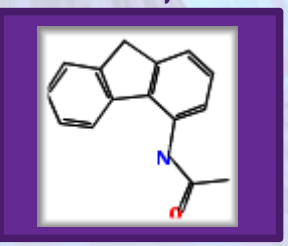
GET
POST
PUT
DELETE

OpenTox

OpenTox datasets: Unified access to data

Everything described by W3C RDF (Resource Description framework)

Compound/ Data	http://myhost.com/feature/21580	http://myhost.com/feature/21589	http://myhost.com/feature/21573	http://myhost.com/feature/21576	http://myhost.com/feature/21588	http://myhost.com/feature/21858	http://myhost.com/feature/22114
http://myhost.com/compound/413	N,N-dimethyl-4-aminoazobenzene	CN(C1=CC=C(C=C1)N=N/C2=CC=CC=C2)C	3	3.31	225.3	YES	3.123
http://myhost.com/compound/44497	4-acetamidofluorene	O=C(Nc3					
...	...						



http://myhost.com/feature/21573

http://myhost.com/feature/21858
dc:title "Structural Alert for genotoxic carcinogenicity"
ot:hasSource
<http://myhost.com/algorithm/Benigni+%2F+Bossa+rulebase+%28for+mutagenicity+and+carcinogenicity> ;

e ;

" ;

df" ;

PUT
DELETE

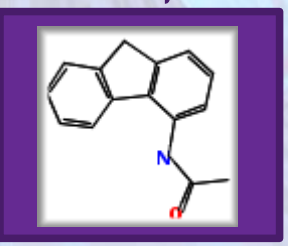
PUT
DELETE

OpenTox

OpenTox datasets: Unified access to data

Everything described by W3C RDF (Resource Description framework)

Compound/ Data	http://myhost.com/feature/21580	http://myhost.com/feature/21589	http://myhost.com/feature/21573	http://myhost.com/feature/21576	http://myhost.com/feature/21588	http://myhost.com/feature/21858	http://myhost.com/feature/22114
http://myhost.com/compound/413	N,N-dimethyl-4-aminoazobenzene	CN(C1=CC=C(C=C1)N=N/C2=CC=CC=C2)C	3	3.31	225.3	YES	3.123
http://myhost.com/compound/44497	4-acetamidofluorene	O=C(Nc3					
...	...						



http://myhost.com/feature/21573

http://myhost.com/feature/22114
a ot:Feature , ot:NumericFeature ;
dc:creator
"http://www.blueobelisk.org/ontologies/chemoinformatics-algorithms/#xlogP" ;
dc:title "XLogP" ;
ot:hasSource
<http://myhost.com/algorithm/org.openscience.cdk.qsar.descriptors.molecular.XLogPDescriptor> ;
= otee:Octanol-water_partition_coefficient_Kow .

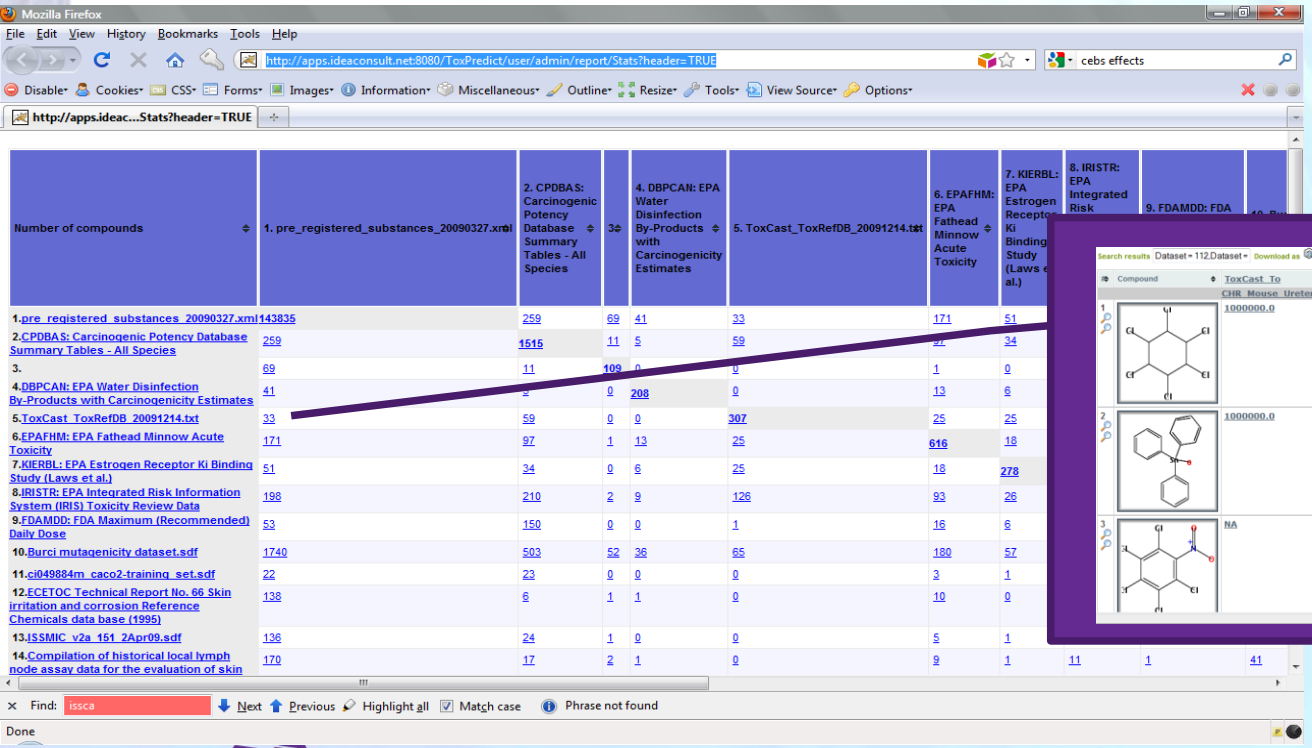
OpenTox

Uniform access to the data

- Ontologies are critical to unambiguously describe data
  - Opentox.owl;
  - Blue Obelisk algorithm ontology (cheminformatics algorithms);
  - OpenTox algotihm types ontology;
  - OpenTox endpoints ontology, based on ECHA endpoints classification;
  - Specific endpoints ontologies, developed by OpenTox partners with toxicology expertise;
  - Existing biomedical ontologies.
- Datasets can be easily merged, compared , and calculations reproduced, regardless of their physical place.
- OWL sameAs construct can be used to denote two links point to the same object
- Authentication and authorization to protect sensitive data

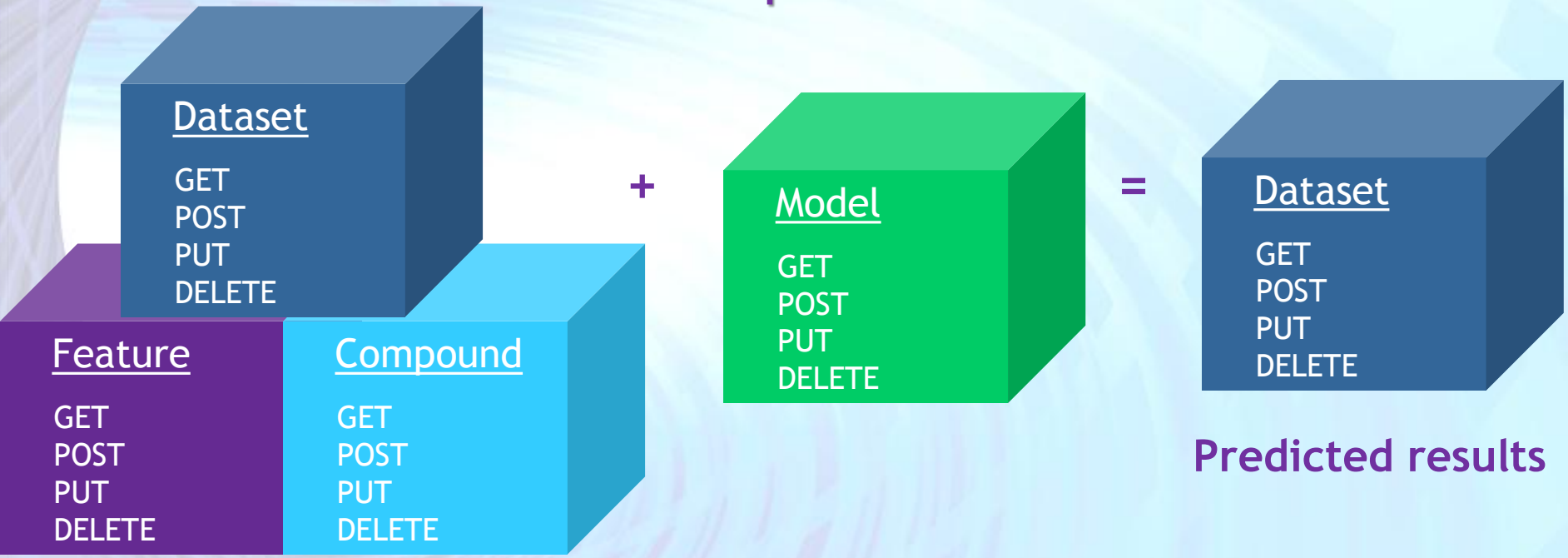
Uniform access to the data

- Datasets can be easily merged, compared , and calculations reproduced, regardless of their physical place.
- The dataset service offers property, compound, substructure and similarity searches via uniform OpenTox Application Programming



Uniform access to calculations

Read data from a web address - process - write to a web address



Predicted results

ttp://myhost.com/dataset/newcompounds

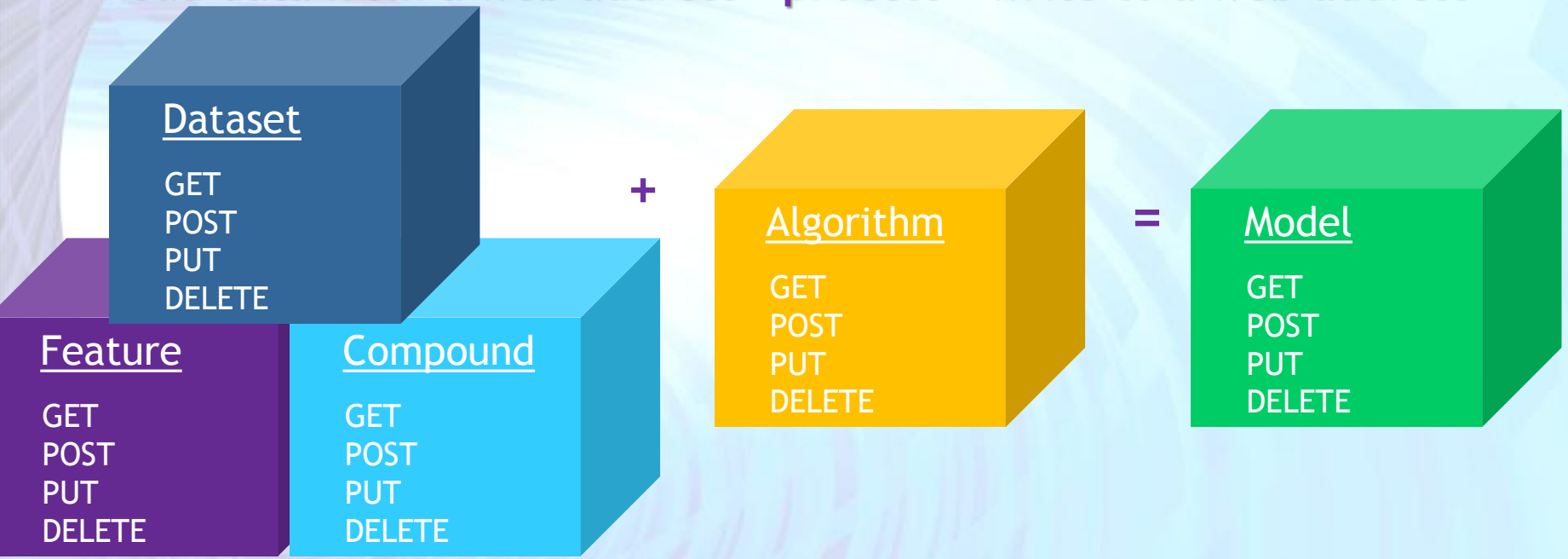
http://myhost.com/model/predictivemodel1

http://myhost.com/dataset/predictedresults1

OpenTox

Uniform approach to models creation

Read data from a web address - process - write to a web address



http://myhost.com/algorithm/neuralnetwork

tp://myhost.com/dataset/trainingset1

http://myhost.com/model/predictivemodel1

OpenTox

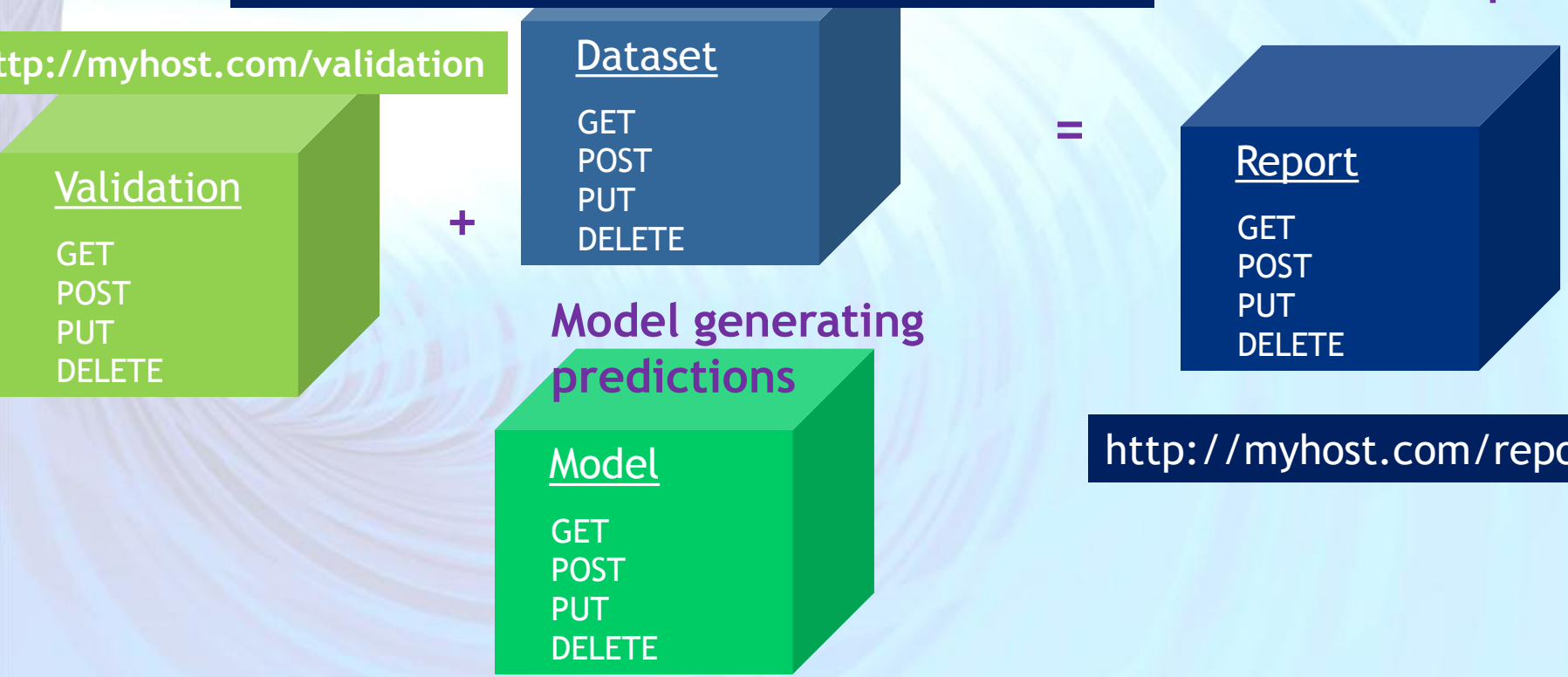
Uniform approach to models validation and report generation

Read data from a web address - process - write to a web address

http://myhost.com/dataset/trainingset1

http://myhost.com/dataset/predictedresults1

Validation report



http://myhost.com/report/1

http://myhost.com/model/predictivemodel1

OpenTox Is A Framework

Framework	• Toxicity data • In silico models • Validation support • Interpretation aids
Unified Access	• Toxicologists • Modellers • API for new algorithm development & integration
Open Source	• To optimise impact • To allow inspection / review • To attract external contributors

OpenTox services can be used to develop specific applications, or embedded in workflow systems

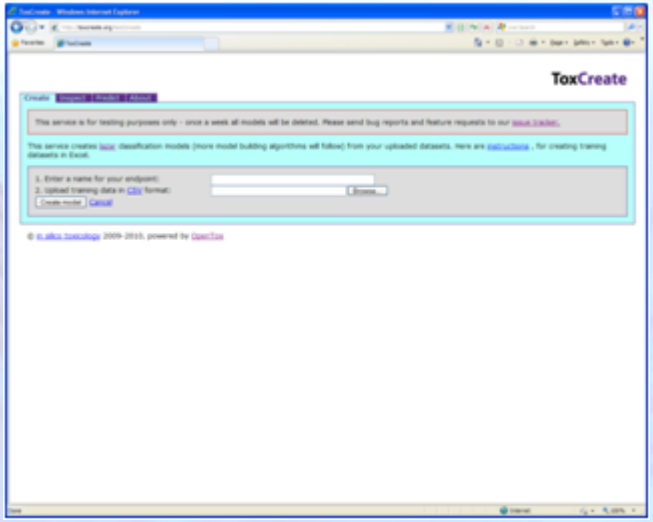
OpenTox

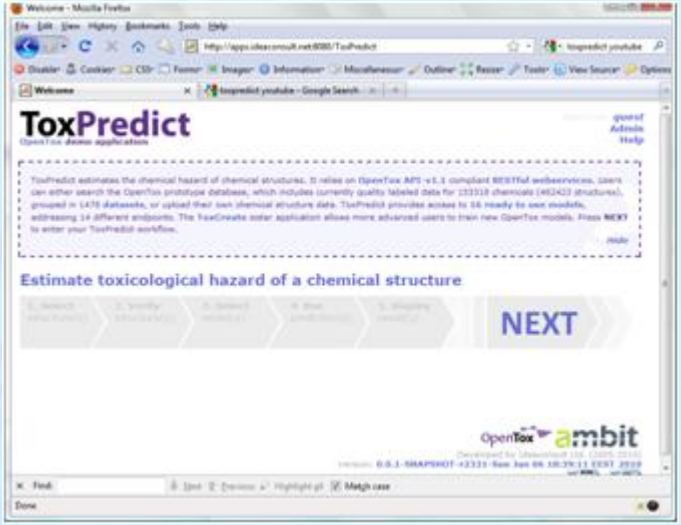
Implementation: key facts

- OpenTox webservices, regardless of their implementation specifics, could be hosted on a single computer (even a laptop or netbook would do) or could be transparently distributed on multiple servers in various physical locations, in particular for better reliability, resilience and scalability;
- OpenTox webservices could be managed by a single or multiple administrative authorities; a key feature of the OpenTox framework is that it has been **designed in a multi-domain friendly way**, which is essential for data and model sharing, repeatability and validation of prediction results (a must in 21st century!);
- OpenTox webservices could be deployed behind firewalls, in Intranets (or even offline), when very tight security policies would have to be met;
- Third parties, willing to deploy OpenTox webservices could select a relevant subset of services to run, tailored to their specific needs;

Implementation: key facts

- Two end user oriented demo applications, making use of OpenTox webservices, have been developed, deployed and are available for testing – ToxCreate and ToxPredict;
- ToxCreate creates models from user supplied datasets;
- ToxPredict uses existing OpenTox models to estimate chemicalcompound properties



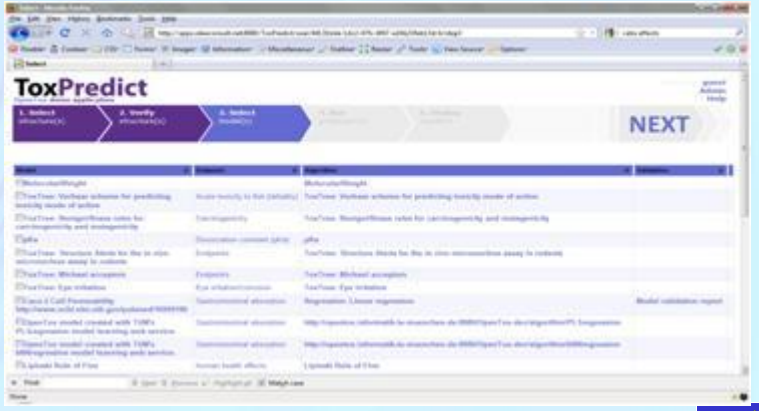


ToxPredict: demo

Screencast at http://www.youtube.com/watch?v=H7gkNBz4Lwo

ToxPredict estimates the chemical hazard of chemical structures. It relies on OpenTox API-v1.1 compliant RESTful webservices. Users can either search the OpenTox prototype database, which includes currently quality labelled data for **~150,000 chemicals**, grouped in more than a dozen datasets, or upload their own chemical structure data. ToxPredict provides access to **16 ready to use models**, addressing **14 different endpoints** (and growing!);

- ToxPredict uses the following OpenTox webservices: Compound, Feature, Dataset, Algorithm, Model, Task and Ontology;
- Any model, available via OpenTox API can be easily integrated in ToxPredict by just publishing its Web address



…and now let the show begin!

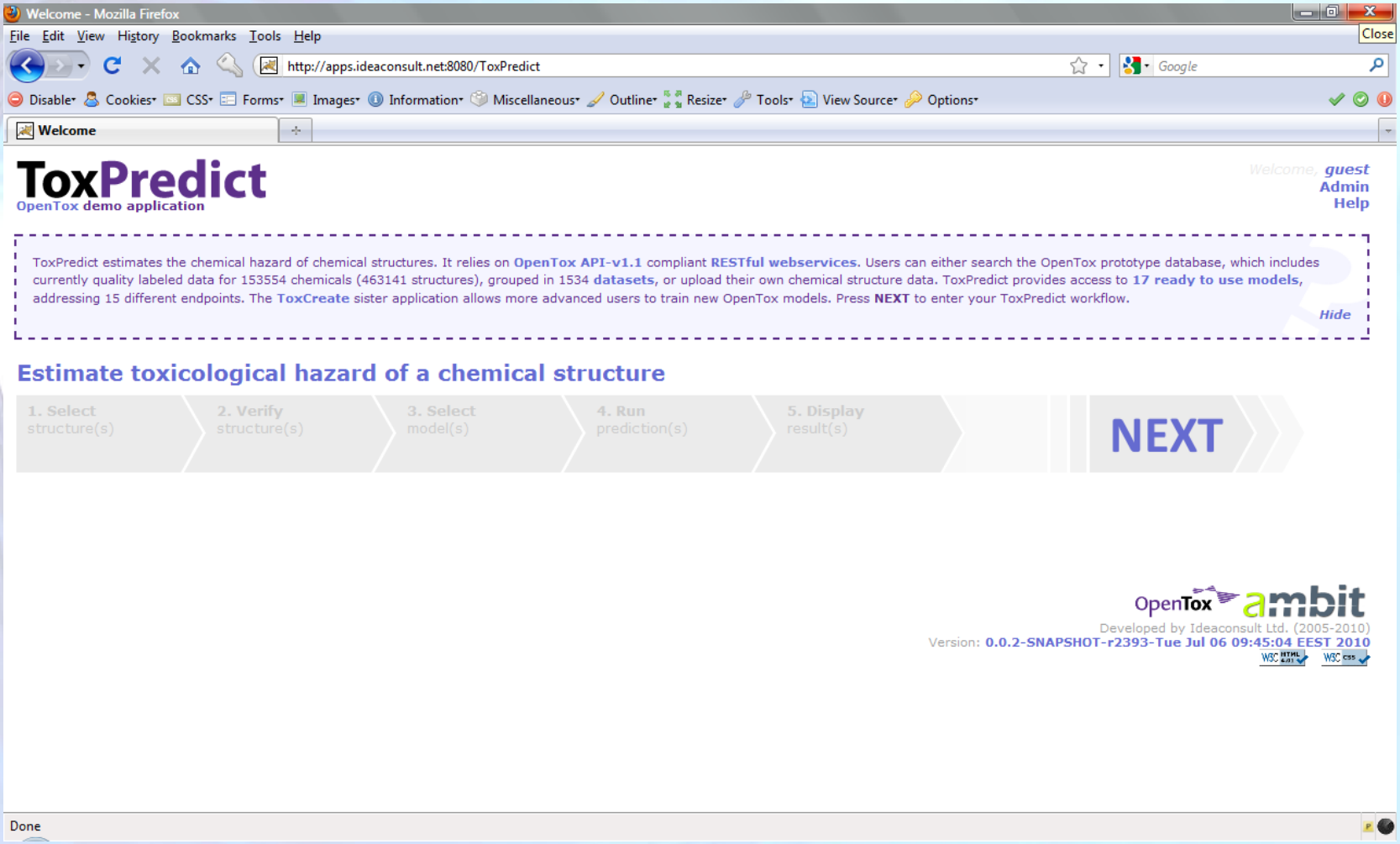
After having followed this rather technical presentation, please enjoy the live demo and feel free to try how it works at toxpredict.org

◊ ◊ ◊

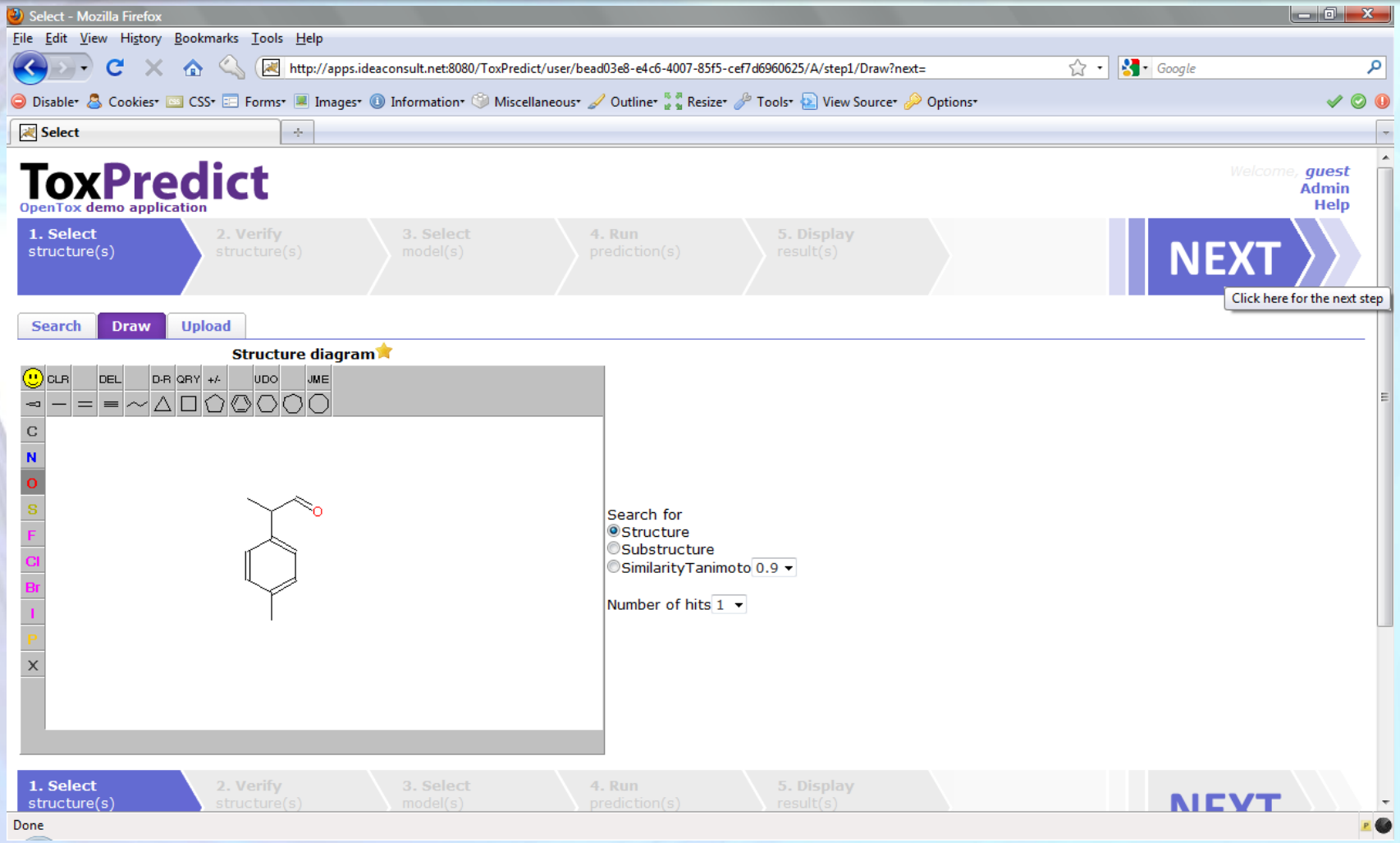
We would also appreciate if you could submit any comments, suggestions, bug reports, feature requests and/or other relevant input you might have to our issue tracker or mailing list.

Many thanks for your attention!

ToxPredict


Welcome - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Close
http://apps.ideaconsult.net:8080/ToxPredict
Google
Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options
Welcome
ToxPredict
OpenTox demo application
Welcome, guest
Admin
Help
ToxPredict estimates the chemical hazard of chemical structures. It relies on OpenTox API-v1.1 compliant RESTful webservices. Users can either search the OpenTox prototype database, which includes currently quality labeled data for 153554 chemicals (463141 structures), grouped in 1534 datasets, or upload their own chemical structure data. ToxPredict provides access to 17 ready to use models, addressing 15 different endpoints. The ToxCreate sister application allows more advanced users to train new OpenTox models. Press NEXT to enter your ToxPredict workflow.
Hide
Estimate toxicological hazard of a chemical structure
1. Select structure(s)
2. Verify structure(s)
3. Select model(s)
4. Run prediction(s)
5. Display result(s)
NEXT
OpenTox ambit
Developed by Ideaconsult Ltd. (2005-2010)
Version: 0.0.2-SNAPSHOT-r2393-Tue Jul 06 09:45:04 EEST 2010
Done

ToxPredict



OpenTox

Ideaconsult Ltd.

ToxPredict

Verify - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://apps.ideaconsult.net:8080/ToxPredict/user/bead03e8-e4c6-4007-85f5-cef7d6960625/A/step2

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

Verify

ToxPredict
OpenTox demo application

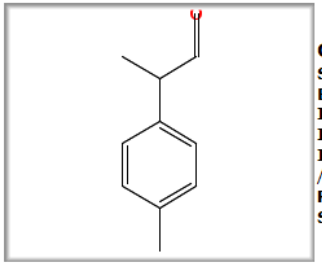
Welcome, *guest*
Admin
Help

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

« Page 1 Records per page 1 »

Structure(s) & Experimental Data SDF CSV PDF

1.



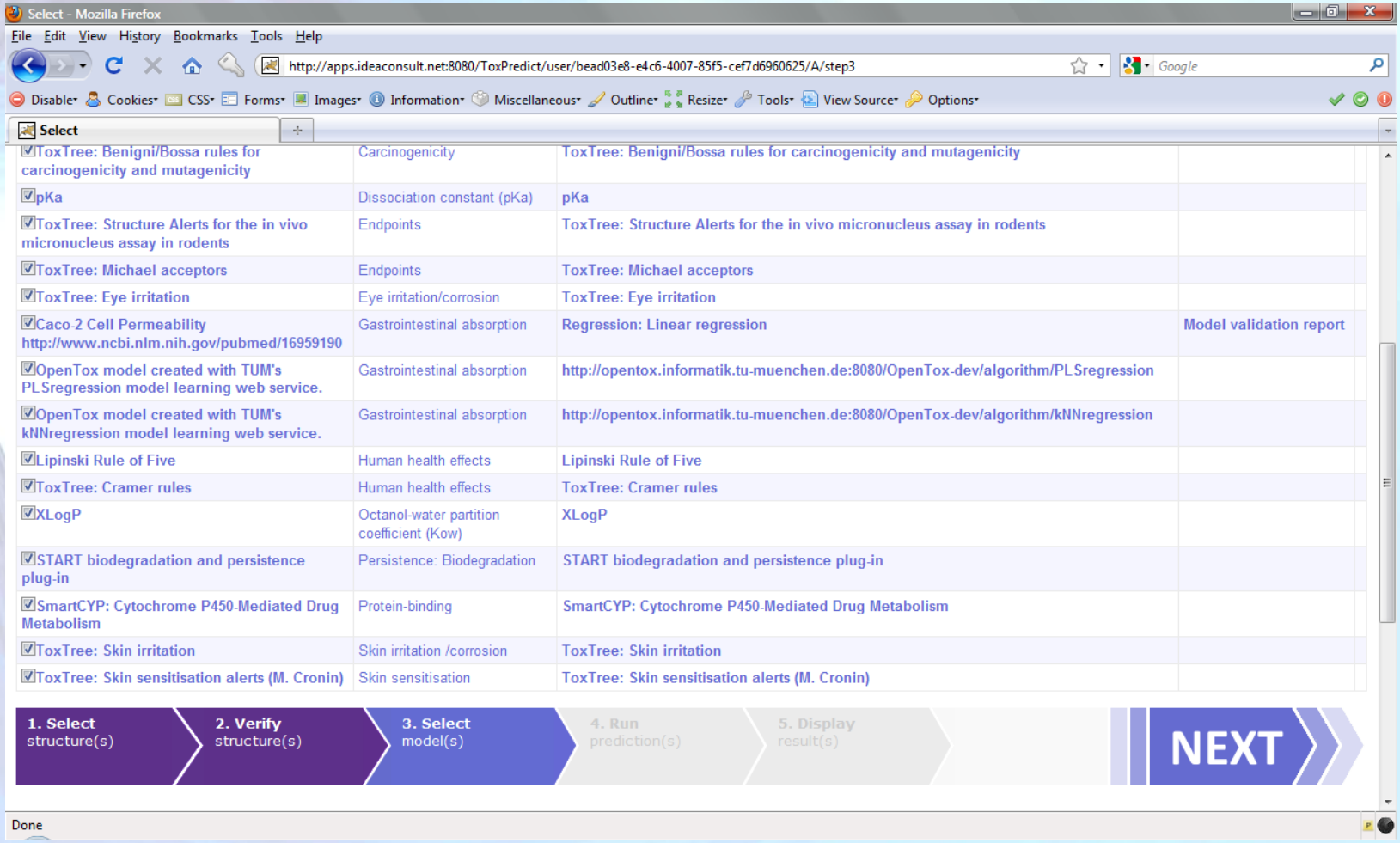
CASRN 99-72-9
Synonym(s) 2-(p-tolyl)propionaldehyde
EINECS 202-782-0
IUPAC name 2-(4-methylphenyl)propanal
InChIKey_std JTZWVKUZBNHSSW-SECBINFHSA-N,JTZWVKUZBNHSSW-UHFFFAOYSA-N
InChI_std InChI=1S/C10H12O/c1-8-3-5-10(6-4-8)9(2)7-11/h3-7,9H,1-2H3/t9-/m1/s1,InChI=1S/C10H12O/c1-8-3-5-10(6-4-8)9(2)7-11/h3-7,9H,1-2H3
REACHRegistrationDate 30.11.2010
SMILES CC(C=O)c1ccc(C)cc1

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

OpenTox ambit
Developed by Ideaconsult Ltd. (2005-2010)
Version: 0.0.2-SNAPSHOT-r2393-Tue Jul 06 09:45:04 EEST 2010

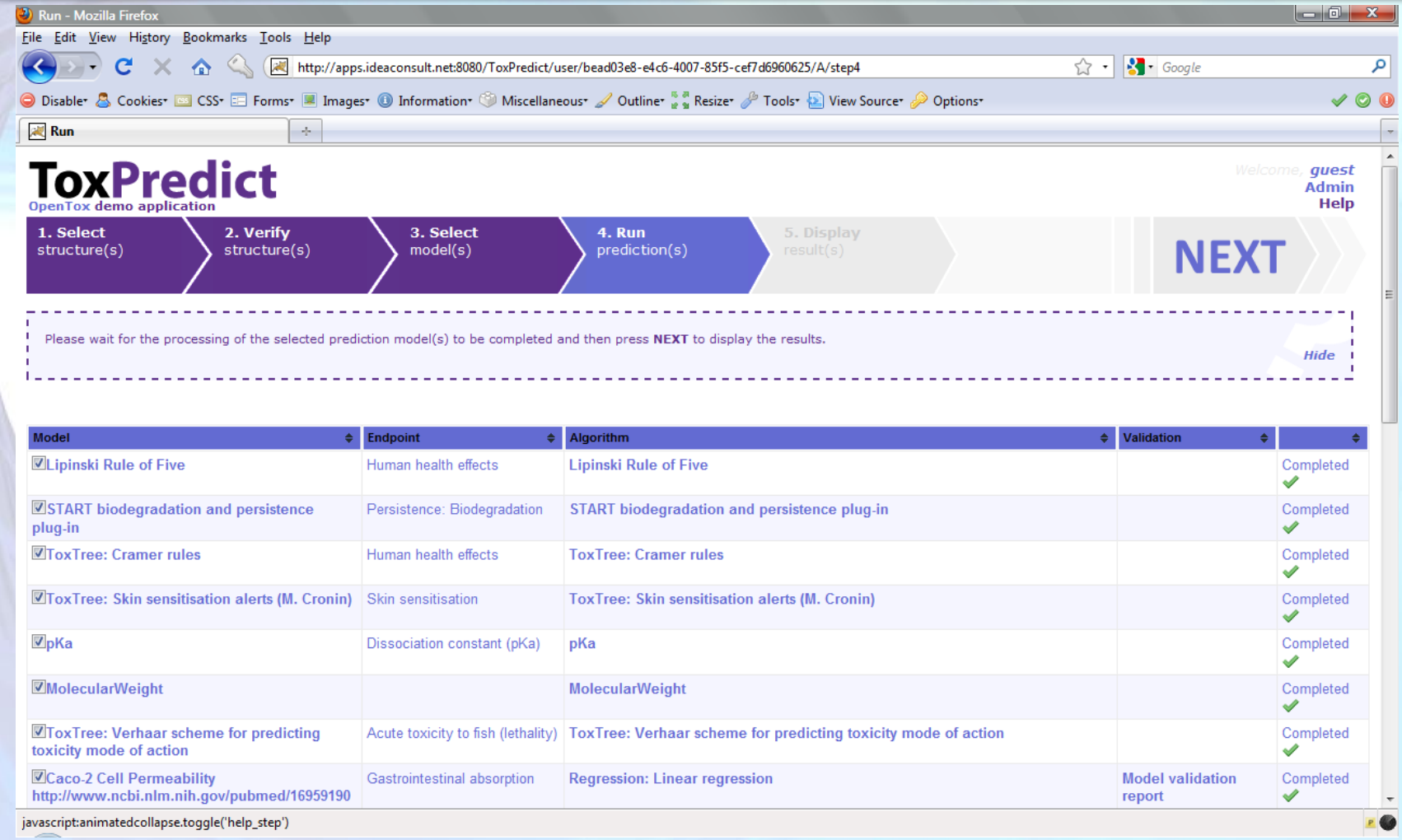
Done

ToxPredict



OpenTox

ToxPredict


Run - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://apps.ideaconsult.net:8080/ToxPredict/user/bead03e8-e4c6-4007-85f5-cef7d6960625/A/step4
Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options
Run
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
Please wait for the processing of the selected prediction model(s) to be completed and then press NEXT to display the results.
Hide
Model
Endpoint
Algorithm
Validation
Lipinski Rule of Five
Human health effects
Lipinski Rule of Five
Completed
START biodegradation and persistence plug-in
Persistence: Biodegradation
START biodegradation and persistence plug-in
Completed
ToxTree: Cramer rules
Human health effects
ToxTree: Cramer rules
Completed
ToxTree: Skin sensitisation alerts (M. Cronin)
Skin sensitisation
ToxTree: Skin sensitisation alerts (M. Cronin)
Completed
pKa
Dissociation constant (pKa)
pKa
Completed
MolecularWeight
MolecularWeight
Completed
ToxTree: Verhaar scheme for predicting toxicity mode of action
Acute toxicity to fish (lethality)
ToxTree: Verhaar scheme for predicting toxicity mode of action
Completed
Caco-2 Cell Permeability http://www.ncbi.nlm.nih.gov/pubmed/16959190
Gastrointestinal absorption
Regression: Linear regression
Model validation report
Completed
javascript:animatedcollapse.toggle('help_step')

ToxPredict

Display - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://apps.ideaconsult.net:8080/ToxPredict/user/bead03e8-e4c6-4007-85f5-cef7d6960625/A/step5#

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

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)

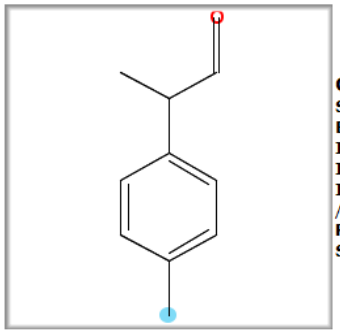
This page lists your ToxPredict workflow results for the structure(s) you have selected and the model prediction(s) you have chosen to run. You could also retrieve the ToxPredict report in various other formats, e.g. SDF, CML, SMI, PDF, CSV, ARFF, RDF/XML or RDF/N3.

Hide

« Page 1 Records per page 1 »

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

1.



CASRN 99-72-9
Synonym(s) 2-(p-tolyl)propionaldehyde
EINECS 202-782-0
IUPAC name 2-(4-methylphenyl)propanal
InChIKey_std JTZWVKUZBNHSSW-SECBINFHSA-N,JTZWVKUZBNHSSW-UHFFFAOYSA-N
InChI_std InChI=1S/C10H12O/c1-8-3-5-10(6-4-8)9(2)7-11/h3-7,9H,1-2H3/t9-/m1/s1,InChI=1S/C10H12O/c1-8-3-5-10(6-4-8)9(2)7-11/h3-7,9H,1-2H3
REACHRegistrationDate 30.11.2010
SMILES CC(C=O)c1ccc(C)cc1

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

Prediction feature for http://apps.ideaconsult.net:8080/ambit2/feature/22200 endpoint prediction -4.53439998626709

LipinskiFailures Lipinski Rule of Five

LipinskiFailures 0.0

Done

OpenTox