

# Data Bases and Data Modeling

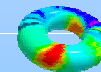
J. Gasteiger

Computer-Chemie-Centrum

Universität Erlangen-Nürnberg

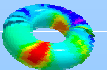
D-91052 Erlangen

<http://www2.chemie.uni-erlangen.de>

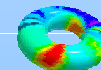
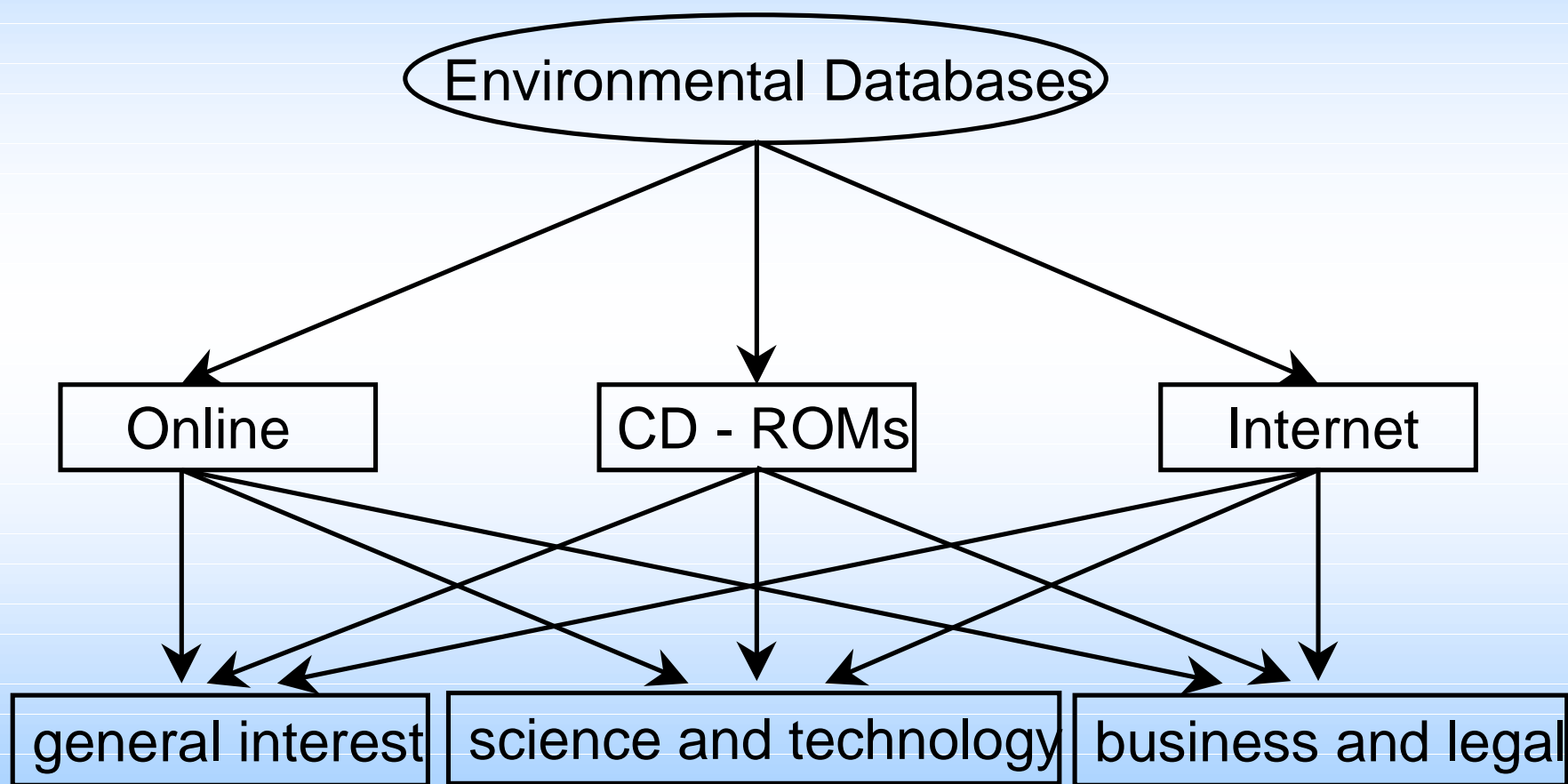


# Overview

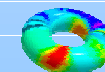
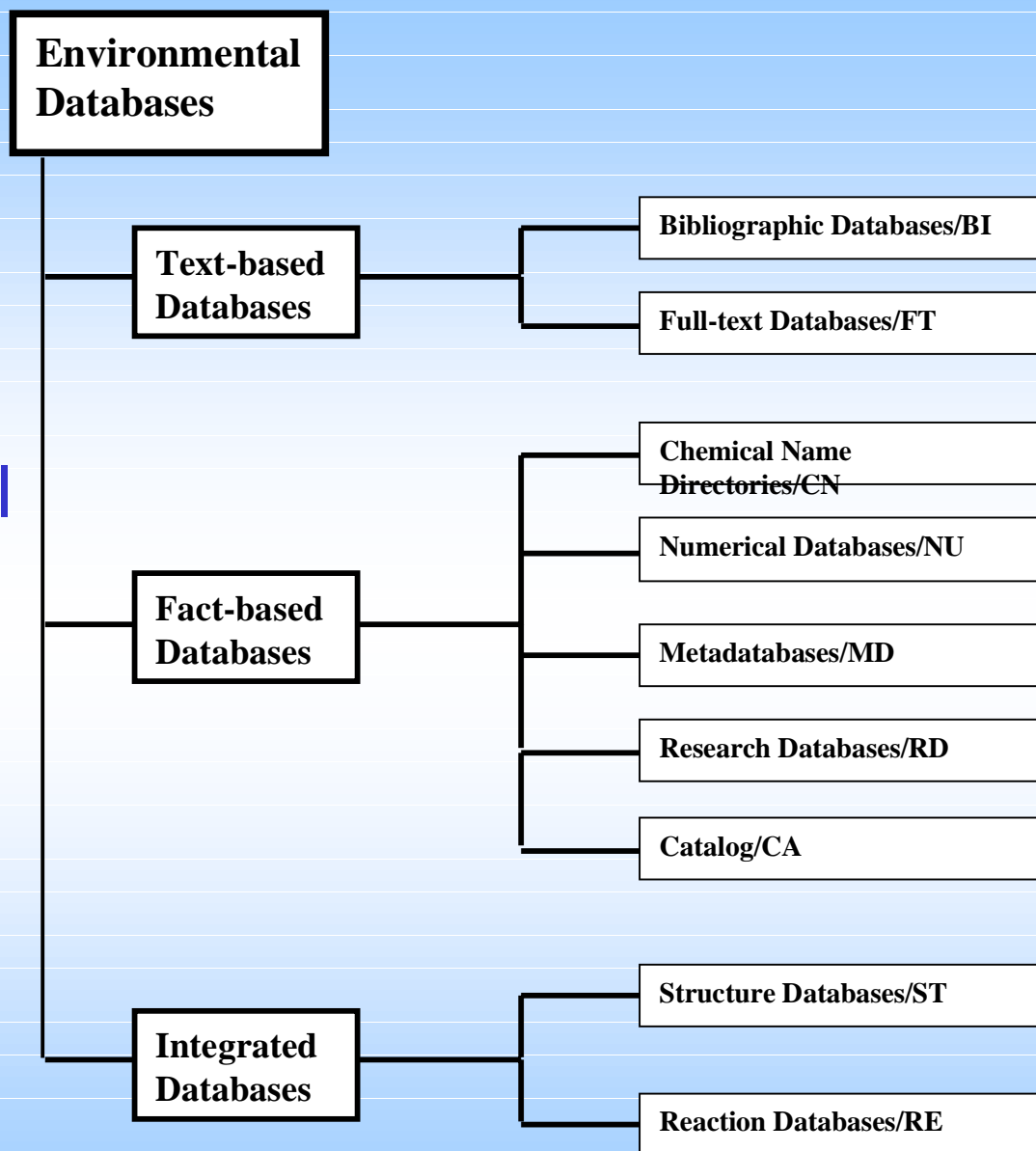
- types of environmental databases
- selection of environmental databases
- metadatabases
- Google search on atrazine
- toxicity information
- prediction of toxic mode of action



# Categories of Environmental Databases

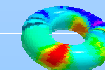


# Types of Environmental Databases



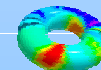
# Environmental Bibliographic Databases

<b>databas e</b>	<b>hosts</b>	<b>free internet</b>	<b>no. of documents</b>
Enviroline	DataStar, DIALOG, DIMDI	No	278,000 (04/02)
Pollutio n	DIALOG, STN	No	284,000 (04/02)
ULIT	DataStar, FIZ-Technik, GBI, STN	<i><a href="http://isis.uba.de:3001/UN/opacs.html">http://isis.uba.de:3001/UN/opacs.html</a></i>	425,000 (11/01)



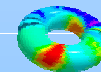
# Selection of Numerical Databases

- Biocatalysis/Biodegradation Database
- Chemicals Information System for Consumer-relevant Substances (CIVS)
- ECOTOX (several databases)
- Envirofacts (several databases)
- Environmental Fate Database (several databases)
- Environmental Health Criteria Monographs (EHCs)



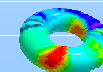
# Selection of Numerical Databases

- Biocatalysis/Biodegradation Database
- EXTOXNET (several databases)
- HSDB
- NRA Chemical Review Program (profiles)
- Pesticide Database, Japan (profiles)
- SIDS (profiles)
- UmweltInfo



# Environmental Databases on the Free Internet

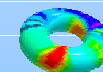
Name	URL	File type	Comments
ULIT	<a href="http://isis.uba.de:3001/UN/opacs.html">http://isis.uba.de:3001/UN/opacs.html</a>	BI	German, English
UFOR	<a href="http://isis.uba.de:3001/VH/opacs.html">http://isis.uba.de:3001/VH/opacs.html</a>	RD	German, English
ECOTOX	<a href="http://www.epa.gov/ecotox/">http://www.epa.gov/ecotox/</a>	NU	Peer-reviewed data
HSDB	<a href="http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB">http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB</a>	NU	Peer-reviewed data
ChemFinder	<a href="http://chemfinder.cambridgesoft.com/">http://chemfinder.cambridgesoft.com/</a>	CN	Structural formula
ChemIDplus	<a href="http://chem.sis.nlm.nih.gov/chemidplus/">http://chem.sis.nlm.nih.gov/chemidplus/</a>	CN	Structural formula
CORDIS Research Database	<a href="http://dbs.cordis.lu/search/en/simple/EN_PROJ_simple.html">http://dbs.cordis.lu/search/en/simple/EN_PROJ_simple.html</a>	RD	Several languages
Sigma, Aldrich	<a href="http://www.sigmaaldrich.com/">http://www.sigmaaldrich.com/</a>	CA	90,000 MSDSs
Fisher Scientific	<a href="http://www.fishersci.com">http://www.fishersci.com</a>	CA	Several companies



# DAIN (metaDAtabase of INternet resources)

<http://www.wiz.uni-kassel.de/dain/>

<b>topic</b>	<b>total number of databases</b>	<b>no. of fact-based databases</b>	<b>no. of text-based databases</b>
Pesticides	48	34	14
Solvents	8	8	0
Warfare Agents	12	12	0
Pharmaceuticals	11	10	1
Polymers	9	8	1
Ecotoxicity	8	8	0
Toxicity	35	27	8
Biodegradation	7	6	1
Detection in air	44	44	0
Physical-chemical parameters	18	17	1



Google-Suche: atrazin - Mozilla

Google  Suche [Erweiterte Suche](#) [Einstellungen](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

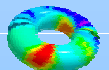
**Web** Ergebnisse 1 - 10 von ungefähr 8,060 für atrazin. (0.14 Sekunden)

[Das Stichwort "Atrazin" im Sonnenseite Umweltlexikon](#)  
... **Atrazin** Von den Herbiziden (Unkrautvernichtungsmitteln) der bekannteste und bisher, vor allem im Maisanbau, am häufigsten eingesetzte Wirkstoff. ...  
[www.sonnenseite.oekoserve.net/fp/archiv/RUBwerkstoffmaterials substanz/ Atrazin.php](http://www.sonnenseite.oekoserve.net/fp/archiv/RUBwerkstoffmaterials substanz/ Atrazin.php) - 37k - [Im Cache](#) - [Ähnliche Seiten](#)

[Atrazin - Lexikon](#)  
**Atrazin** ist ein Pflanzenbehandlungsmittel. Die Gründe ... **Atrazin**. Erklärungen und Begriffe aus dem Lexikon (Glossar). Das Schlagwortverzeichnis ...  
[www.umweltdatenbank.de/lexikon/atrazin.htm](http://www.umweltdatenbank.de/lexikon/atrazin.htm) - 10k - [Im Cache](#) - [Ähnliche Seiten](#)

[Nie wieder Atrazin und Simazin](#)  
PAN Germany. Drucken. Nie wieder **Atrazin** und Simazin. Lange hat es gedauert, aber jetzt hat der Umweltschutz offensichtlich doch gegen ...  
[www.pan-germany.org/info/gesetz-eu/atrazin.htm](http://www.pan-germany.org/info/gesetz-eu/atrazin.htm) - 13k - [Im Cache](#) - [Ähnliche Seiten](#)

[Atrazin](#)  
... Notfalldepots ] [ Pflanzenschutzmittel ] [ Recht ] [ Toxikologie ]. Zurück  
Nach oben Weiter. **Atrazin**. Handelsnamen: Aatrex, Actinite, Aktikon ...  
[www.gifte.de/atrazin.htm](http://www.gifte.de/atrazin.htm) - 17k - [Im Cache](#) - [Ähnliche Seiten](#)



Google-Suche: atrazine - Mozilla

Web Bilder Groups Verzeichnis News

Google atrazine Suche [Erweiterte Suche](#) [Einstellungen](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

**Web** Ergebnisse 1 - 10 von ungefähr 170,000 für **atrazine**. (0.48 Sekunden)

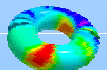
[US EPA - Documents for Atrazine](#) - [ [Diese Seite übersetzen](#) ]  
**Atrazine**. Chemical Review Manager: Eric Olson (703) 308-8067. Many ... Sources".  
**Atrazine** IRED Questions and Answers (January 31, 2003). ...  
[www.epa.gov/oppsrrd1/reregistration/atrazine/](http://www.epa.gov/oppsrrd1/reregistration/atrazine/) - 28k - [Im Cache](#) - [Ähnliche Seiten](#)

[EPA Ground Water & Drinking Water > breadcrumb? > Consumer ...](#) - [ [Diese Seite übersetzen](#) ]  
 ... Consumer Factsheet on: **ATRAZINE**. List of Contaminants. ... What is **Atrazine** and how is it used? **Atrazine** is a white, crystalline solid organic compound. ...  
[www.epa.gov/safewater/dwh/c-soc/atrazine.html](http://www.epa.gov/safewater/dwh/c-soc/atrazine.html) - 19k - 2. Mai 2004 - [Im Cache](#) - [Ähnliche Seiten](#)  
 [ [Weitere Ergebnisse von www.epa.gov](#) ]

[NRDC: EPA Won't Restrict Toxic Herbicide Atrazine, Despite Health ...](#) - [ [Diese Seite übersetzen](#) ]  
**Atrazine**, the nation's most widely used weed-killer, is linked to sexual abnormalities in frogs and high rates of prostate cancer among workers at an **atrazine** ...  
[www.nrdc.org/health/pesticides/natrazine.asp](http://www.nrdc.org/health/pesticides/natrazine.asp) - 28k - 2. Mai 2004 - [Im Cache](#) - [Ähnliche Seiten](#)

[Atrazine Poisoning Worse Than Suspected](#) - [ [Diese Seite übersetzen](#) ]  
 ... USERNAME, PASSWORD, **Atrazine** Poisoning Worse Than Suspected. Controversy erupted over new findings that **atrazine** may be linked to global demise of frogs. Prof. ...  
[www.i-sis.org.uk/atrazine.php](http://www.i-sis.org.uk/atrazine.php) - 24k - 2. Mai 2004 - [Im Cache](#) - [Ähnliche Seiten](#)

[EXTOXNET PIP - ATRAZINE](#) - [ [Diese Seite übersetzen](#) ]  
 ... EXTOXNET primary files maintained and archived at Oregon State University. Revised June 1996. **Atrazine**. ... **Atrazine** is toxicity class III - slightly toxic. ...  
[extoxnet.orst.edu/pips/atrazine.htm](http://extoxnet.orst.edu/pips/atrazine.htm) - 13k - [Im Cache](#) - [Ähnliche Seiten](#)



# U.S. EPA: Atrazine: Interim Reregistration Eligibility Decision

The screenshot shows a Microsoft Internet Explorer browser window displaying the U.S. Environmental Protection Agency's website for Atrazine reregistration. The page title is "Pesticides: Reregistration" and the specific chemical is "Atrazine". The page includes a search bar, a list of recent additions, and a section for "Find Electronic Dockets" with links to various documents and press releases.

**U.S. Environmental Protection Agency**

## Pesticides: Reregistration

Recent Additions | Contact Us | Print Version Search:

EPA Home > Pesticides > Resampling Pesticides > Reregistration > Chemical Status > Atrazine

### Atrazine

Chemical Review Manager: [Eric Olson](#) (703) 308-6067

Many documents are in PDF format and require the [Acrobat Reader](#). New documents may be available from [EDockets](#), EPA's electronic docket system. To find open dockets related to this chemical, click the button below.

- [October 31, 2003 Revised Atrazine IRED](#) (690 KB PDF)
  - [IRED Appendices](#) (1.6 MB ZIP file - may be expanded with [South Ecuador](#) or [Winnipeg](#))
  - [IRED Q and A's](#)
  - [IRED Press Announcement](#)
- [January 2003 Atrazine IRED](#) (amended March 1, 2004) (4200 KB PDF)

[January 31, 2003 Press Release](#): "EPA Adopts Aggressive Measures on Herbicide Atrazine - Approach Ensures Protection of Nation's most Vulnerable Drinking Water Sources"

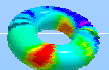
[Atrazine IRED Questions and Answers](#) (January 31, 2003)

Slides from April 16, 2002 Technical Briefing ([Agenda](#) 64 KB PDF):

- [DownPoint file](#) (1,376 KB)
- [EOP file](#) (5,269 KB)

[Atrazine Summary](#) (48 KB PDF)

[Atrazine Overview](#) (139 KB PDF)

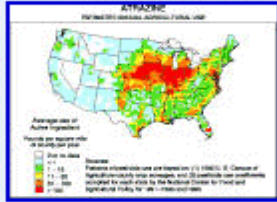


Google-Suche: atrazine - Mozilla


Google Bilder [Web](#) [Bilder](#) [Groups](#) [Verzeichnis](#) [News](#)

atrazine  [Erweiterte Bildsuche](#)  
[Einstellungen](#)

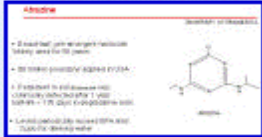
**Bilder** Ergebnisse 1 - 20 von ungefähr 910 für atrazine. (0.10 Sekunden)  
 Anzeigen: **Alle Größen** - [groß](#) - [mittel](#) - [klein](#)




**atrazine.gif**  
632 x 495 Pixel - 19k  
[www.cee.vt.edu/.../gwprimer/agric/atrazine.gif](http://www.cee.vt.edu/.../gwprimer/agric/atrazine.gif)




**atrazine.gif**  
431 x 402 Pixel - 5k  
[ipmworld.umn.edu/chapters/wareherb.htm](http://ipmworld.umn.edu/chapters/wareherb.htm)




**atrazine.gif**  
507 x 282 Pixel - 5k  
[www.cbs.umn.edu/cbri/lisa/web/atrazine.html](http://www.cbs.umn.edu/cbri/lisa/web/atrazine.html)



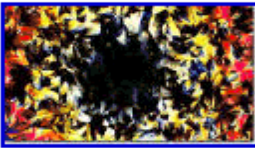
**bioaug1.gif**  
609 x 402 Pixel - 10k  
[www.cbs.umn.edu/cbri/lisa/web/bioaug.html](http://www.cbs.umn.edu/cbri/lisa/web/bioaug.html)  
[ [Weitere Ergebnisse von www.cbs.umn.edu](http://www.cbs.umn.edu) ]



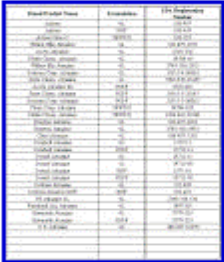
**atrazine.gif**  
949 x 548 Pixel - 19k  
[water.usgs.gov/.../results/bubble/atrazine.gif](http://water.usgs.gov/.../results/bubble/atrazine.gif)



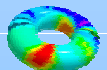
**atrazine3d.gif**  
500 x 305 Pixel - 23k  
[www.chem.ox.ac.uk/mom/lindane/lindane.html](http://www.chem.ox.ac.uk/mom/lindane/lindane.html)



**atrazine.jpg**  
489 x 290 Pixel - 52k  
[micro.magnet.fsu.edu/.../pages/atrazine.html](http://micro.magnet.fsu.edu/.../pages/atrazine.html)



**atrazine.gif**  
579 x 749 Pixel - 15k  
[pested.unl.edu/record.htm](http://pested.unl.edu/record.htm)

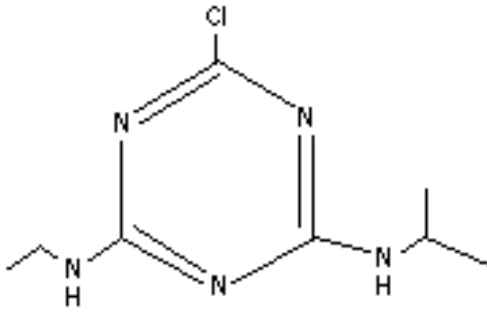


atrazine - Mozilla

## Atrazine

UNIVERSITY OF MINNESOTA

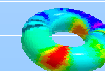
- Broad-leaf, pre-emergent herbicide  
Widely used for 50 years
- 80 million pounds/yr applied in USA
- Persistent in soil (Radosevich 1993):  
commonly detected after 1 year  
half-life > 170 days in degradative soils
- Levels periodically exceed EPA limit  
3 ppb for drinking water



Atrazine

[<previous page>](#) | [<main menu>](#) | [<next page>](#)

Done



Google-Suche: atrazine - Mozilla

**atrazine.jpg**  
280 x 209 Pixel - 10k  
[turfscience.com/weeds/broadleaf.html](http://turfscience.com/weeds/broadleaf.html)

**atrazine.jpg**  
700 x 534 Pixel - 110k  
[www.stjoeriver.net/images/atrazine.jpg](http://www.stjoeriver.net/images/atrazine.jpg)

**atr\_map2.gif**  
609 x 747 Pixel - 17k  
[umbbd.ahc.umn.edu/atr/atr\\_image\\_map2.html](http://umbbd.ahc.umn.edu/atr/atr_image_map2.html)

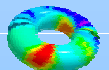
**atrazine.gif**  
207 x 185 Pixel - 1k  
[www.alanwood.net/pesticides/atrazine.html](http://www.alanwood.net/pesticides/atrazine.html)

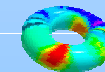
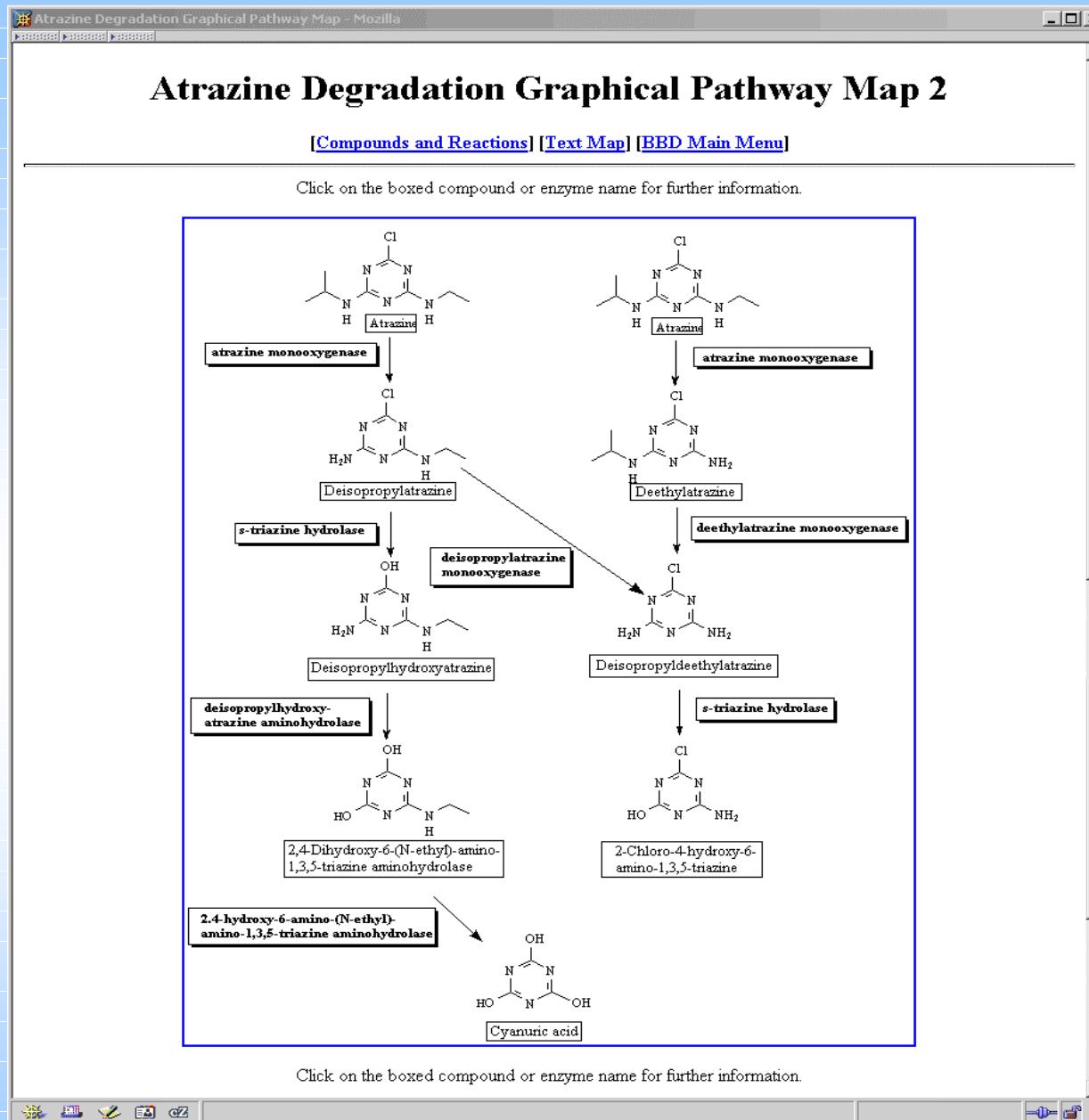
**atrazine.gif**  
1198 x 898 Pixel - 17k  
[nwql.usgs.gov/Public/pubs/QC\\_Fact/atrazine.gif](http://nwql.usgs.gov/Public/pubs/QC_Fact/atrazine.gif)

**marchatrazine.jpg**  
720 x 540 Pixel - 68k  
[la.water.usgs.gov/nawqa/liaison/marchatrazine.jpg](http://la.water.usgs.gov/nawqa/liaison/marchatrazine.jpg)

**augatrazine.jpg**  
720 x 540 Pixel - 68k  
[la.water.usgs.gov/nawqa/liaison/augatrazine.jpg](http://la.water.usgs.gov/nawqa/liaison/augatrazine.jpg)  
[ [Weitere Ergebnisse von la.water.usgs.gov](#) ]

**atra\_loads.gif**  
450 x 300 Pixel - 6k  
[www.epa.gov/glnpo/lmmb/results/loadatra.html](http://www.epa.gov/glnpo/lmmb/results/loadatra.html)

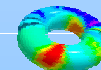




# Results

Considerable Data-gaps especially in:

- Environmental Fate and Pathway Parameters
- Chronic Ecotoxicity Parameters
- Even High Production Volume (HPV) chemicals are not covered in each environmental database
- Even the largest Internet databases provide insufficient data



# Publication

Kristina Voigt,

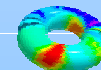
Databases on Environmental Information, in:

*Handbook of Chemoinformatics*, 4 volumes,

J. Gasteiger (editor),

Wiley-VCH, Weinheim, 2003,

pp. 722-742.



EPA/NHEERL: Distributed Structure-Searchable Toxicity (DSSTox) - Mozilla

**U.S. Environmental Protection Agency**

**Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network**

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) Search:  **GO**

[EPA Home](#) > [Research & Development](#) > [Health and Environmental Effects Research](#) > [DSSTox](#) > DSSTox Flowchart Graphic

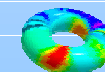
**DSSTox Graphic Flowchart**

Molecular structures are content-rich, unique chemical identifiers with widely understood meaning.

Publicly available toxicity databases exist in many locations and file formats, often are not downloadable in their entirety, and most do not contain chemical structures.

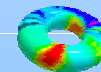
**Chemical Structures** + **Public Toxicity Data Files**

<http://www.epa.gov/nheerl/dsstox/>

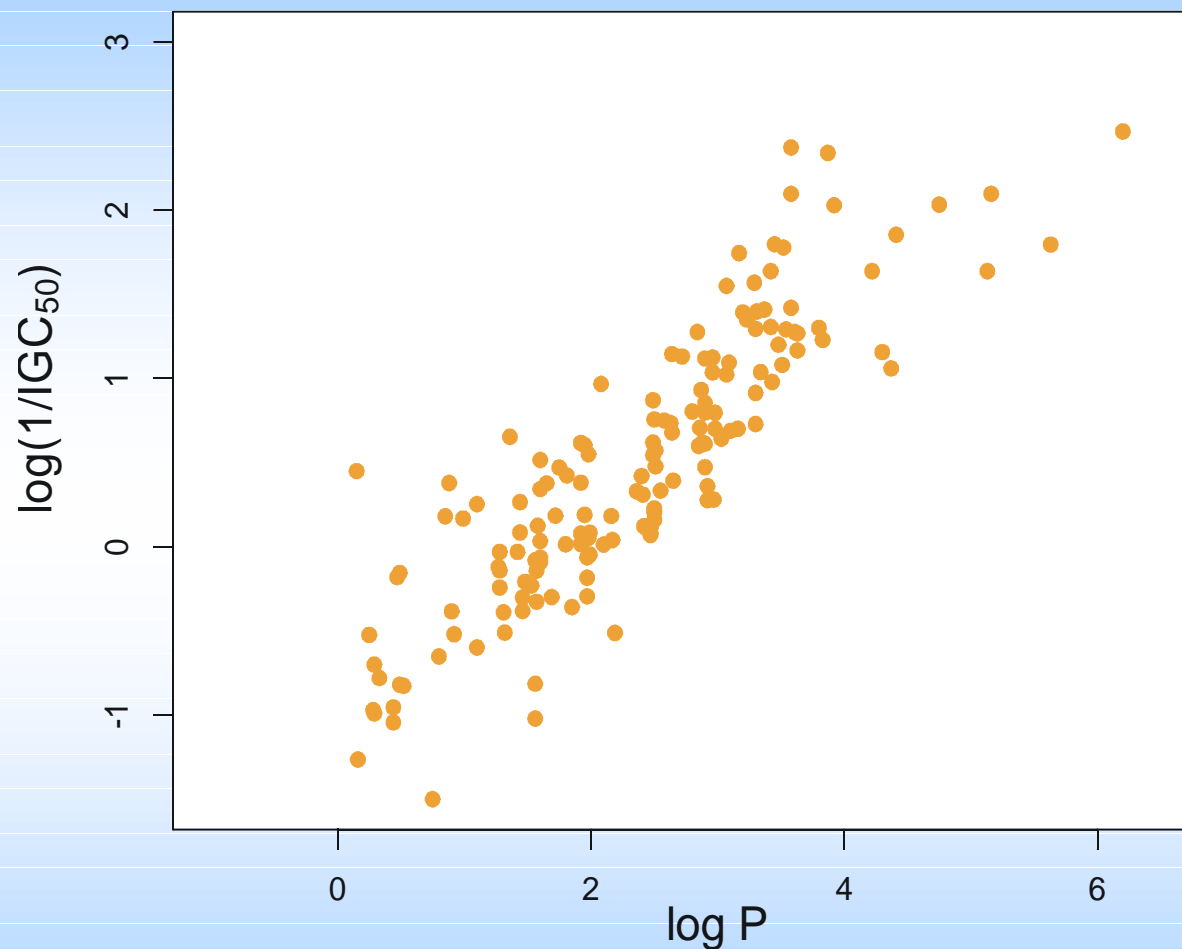


# DSSTox Databases

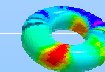
- CPDB:  
Carcinogenic Potency Database
- DBPCAN:  
Water Disinfection By-Products Database with  
Carcinogenicity Estimates
- EPAFHM:  
EPA Fathead Minnow Aquatic Toxicity Database
- NCTRER:  
FDA's National Center for Toxicological Research -  
Estrogen Receptor Binding Database



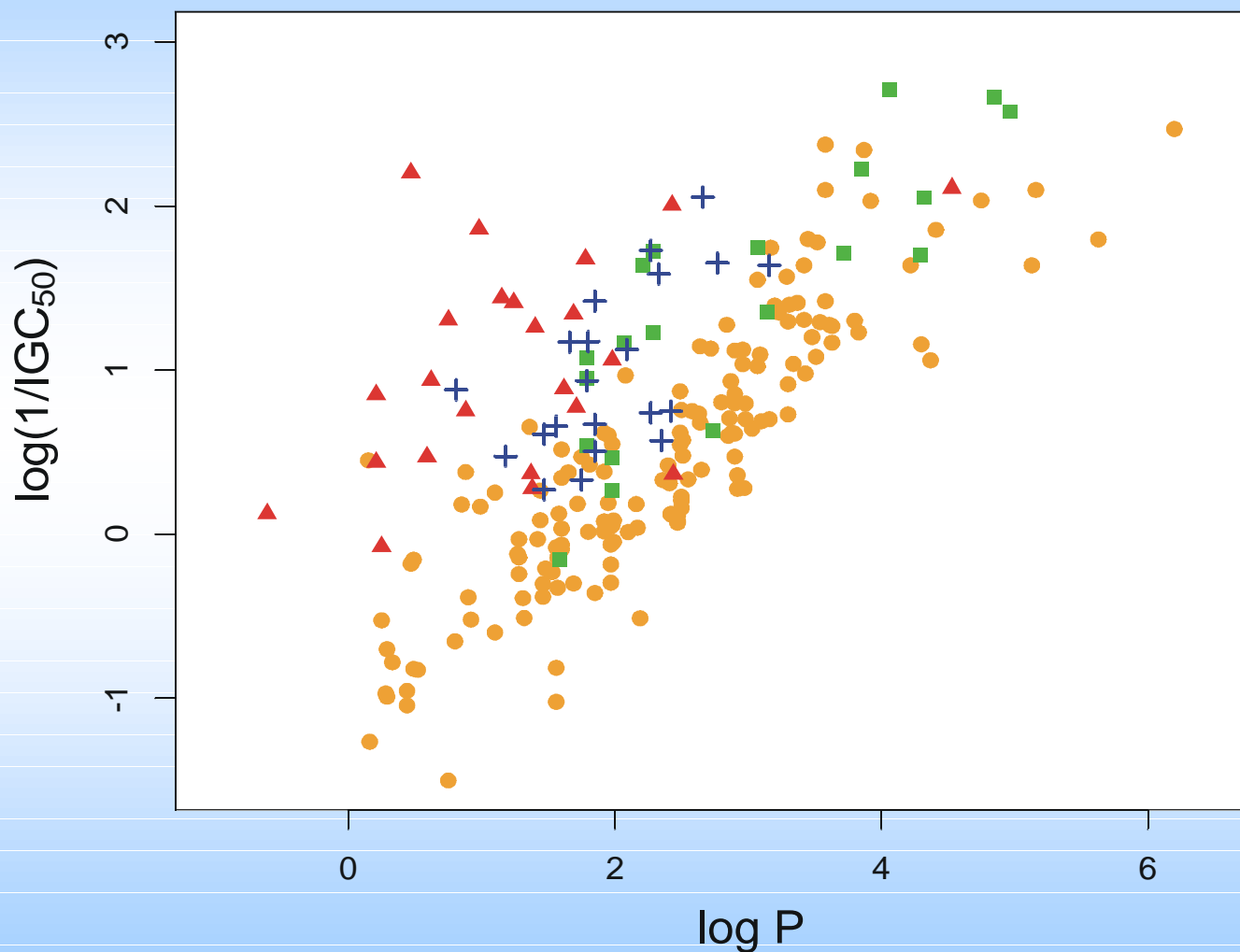
# Polar Narcotics



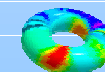
IGC<sub>50</sub> of *Tetrahymena pyriformis* with polar narcotic phenols versus log P



# The Larger Picture

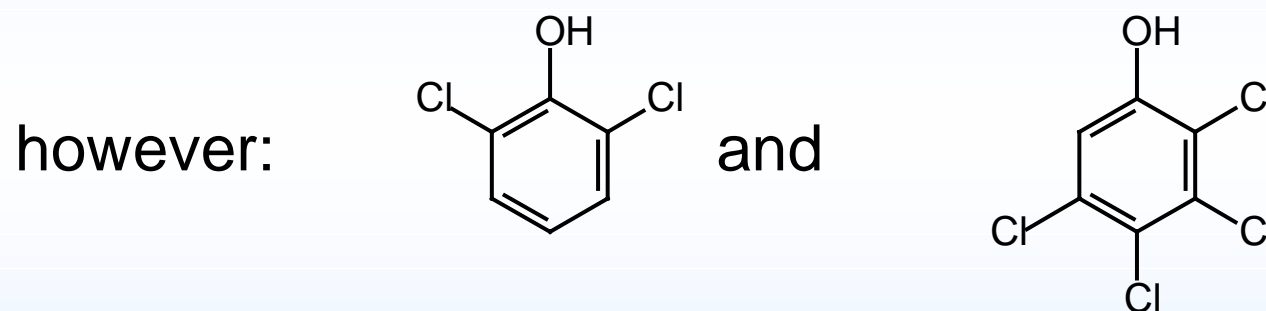


- polar narcotics
- uncouplers of oxidative phosphorylation
- ▲ precursors to soft electrophiles
- + soft electrophiles

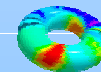


# Why Prediction of Modes of Action (MOA)?

most QSARs in toxicology focus on  
a certain **class of compounds**



have different **biological responses**,  
require different QSAR equations.



# Definition of Mode of Action (MOA)

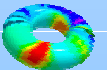
MOA:

common set of *physiological and behavioural signs* characterizing a type of adverse biological response

toxic mechanism:

*biochemical process* underlying a given MOA

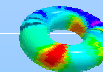
Rand G. (1995) Fundamentals of Aquatic Toxicology



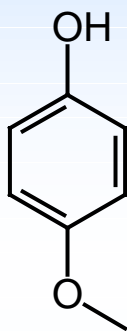
# Data Set: MOA of Phenols

1. polar narcotics	(157 cpds)
2. uncouplers of oxidative phosphorylation	(20 cpds)
3. precursors to soft electrophiles	(23 cpds)
4. soft electrophiles	(21 cpds)
	<hr/>
	221 cpds

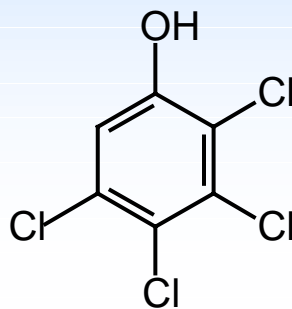
Aptula A. O., Netzeva T. I., Valkova I. V., Cronin M. T. D., Schultz T. W. D., Kühne R. and Schüürmann G.(2002): *Quant.Struct.-Act. Relat.*, 21, 12-22.



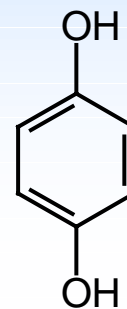
# Typical Structures



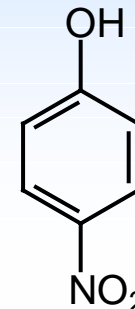
polar  
narcotics



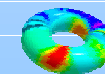
uncouplers



precursors to  
soft electrophiles

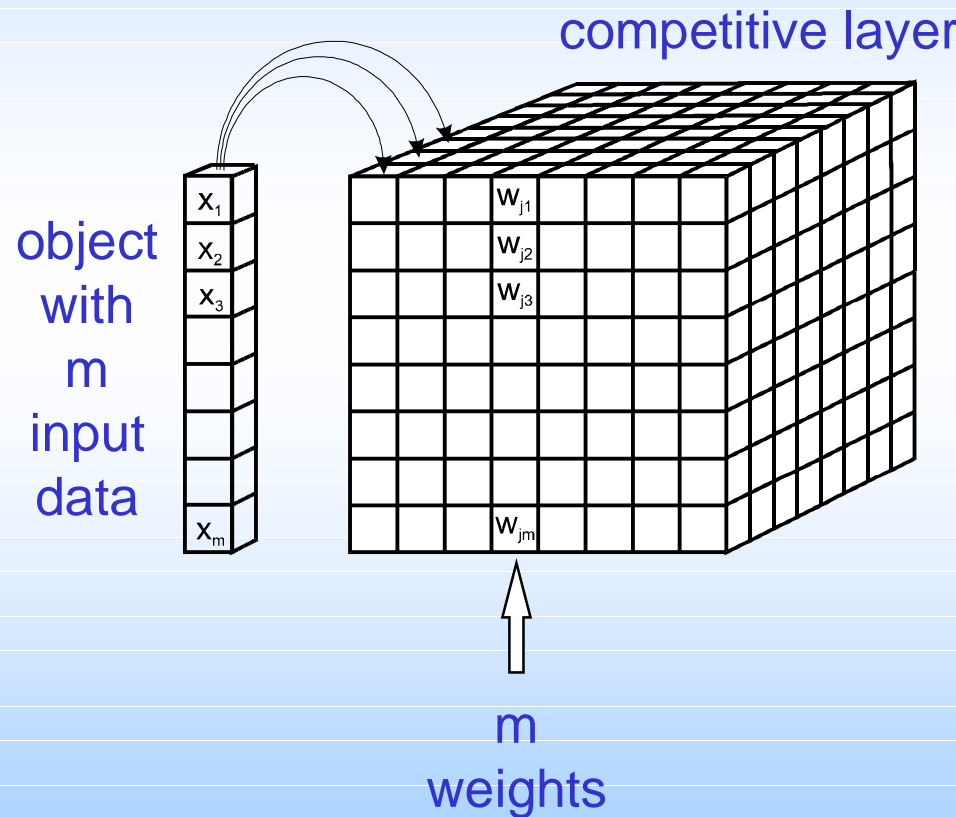


soft  
electrophiles



# Architecture of the Kohonen Network

## Self-Organizing Map (SOM)

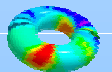


winning neuron

$$out_c \leftarrow \min \left\{ \sum_{i=1}^m (x_{si} - w_{ji})^2 \right\}$$

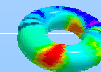
Euclidean distance

$$d_{sj} = \sqrt{\sum_{i=1}^m (w_{si} - w_{ji})^2}$$

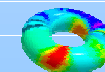
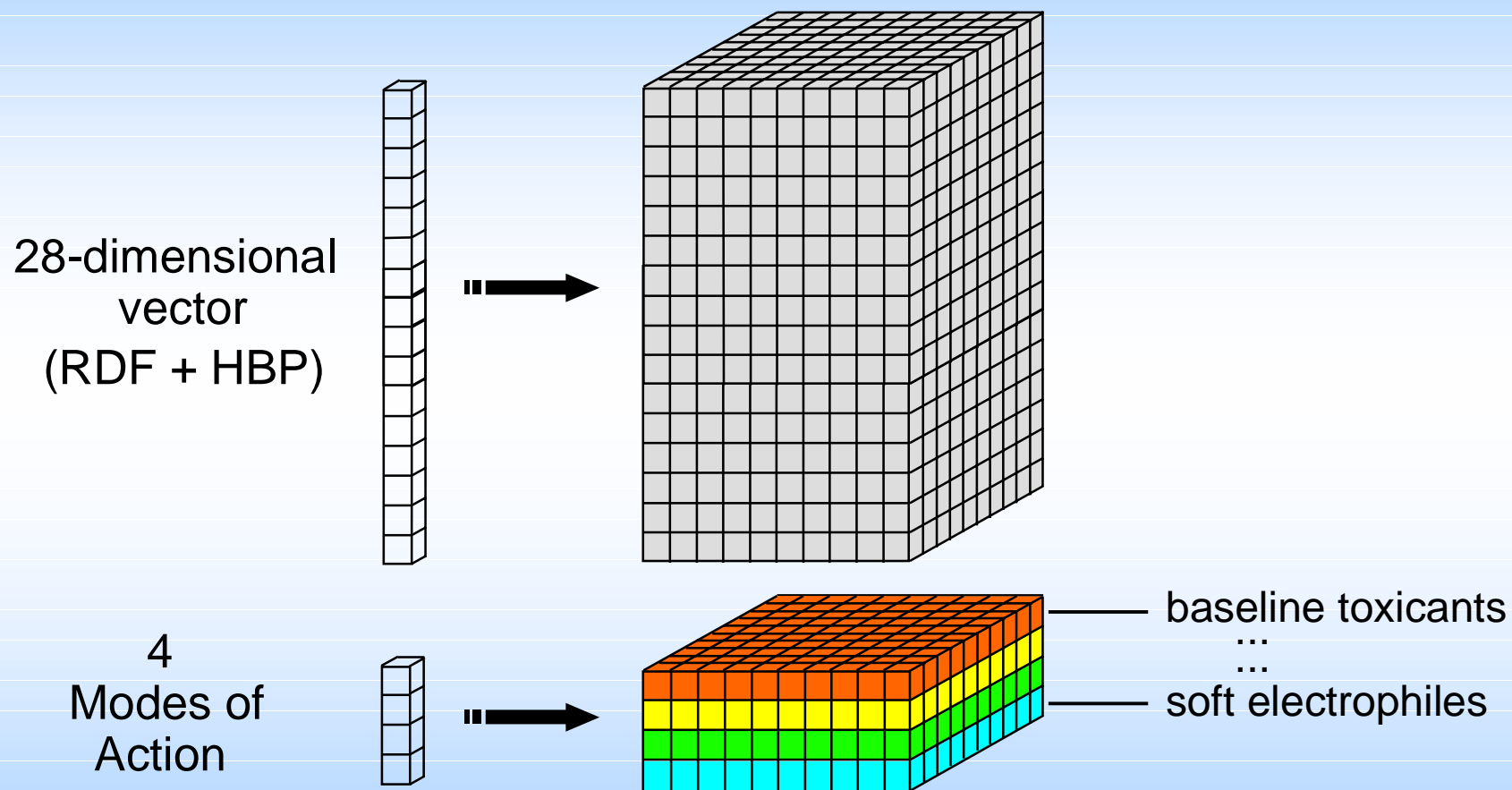


# Structure Representation

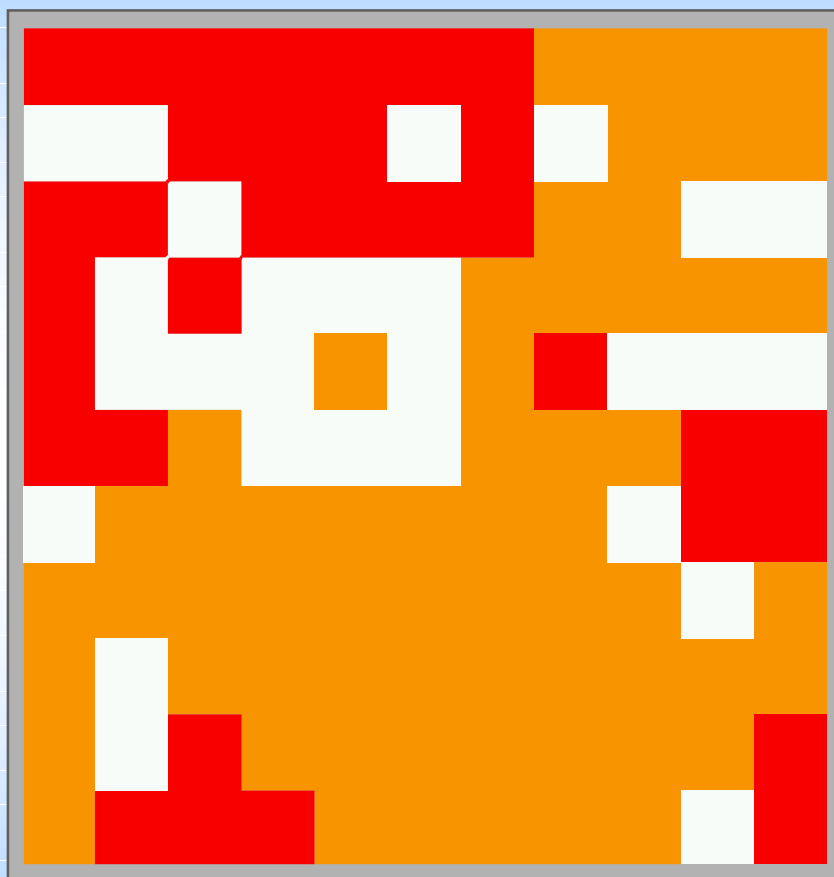
- 3D structure  
radial distribution function of  
sigma- and lone pair-electronegativity  
(16 values)
- molecular surfaces  
autocorrelation of hydrogen bonding potential  
(12 values)



# Classification According to Mode of Action (4 output layers)

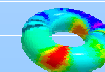


# Polar Narcotics

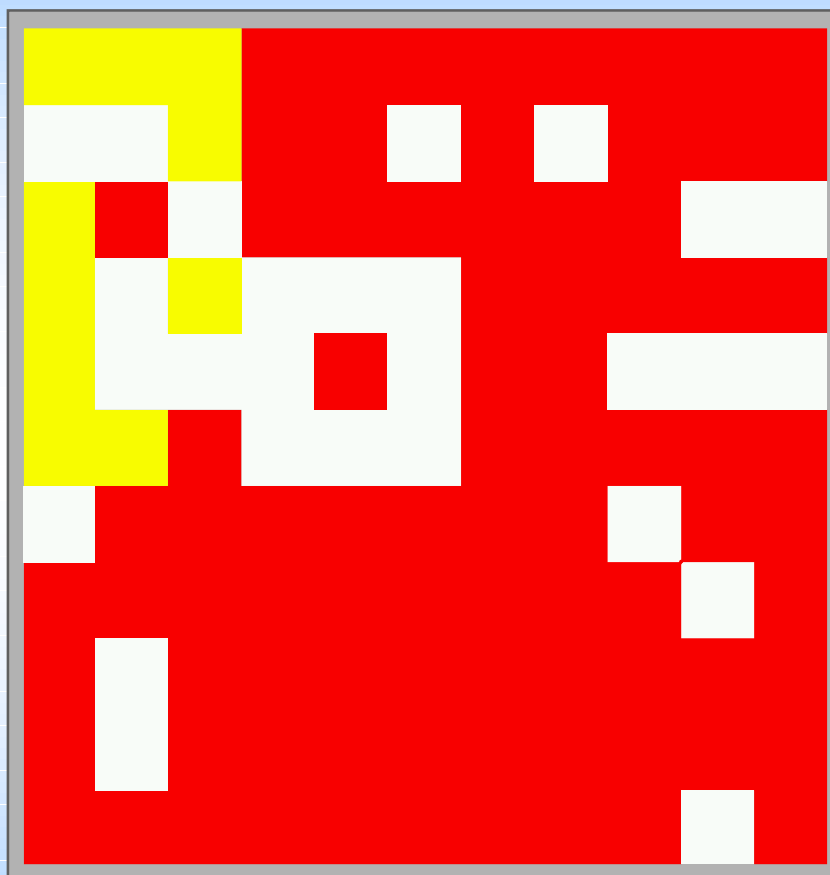


■ active  
■ non-active

221 phenols with 12 Surface HBP AC, 2x8 RDF ( $\chi_{LP}$ ,  $\chi_{\sigma}$ ), rectangular

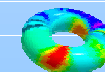


# Uncouplers of Oxidative Phosphorylation

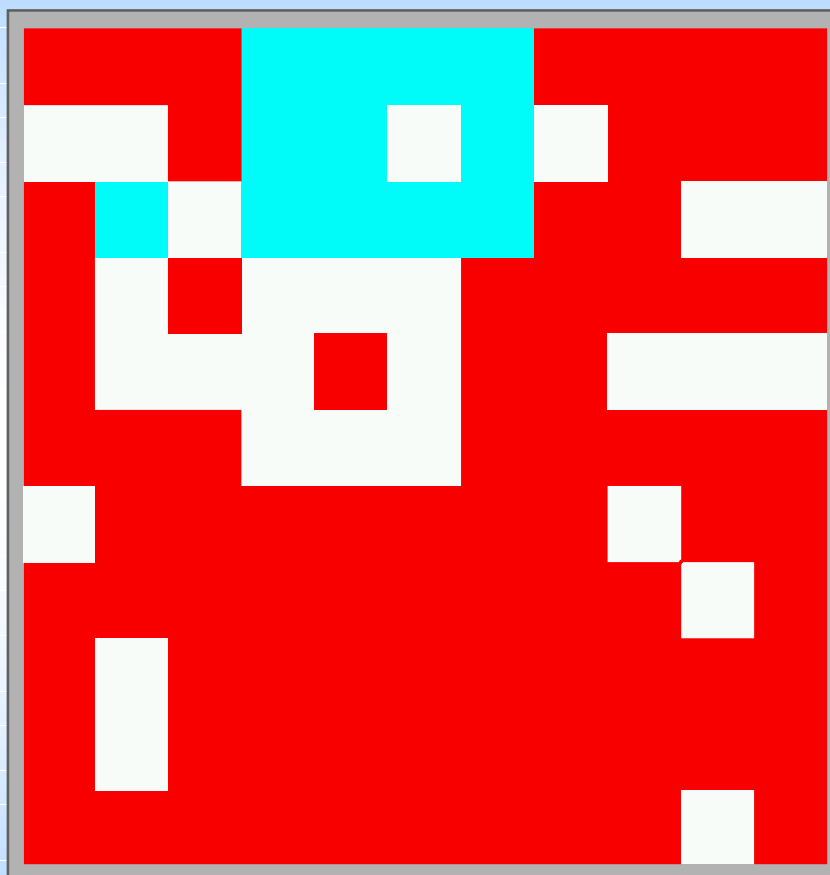




■ active  
■ non-active

221 phenols with 12 Surface HBP AC, 2x8 RDF ( $\chi_{LP}$ ,  $\chi_{\sigma}$ ), rectangular

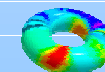


# Soft Electrophiles

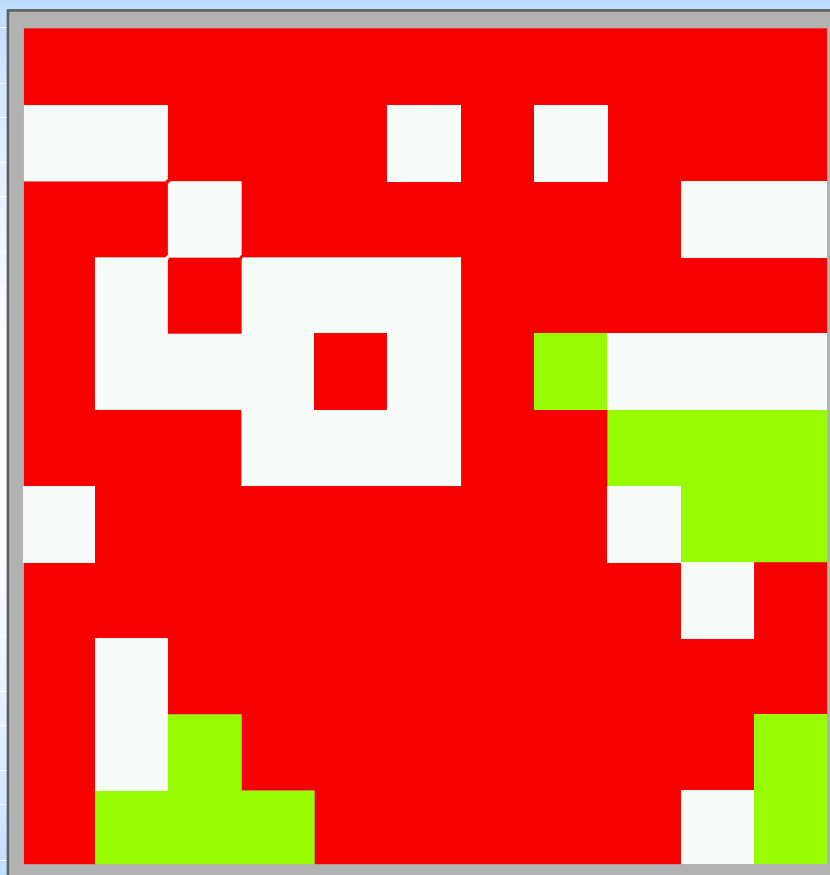




 active  
 non-active

221 phenols with 12 Surface HBP AC, 2x8 RDF ( $\chi_{LP}$ ,  $\chi_{\sigma}$ ), rectangular

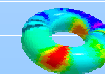


# Pro-electrophiles



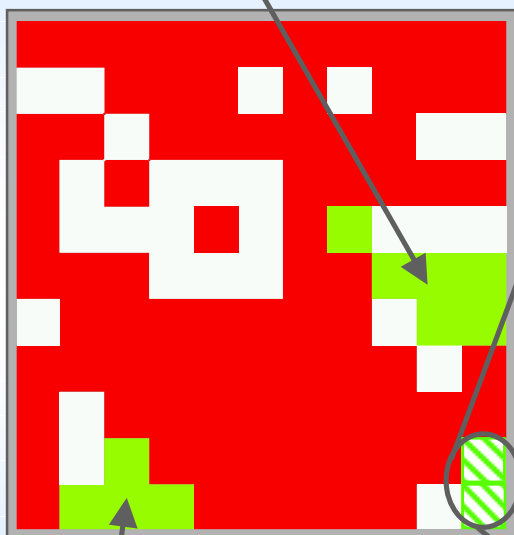
-  active
-  non-active

221 phenols with 12 Surface HBP AC, 2x8 RDF ( $\chi_{LP}$ ,  $\chi_{\sigma}$ ), rectangular

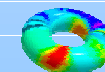
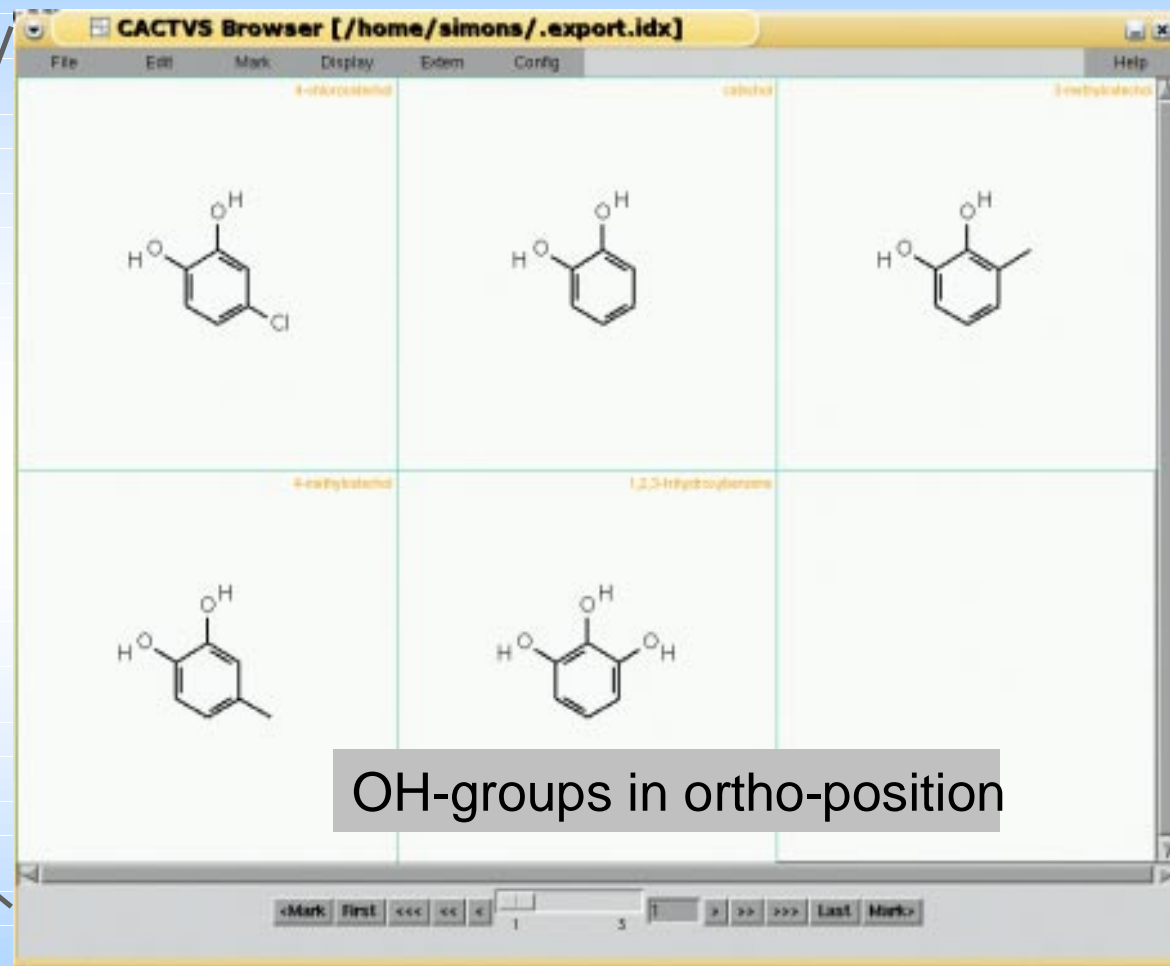


# Focus into the Neural Network

NH<sub>2</sub>-groups in ortho- and para-position



OH-groups in para position



# Final Model

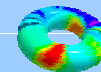
12 HBP surface autocorrelation, 2x8 RDF values ( $\chi_{LP}$ ,  $\chi_{\sigma}$ )

5-fold cross-validation:

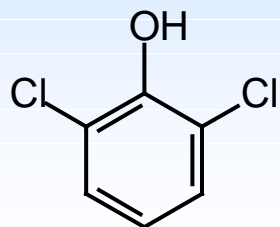
predicted \ actual	narcotic	uncouplers	pro-electroph.	soft electroph.
narcotic	155	0	4	0
uncouplers	0	19	0	1
pro-electroph.	1	0	19	0
soft electroph.	1	1	0	19
undecided	0	0	0	1

overall 95.9 % correct

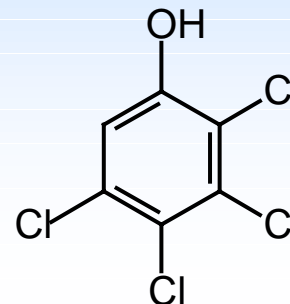
221 phenols with 12 Surface HBP AC, 2x8 RDF ( $\chi_{LP}$ ,  $\chi_{\sigma}$ ), rectangular



# Classification in 5-fold Cross-Validation

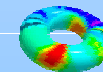


polar narcotic



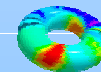
uncoupler of oxidative phosphorylation

classification successfull!



# Conclusions

- good predictive models for the MOA classification
- self-organizing maps can be used as transparent and robust neural network method
- the two step process, first MOA classification then building of a QSAR model is the right way to go



# Acknowledgements

Toxic Modes of Action

Simon Spycher, CCC, Erlangen

Environmental Databases

Dr. Kristina Voigt, GSF, Neuherberg

bmb+f

