DSSTox: A New On-Line Resource for Publishing Structure-Standardized Toxicity Databases

Ann Richard

Environmental Carcinogenesis Division
National Health & Environmental Effects Research Lab
US Environmental Protection Agency

In reference to technologies applied to analyzing DNA, RNA, proteins, metabolites in diverse biological systems:

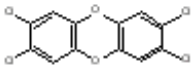
"Poor data management and a lack of standards has caused under-utilization of the data such that extracting fundamental knowledge and applications … requires extensive reformatting, repackaging, manual integration, etc."

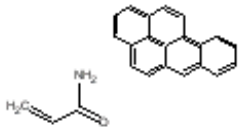
J.D. Eckart and B.W.S. Sobral (2003) A life scientist's gateway to distributed data management and computing: the PathPort/ToolBus framework, *Omics*, 7:79-88.

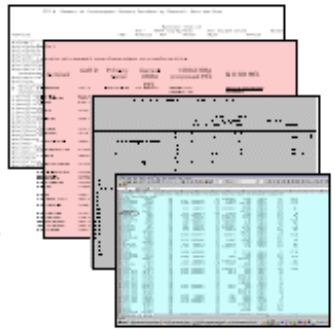
Distributed Structure-Searchable Toxicity (DSSTox) Network

Glossary

DSSTox







Welcome to the Distributed Structure-Searchable Toxicity (DSSTox) public database network.

DSSTox is an EPA-sponsored, community-wide project whose goals are to:

- Create and promote the use of SDF standard format, chemical structure-inclusive data files for storing public chemical toxicity data.

- Facilitate structure-searchability across toxicity databases and more complete access to data for use in toxicity prediction model development

- Involve the user community in the effort to migrate more public toxicity data into the SDF standard format for sharing

Goals of DSSTox Project:

- Provide open and full access to public toxicity data
- Include chemical structures and improve chemistry annotation of toxicity databases
- Promote standards for file formats, content and documentation
- Encourage broader SAR modeling participation, varied & flexible solutions
- Improve communication between tox, chemistry & modeling communities

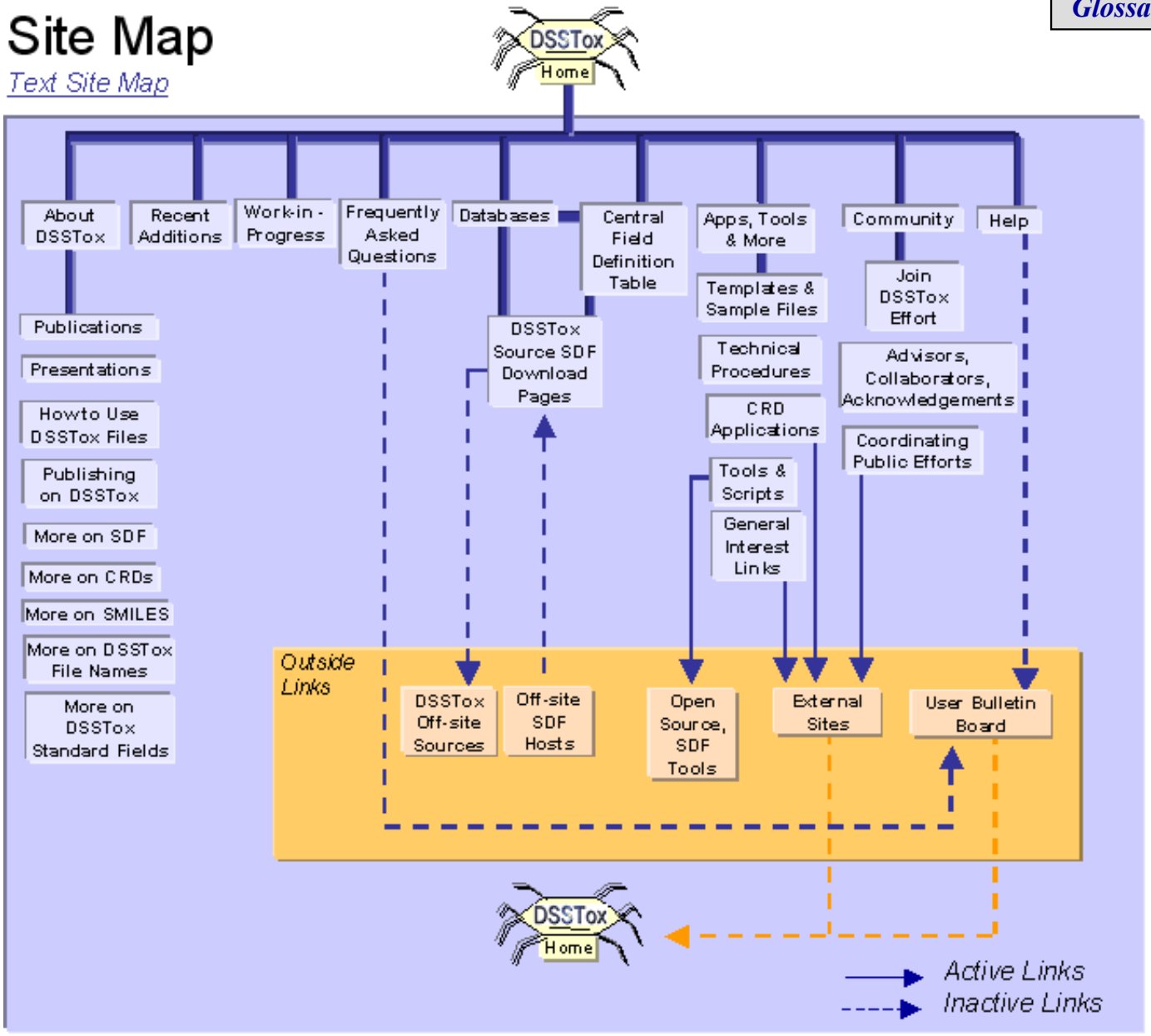
Distributed Structure-Searchable Toxicity (DSSTox) Network

Glossary

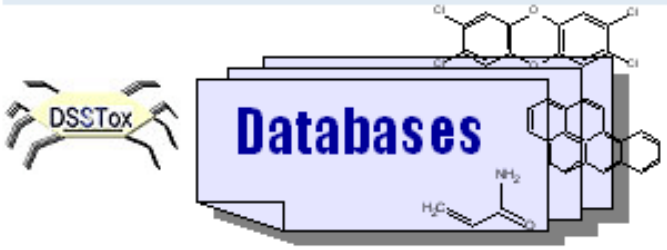
- Site Map
- About DSSTox
- Recent Additions
- Work-in-Progress
- Frequently Asked Questions
- Databases
- Central Field Definition Table
- Apps, Tools & More
- Community
- Help

Site Map

Text Site Map



Distributed Structure-Searchable Toxicity (DSSTox) Network



- Site Map
- About DSSTox
- Recent Additions
- Work-in-Progress
- Frequently Asked Questions
- Databases
- Central Field Definition Table
- Apps, Tools & More
- Community
- Help

- **CPDBRM, CPDBHA, CPDBDO, CPDBPR**: Carcinogenic Potency Database Summary Tables for Rat&Mouse, Hamster, Dog, and Non-human Primates
  Tumor target site incidence and TD50 potencies for 1354 chemical substances tested in rats and mouse, 80 chemical substances tested in hamsters, 5 chemicals tested in dogs, and 27 chemical substances tested in non-human primates; data reviewed and compiled from literature and NTP studies.
  (SDF last updated 15Oct03)

- **DBPCAN**: Water Disinfection By-Products Database with Carcinogenicity Estimates
  Carcinogenicity estimates (high, moderate, low concern) by EPA experts using a mechanism-based analog SAR approach on a set of 209 water disinfection by-products, mostly small halogenated organics.
  (SDF last updated 12Sep03)

- **EPAFHM**: EPA Fathead Minnow Aquatic Toxicity Database
  Acute toxicities of 617 chemicals tested in common assay, with mode-of-action assessments and confirmatory measures.
  (SDF last updated 15Oct03)

- **NCTRER**: FDA's National Center for Toxicological Research - Estrogen Receptor Binding Database
  Estrogen receptor relative binding affinities tested in a common in vitro assay for 232 chemicals, listed with chemical class-based structure activity features.
  (SDF last updated 7Nov03)

DSSTox Toxicity Database Standards:

- SDF data file format
- File naming convention
- Chemical structure information fields
- Documentation requirements

Sample "SDF" FILE:

"mol" file + text/data fields

```
csChmFindW05030111462D

 14 16  0  0  0  0  0  0  0  0999 V2000
   0.1283    2.1977    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   0.0000    0.7780    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   1.0347    0.0000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   2.3261    0.5213    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   2.4544    1.9411    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   1.4197    2.7191    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   3.6254    0.0000    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   4.5318    1.0347    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   3.8821    2.1977    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   5.9516    1.0347    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   6.7295    2.1977    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   5.9516    3.4891    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   4.5318    3.4891    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   8.0209    2.1977    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  1  6  2  0  0  0  0
  2  3  2  0  0  0  0
  3  4  1  0  0  0  0
  4  5  1  0  0  0  0
  4  7  2  0  0  0  0
  5  6  1  0  0  0  0
  5  9  1  0  0  0  0
  7  8  1  0  0  0  0
  8  9  2  0  0  0  0
  8 10  1  0  0  0  0
  9 13  1  0  0  0  0
 10 11  2  0  0  0  0
 11 12  1  0  0  0  0
 11 14  1  0  0  0  0
 12 13  2  0  0  0  0
M  END
>  <Last Updated> (1)
5/3/01

>  <Source> (1)
http://potency.berkeley.edu/cpdb.html

>  <Chemical> (1)
A-alpha-C

>  <CAS> (1)
26148-68-5

>  <Tested Form> (1)
neutral
```

- Industry standard export, import file
- Simple ASCII text file
- 2D structure, text, and data fields
- Supports "unlimited" # of records, fields
- SDF tools, scripts available

DSSTox File Naming Standard:

EPAFHM_v1a_617_15Oct03.sdf (.xls)

6 letter NAMEID

Number of records

Date of last modification

File type

version 1: major modifications or data additions

revision a: error corrections, minor additions

DSSTox Standard Chemical Fields:

- Structure — *2D chemical structure*
- StructureShown — *Description of displayed 2D structure*
  - tested form, simplified to parent, predicted form, monomer, active ingredient of formulation
- Formula — *Empirical molecular formula*
- MolWeight — *Molecular weight in atomic units*
- CAS — *Chem Abstracts Service No. for StructureShown*
- SMILES — *Linear text notation for 2D StructureShown*
- DSSTox_ID — *Counter allows unique identification of record*
- DSSTox_FileName — *Name of file included in each record*
- ChemName — *Chemical name from original data base*
- SubstanceType — *Broad substance classification*
  - defined organic, inorganic, organometallic, polymer, mixture or unknown
- TestedForm — *Tested form of chemical*
  - parent, salt, complex, unknown or multiple forms
- AddToParent — *Salt counterions or complexed moieties*
- CAS_TestedForm — *CAS No. for tested form of chemical*
- SMILES_TestedForm — *SMILES code for the tested form of the chemical*
- ChemNote — *Additional qualifier info for chemical fields*
  - defined mixture characteristics, uncertainty in structure or CAS, stereochem, replicate, etc.
- ChemCount — *Counter for structure or CAS duplications in database*

DSSTox Standard Chemical Field Definition Table

(last updated 12 November 03)

The following table is intended to serve as a detailed reference document for the definition and use of the DSSTox Standard Chemical Fields in DSSTox SDF data files. Additional notes and documents pertaining to the use of these fields follow the abbreviated field definition table located on the More on DSSTox Standard Chemical Fields page of the main DSSTox website. Abbreviated versions of the following field definitions are also included in each DSSTox SDF Field Definition File. In addition, a complete alphabetical listing of all fields contained within current DSSTox SDF files, including both Standard Chemical Fields and Source-Specific Fields, are offered in the DSSTox Central Field Definition File.

Field Name	Allowable Values	Description	Comments
Structure	2D (or 3D) "mol" file coordinates for defined molecular structure	Main DSSTox file includes 2D graphical representation of molecular structure (i.e., x,y coordinates only). Form of structure is identified in the **StructureShown** field and always corresponds to the fields: **CAS**, **SMILES**, **MolWeight**, **Formula**. Structure entry is *blank* when **SubstanceType** entry is "mixture or unknown".	Chemical structure shown may be a single molecular entity or, if the "tested form" is displayed, may include salt counter ions, complexed waters or other complexed molecular species. Structures are obtained from a variety of public databases and sources and are verified with CAS numbers whenever possible. In some cases, structures are generated from ChemName and/or SMILES information provided in original Source database. Details of file construction are provided in the LogFile for each DSSTox database, available for viewing and download from the DSSTox Source SDF Download Page. Aromatic rings are drawn with explicit alternating double bonds in all cases. Where tautomeric forms are known to exist, these are drawn in a consistent tautomeric representation throughout the database. Nitro groups are represented in the charge separated form. **StructureShown**= "simplified to parent" structures are represented in their neutral or protonated forms, i.e. without counterions or complexed chemical entities. An exception is quaternary ammonium ions, which are represented in the positive charged state even when "simplified to parent". DSSTox SDF files, as a rule, do not have a **Structure** entry for any substances classified as "mixture or unknown", even if the substance is a defined mixture of two or more defined organic chemicals. To do so would generate meaningless MolWeight and Formula field entries and would potentially lead to misuse of the approximate **Structure** in applications where a single chemical entity is assumed (such as in a chemical property calculation module or structure-activity prediction). If available, additional information on the chemical components or purity of defined mixtures is provided in the **ChemNote** field. SDF format supports display of triangular bonds and cis/trans orientations of double bonds in most SDF viewing applications. SDF can also store and display 3D structures. For selected databases, 3D structures are provided in a supplementary file for specialized use in 3D modeling applications. This file will generally be a mirror of the corresponding DOP (Defined Organic Parent) file, containing simplified structures of only the defined organics in the database (i.e., excluding inorganics, organometallics, and mixtures or unknowns). More on SDF
Structure Shown *(no spaces)*	tested form/ simplified to parent/ predicted form/ general form/ active ingredient	Identifies form of graphical 2D structure displayed in the **Structure** field. "tested form" - structure displayed is the actual form of the structure experimentally tested in the toxicity assay; "simplified to parent" - only in specialized DOP (defined organic parent) SDF files; "predicted form" - a theoretical prediction of activity, rather than actual test data, is provided in the chemical record; "general form" - chemical record contains toxicity data fields summarized from multiple experiments, where either multiple tested forms of the chemical were used, or the tested form is not specified; "active ingredient" - the actual tested form of the chemical substance was a mixture or formulation for the majority of chemicals in the database, and where only the active ingredient is represented in the structure field; if a DOP file is created, entry can be followed by "simplified to parent" for active ingredient salts or complexes.	In files containing test data of dis… chemical entities, field entry is "tested form" for all records in Main SDF. For DOP SDF file, entry is "tested form" … organic" a "parent" structure". Entry is "simplified to parent" only for "salt" or "complex", where displayed … rm. In the latter case, **CAS** and **SMILES** field entries will differ … ts. The field entry "predicted form" is … redictions, i.e. that contain no actual test data. Generally, such database… no salts or complexes). The field entry "general form" is u… chemicals tested in many experiments and/or endpoints. In these cases, the chemical could possibly be tested in multiple forms (i.e., parent, salt, or complex). This field entry will always correspond to a **TestedForm** entry of "unknown or multiple forms". For databases using the field entry "active ingredient", the original CAS provided by the Source may correspond to the original formulation or to the active ingredient. If the CAS number for the original formulation is provided by the Source for a significant fraction of the database, the field **CAS TestedForm** will be included in addition to **CAS**, since the latter will correspond to the "active ingredient" in the **Structure** field.
Formula	*Text*	Empirical formula of displayed **Structure**.	Empirical formula field entry is automatically generated within the CambridgeSoft ChemFinder application based on the **Structure** field entry.

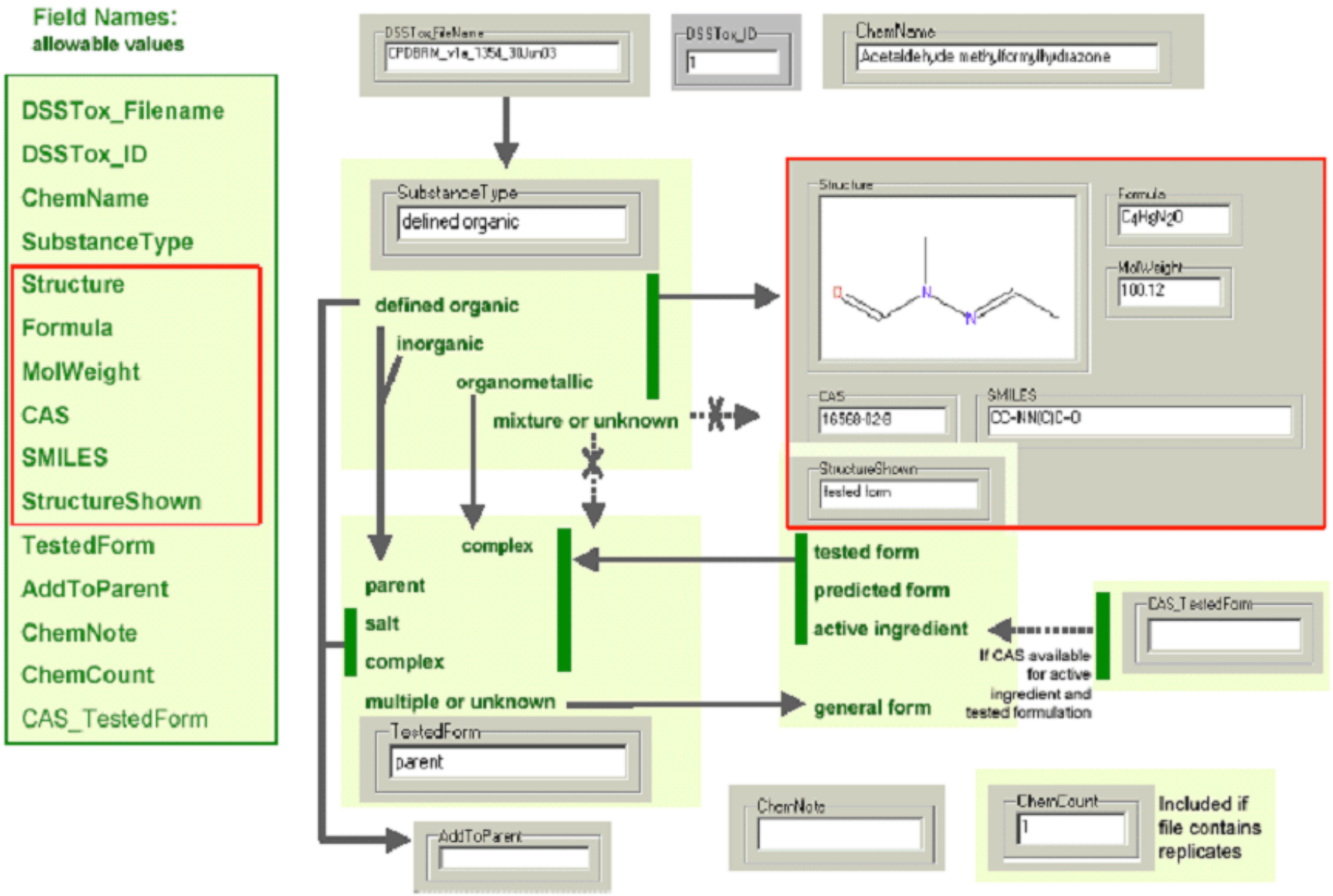
All Allowable Values listed and defined

Detailed description of standard fields

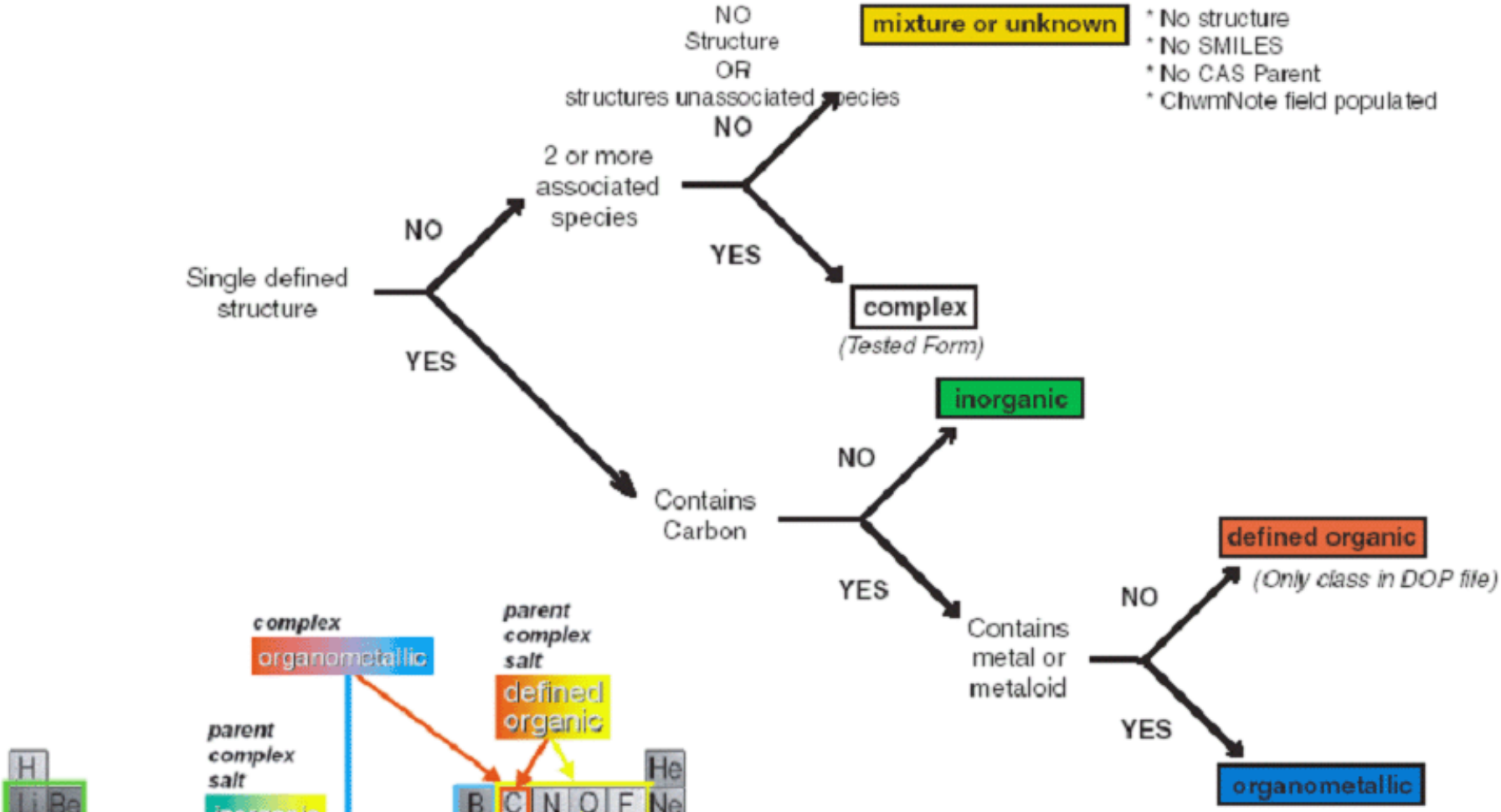
Links to information pages

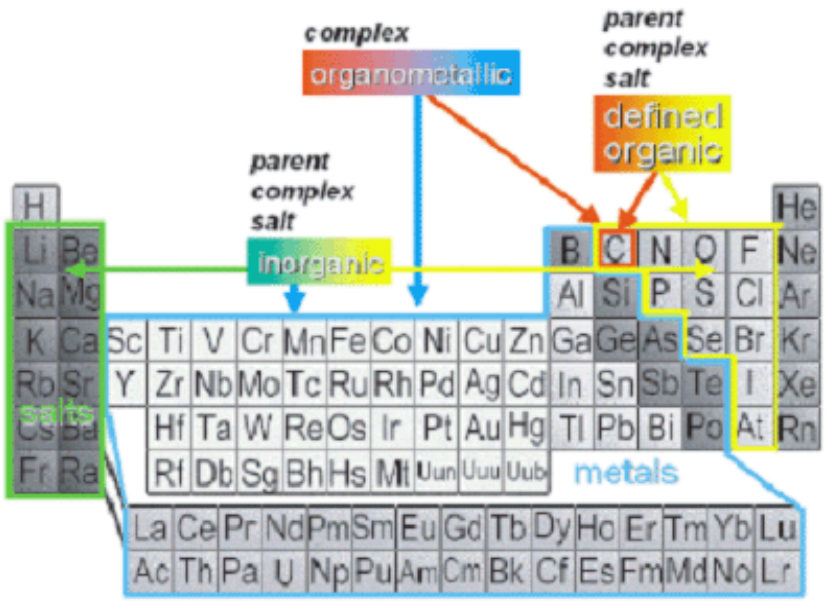
DSSTox Standard Chemical Fields: Decision Path (Main Files)

(last updated 10 Nov 03)



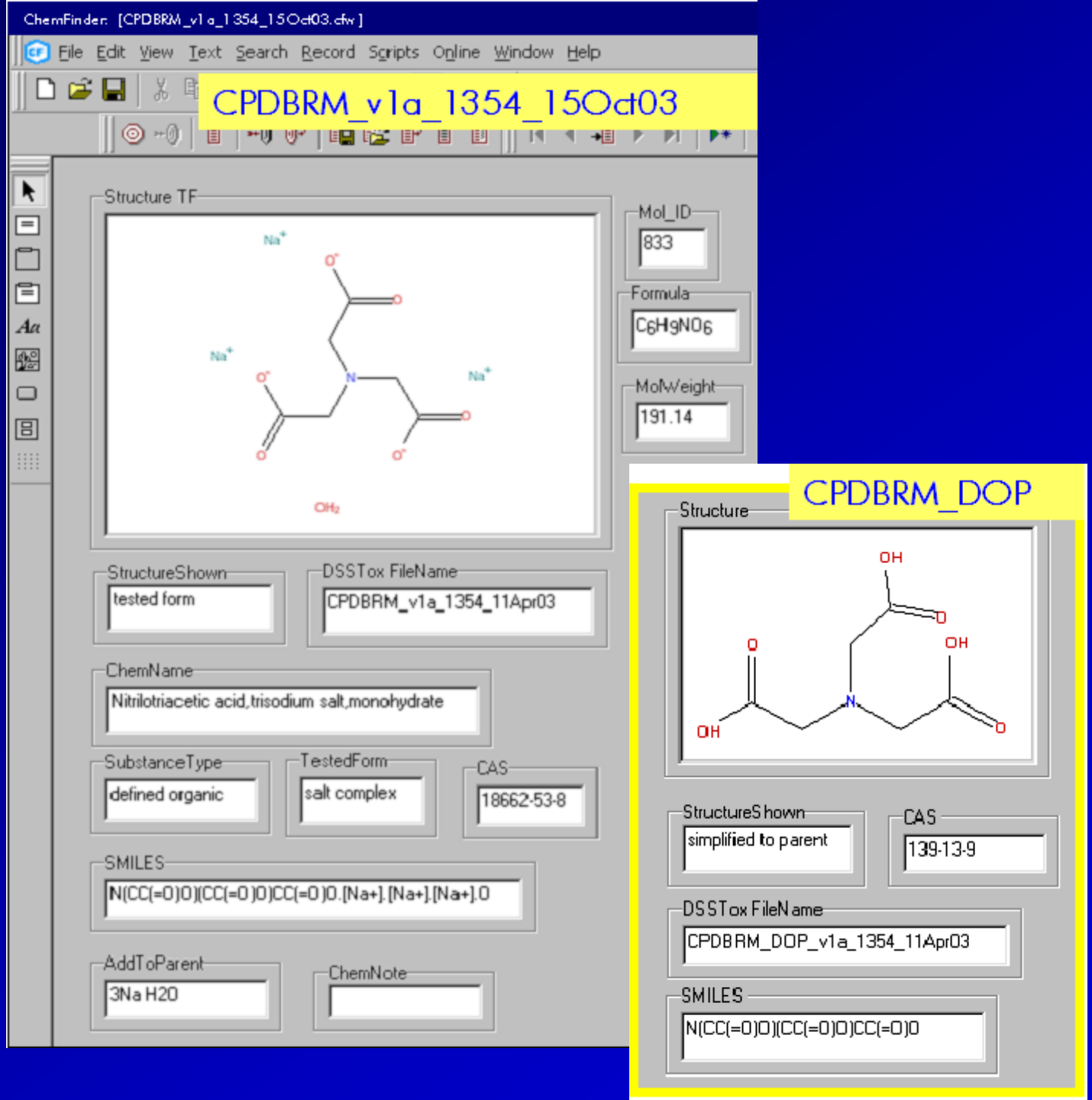
Substance Type





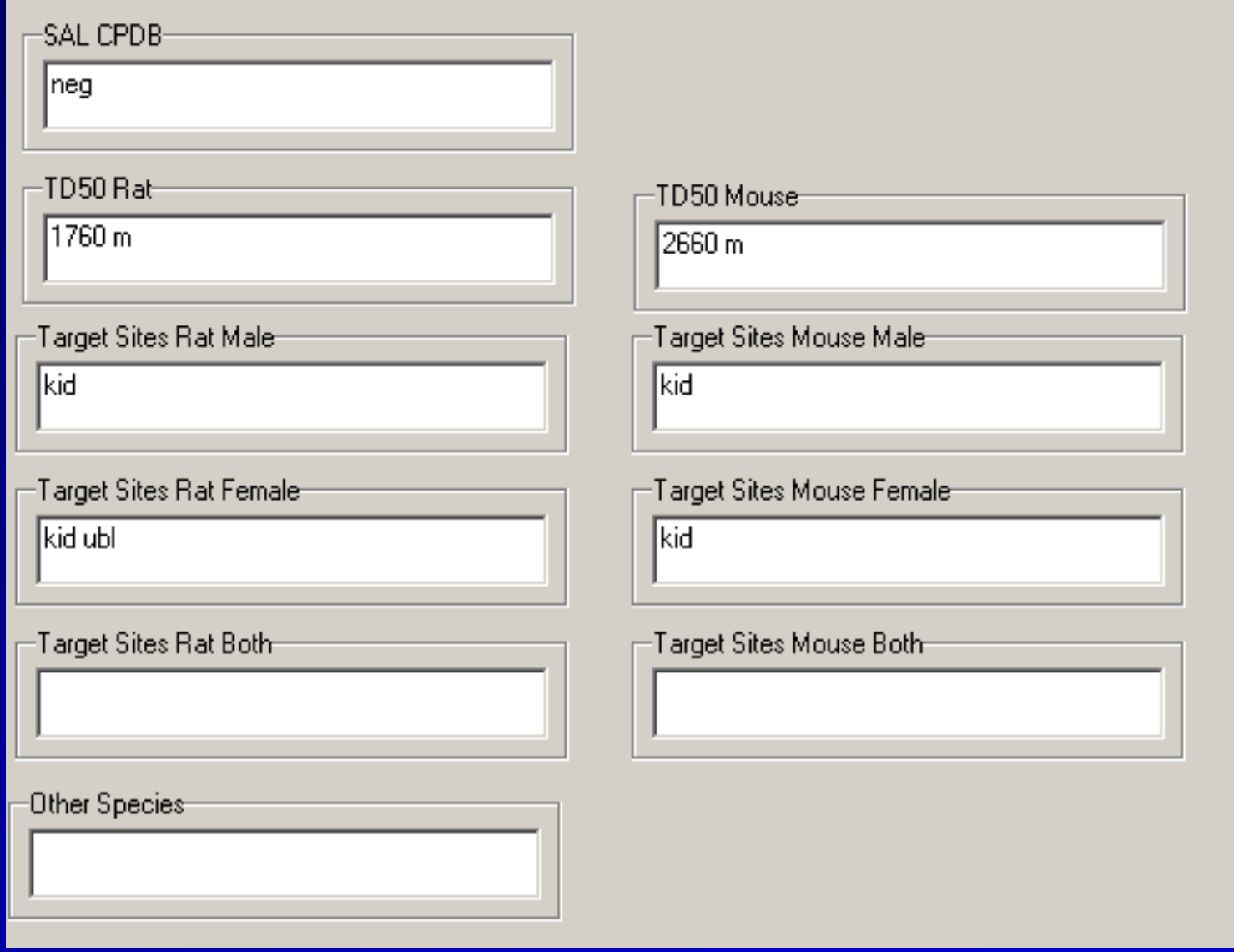
DSSTox Standard Chemical Fields:

CambridgeSoft ChemFinder Application view after SDF import



DSSTox Source-Specific Fields:

CambridgeSoft ChemFinder Application view after SDF import



CPDBRM_v1a_1354_15Oct03

CPDBRM_DOP_v1a_1354_15Oct03

DSSTox Database Documentation:

- Source Download Page
- Log File
- Field Definition File

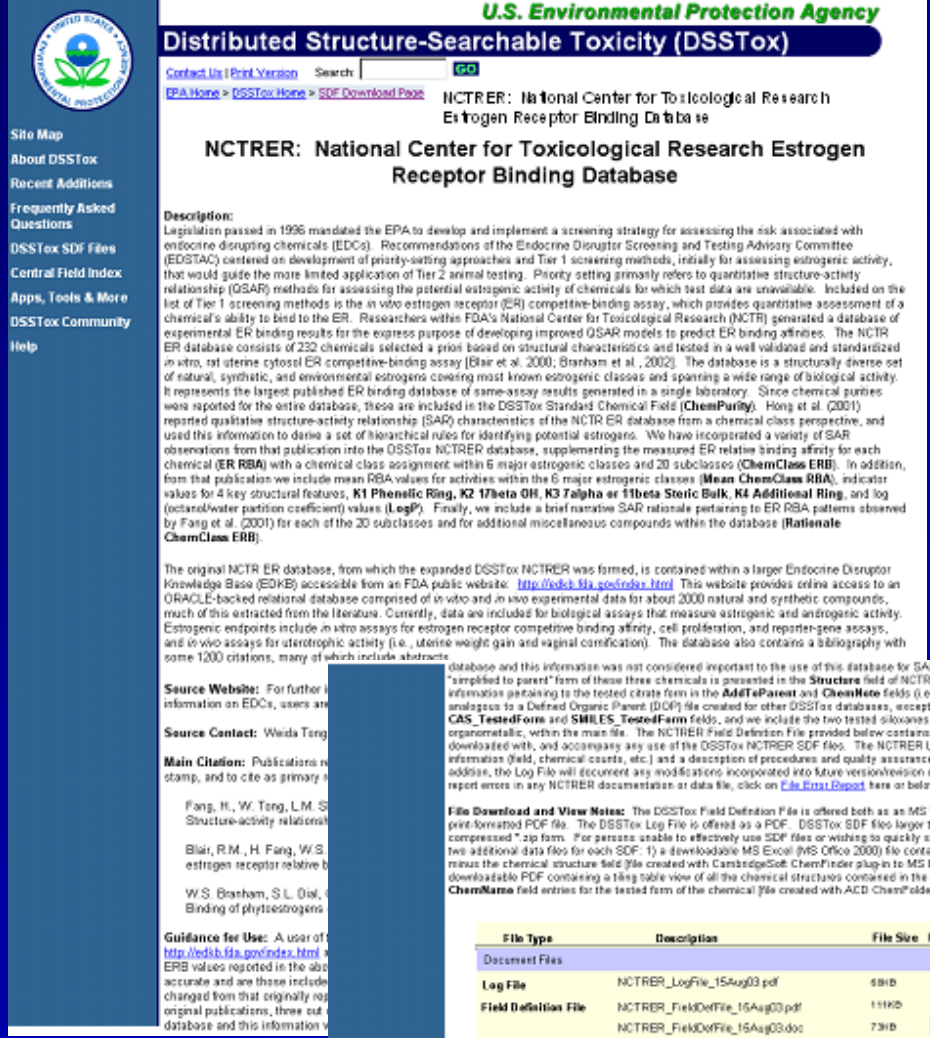
DSSTox Source SDF Download Page:

- NAMEID: Database Title

- Description
- Source Website
- Source Contact
- Main Citation
- Guidance for Use
- File Download & View Notes

- Documentation files
- SDF Structure-Data file
- Excel Data Table file
- PDF Structures file
- File Error Report

- Acknowledgements
- DSSTox Citation



U.S. Environmental Protection Agency

Distributed Structure-Searchable Toxicity (DSSTox)

Contact Us | Print Version Search: GO

EPA Home > DSSTox Home > SDF Download Page

NCTRER: National Center for Toxicological Research Estrogen Receptor Binding Database

Site Map
About DSSTox
Recent Additions
Frequently Asked Questions
DSSTox SDF Files
Central Field Index
Apps, Tools & More
DSSTox Community
Help

NCTRER: National Center for Toxicological Research Estrogen Receptor Binding Database

Description:
Legislation passed in 1996 mandated the EPA to develop and implement a screening strategy for assessing the risk associated with endocrine disrupting chemicals (EDCs). Recommendations of the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) centered on development of priority-setting approaches and Tier 1 screening methods, initially for assessing estrogenic activity, that would guide the more limited application of Tier 2 animal testing. Priority setting primarily refers to quantitative structure-activity relationship (QSAR) methods for assessing the potential estrogenic activity of chemicals for which test data are unavailable. Included on the list of Tier 1 screening methods is the *in vitro* estrogen receptor (ER) competitive-binding assay, which provides quantitative assessment of a chemical's ability to bind to the ER. Researchers within FDA's National Center for Toxicological Research (NCTR) generated a database of experimental ER binding results for the express purpose of developing improved QSAR models to predict ER binding affinities. The NCTR ER database consists of 232 chemicals selected a priori based on structural characteristics and tested in a well validated and standardized *in vitro*, rat uterine cytosol ER competitive-binding assay (Blair et al. 2000; Branham et al., 2002). The database is a structurally diverse set of natural, synthetic, and environmental estrogens covering most known estrogenic classes and spanning a wide range of biological activity. It represents the largest published ER binding database of same-assay results generated in a single laboratory. Since chemical purities were reported for the entire database, these are included in the DSSTox Standard Chemical Field (**ChemPurity**). Hong et al. (2001) reported qualitative structure-activity relationship (SAR) characteristics of the NCTR ER database from a chemical class perspective, and used this information to derive a set of hierarchical rules for identifying potential estrogens. We have incorporated a variety of SAR observations from that publication into the DSSTox NCTRER database, supplementing the measured ER relative binding affinity for each chemical (**ER RBA**) with a chemical class assignment within 6 major estrogenic classes and 20 subclasses (**ChemClass ERB**). In addition, from that publication we include mean RBA values for activities within the 6 major estrogenic classes (**Mean ChemClass RBA**), indicator values for 4 key structural features, **K1 Phenolic Ring, K2 17beta OH, K3 7alpha or 11beta Steric Bulk, K4 Additional Ring**, and log (octanol/water partition coefficient) values (**LogP**). Finally, we include a brief narrative SAR rationale pertaining to ER RBA patterns observed by Fang et al. (2001) for each of the 20 subclasses and for additional miscellaneous compounds within the database (**Rationale ChemClass ERB**).

The original NCTR ER database, from which the expanded DSSTox NCTRER was formed, is contained within a larger Endocrine Disruptor Knowledge Base (EDKB) accessible from an FDA public website: http://edkb.fda.gov/index.html This website provides online access to an ORACLE-backed relational database comprised of *in vitro* and *in vivo* experimental data for about 2000 natural and synthetic compounds, much of this extracted from the literature. Currently, data are included for biological assays that measure estrogenic and androgenic activity. Estrogenic endpoints include *in vitro* assays for estrogen receptor competitive binding affinity, cell proliferation, and reporter-gene assays, and *in vivo* assays for uterotrophic activity (i.e., uterine weight gain and vaginal cornification). The database also contains a bibliography with some 1200 citations, many of which include abstracts.

Source Website: For further i

Source Contact: Weida Tong

Main Citation: Publications r
stamp, and to cite as primary r

Fang, H., W. Tong, L.M. S
Structure-activity relationsh

Blair, R.M., H. Fang, W.S.
estrogen receptor relative b

W.S. Branham, S.L. Dial,
Binding of phytoestrogens

Guidance for Use: A user of
http://edkb.fda.gov/index.html
ERB values reported in the ab
accurate and are those include
changed from that originally re
original publications, three out
database and this information

database and this information was not considered important to the use of this database for SAR investigations, the neutralized parent, or "simplified to parent" form of these three chemicals is presented in the **Structure** field of NCTRER. For these three cases, we include basic information pertaining to the tested citrate form in the **AddToParent** and **ChemNote** fields (i.e., CAS of the citrate form). This would be analogous to a Defined Organic Parent (DOP) file created for other DSSTox databases, except that we do not include here the **CAS_TestedForm** and **SMILES_TestedForm** fields, and we include the two tested siloxanes, operationally classified within DSSTox as organometallic, within the main file. The NCTRER Field Definition File provided below contains essential documentation and should be downloaded with, and accompany any use of the DSSTox NCTRER SDF files. The NCTRER Log File provides database summary information (field, chemical counts, etc.) and a description of procedures and quality assurance checks used in SDF file creation. In addition, the Log File will document any modifications incorporated into future version/revision updates of the DSSTox NCTRER SDF file. To report errors in any NCTRER documentation or data file, click on File Error Report here or below.

File Download and View Notes: The DSSTox Field Definition File is offered both as an MS Word (MS Office 2000) document and as a print-formatted PDF file. The DSSTox Log File is offered as a PDF. DSSTox SDF files larger than 1MB are offered for download in compressed ".zip form. For persons unable to effectively use SDF files or wishing to quickly survey the content of the SDF files, we provide two additional data files for each SDF: 1) a downloadable MS Excel (MS Office 2000) file containing the full SDF data contents in table form, minus the chemical structure field [file created with CambridgeSoft ChemFinder plug-in to MS Excel 2000]; and 2) a viewable and downloadable PDF containing a tiling table view of all the chemical structures contained in the database, annotated with **CAS** and truncated **ChemName** field entries for the tested form of the chemical [file created with ACD ChemFolder ver. 6.0].

File Type	Description	File Size	Format
Document Files			
Log File	NCTRER_LogFile_15Aug03.pdf	68KB	
Field Definition File	NCTRER_FieldDefFile_15Aug03.pdf	111KB	
	NCTRER_FieldDefFile_15Aug03.doc	73KB	
Data Files			
SDF Structure/Data File	NCTRER_v1a_232_15Aug03.sdf	329KB	
- Data Table (no structures)	NCTRER_v1a_232_15Aug03_nostructures.xls	98KB	
- Structures Table	NCTRER_v1a_232_15Aug03_structures.pdf	359KB	

File Error Report

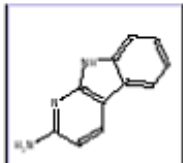
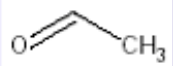
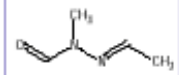
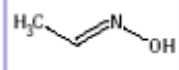
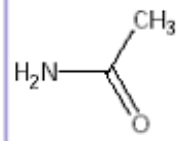
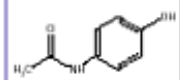
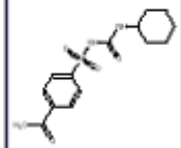
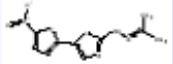
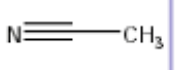
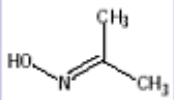
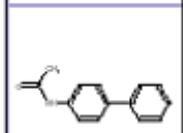
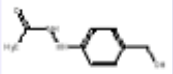
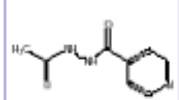
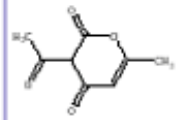
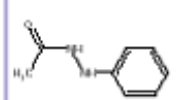
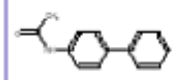
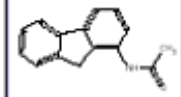
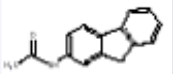
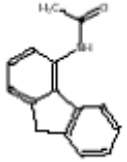
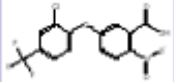
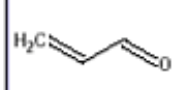
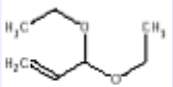
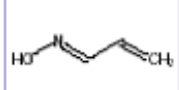
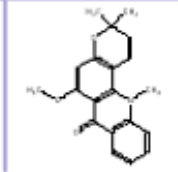
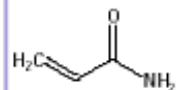
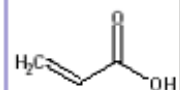
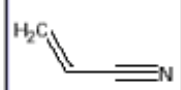
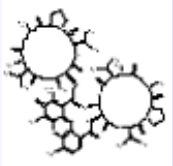
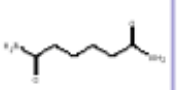
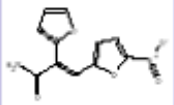
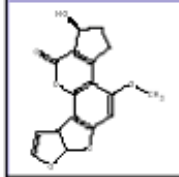
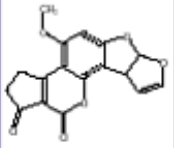
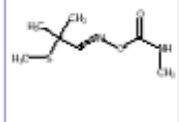
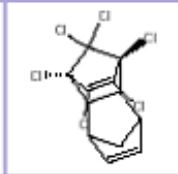
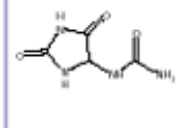
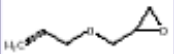
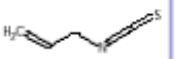
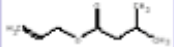
Acknowledgements: The DSSTox SDF file for the NCTRER was expanded from an original SDF file kindly provided by the NCTR Source, Weida Tong. The file was converted to DSSTox format by ClarLynda Williams (EPA/NC Central Univ Student COOP; EPA) with the assistance of Jamie Burch (EPA/NC Central Univ Student COOP). Additional ER-related data fields were added by Ann Richard (EPA) and the Source collaborators, Weida Tong and Hong Fang, both of NCTR.

DSSTox Citation: Tong, W., H. Fang, C.R. Williams, J.M. Burch, and A.M. Richard (2003) National Center for Toxicological Research Estrogen Receptor Binding Database (NCTRER): SDF files and website documentation, www.epa.gov/nheerl/dsstox/

Disclaimer: Every effort is made to ensure that DSSTox SDF files and associated documentation are error-free, but neither the DSSTox Source collaborators nor the EPA DSSTox project team make guarantees of accuracy, nor are any of these persons to be held liable for any subsequent use of these public data. The contents of this webpage and supporting documents have been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Content of this page last updated 15 August 03.

CPDBRM_v1a_1354_01Apr03: Structure tiling view pdf (ACD ChemFolder)

									
A-alpha-C 26148-68-5	Acetaldehyde 75-07-0	Acetaldehyde methyl 16568-02-8	Acetaldoxime 107-29-9	Acetamide 60-35-5	Acetaminophen 103-90-2	Acetohexamide 968-81-0	Acetone[4-(5-nitro-2 18523-69-8	Acetonitrile 75-05-8	Acetoxime 127-06-0
									
1'-Acetoxysafrole 34627-78-6	N'-Acetyl-4-(hydrox 65734-38-5	1-Acetyl-2-isonicotin 1078-38-2	3-Acetyl-6-methyl-2, 520-45-6	1-Acetyl-2-phenylhy 114-83-0	4-Acetylaminobiphe 4075-79-0	1-Acetylaminofluore 28314-03-6	2-Acetylaminofluore 53-96-3	4-Acetylaminofluore 28322-02-3	Acifluorfen 50594-66-6
									
Acrolein 107-02-8	Acrolein diethylacet 3054-95-3	Acrolein oxime 5314-33-0	Acronycine 7008-42-6	Acrylamide 79-06-1	Acrylic acid 79-10-7	Acrylonitrile 107-13-1	Actinomycin D 50-76-0	Adipamide 628-94-4	AF-2 3688-53-7
									
Aflatoxicol 29611-03-8	Aflatoxin B1 1162-65-8	Aldicarb 116-06-3	Aldrin 309-00-2	Allantoin 97-59-6	Allyl alcohol 107-18-6	Allyl chloride 107-05-1	Allyl glycidyl ether 106-92-3	Allyl isocyanate 57-06-7	Allyl isovalerate 2835-39-4
1-Allyl-1-nitrosoure 760-56-5	Allylhydrazine.HCl 52207-83-7	Aluminum potassiu 10043-67-1	1-Amino-2,4-dibrom 81-49-2	3-Amino-4-ethoxyac 17026-81-2	3-Amino-9-ethylcarb 6109-97-3	1-Amino-2-methylan 82-28-0	2-Amino-5-(5-nitro- 3775-55-1	2-Amino-5-(5-nitro- 712-68-5	2-Amino-4-(5-nitro- 38514-71-5

DSSTox Log File:

- NAMEID: Database Title
- Description
- QA & Development Notes
- Field and Data Counts in SDF
- Replicate Information in SDF
- Missing CAS or structure info
- Log of SDF Modifications

DSSTox Log File:

Carcinogenic Potency Database Summary Tables
(CPDBRM, CPDBHA, CPDBDG, CPDBPR)
(last updated 25 July 03)

Description: Information in this file documents the creation, review, and update process for the DSSTox CPDB SDF files, provides summary information on database contents, and lists currently unavailable CAS information for known structures. The first section summarizes the process used for creating the initial DSSTox SDF files and the quality assurance checks and procedures employed. A table providing field and data counts offers summary overview of CPDB file contents and chemical composition. A second table provides summary counts of various types of replicate chemical information in the various CPDB files. The Log table will document any future modifications and revisions to the database content or format. For the most current version of this Log File and a record of any new modifications, a user should periodically consult the central DSSTox website: http://www.epa.gov/nheerl/dsstox/

QA and Development Notes:
CPDB SDF files underwent an extensive series of quality review checks prior to publication of initial launch versions. Source field entries (i.e. non-DSSTox Standard fields) were thoroughly checked by visual inspection for correspondence to original CPDB Summary Tables. We thank Lois Swirsky Gold and Thomas H. Slone for valuable assistance in ongoing quality review of the DSSTox CPDB files, helping to ensure that data are accurately extracted and represented from the original CPDB Summary Tables. They pointed out numerous systematic and human-error problems early in the DSSTox project and early in the process of CPDB SDF development, carefully reviewed DSSTox field definitions and offered suggestions for improving and finalizing all documentation files, and worked with the DSSTox team to find missing structures and reconcile remaining discrepancies in CAS numbers from the original CPDB Summary Tables. Chemical structures were initially obtained by automated filling from large in-house databases of CAS-referenced structures (American Chemicals Directory, NCI Structure Database). The ChemFinder website (http://chemfinder.cambridgesoft.com/) was used extensively for checking CAS-to-structures and for retrieving CAS numbers for parent forms of salts and complexes. CambridgeSoft's ChemOffice 2002 ChemFinder (ver 7.0 for Windows) was used for automatic generation of SMILES codes from structures and both ChemFinder and ACD ChemFolder (ver 6.0 for Windows) were employed for "Structure-to-Name" or "Name-to-Structure" features. **ChemName**, **SMILES**, **CAS** and **Structure** field contents were checked by cross-referencing wherever possible. The CPDBRM_DOP (defined organic parent) SDF file was created by exporting only defined organics to SDF from the Main ChemFinder file for CPDBRM, and converting salts and complexes to their simplified form, with changes to corresponding Standard Chemical Fields.

Field and Data Counts in DSSTox SDF files: Refer to CPDB_FieldDefFile for definitions and explanations of all terms.

DSSTox SDF	Standard Chemical Fields	Source-specific fields	Chemical records total	Defined organic	Inorganic	Organo-metallic	Mixture or unknown	Parent	Salt or Salt complex	Complex
CPDBRM_v1a	13	10	1354	1189	52	39	74	1016	99	165
CPDBRM_DOP_v1a	15	10	1189	1189	0	0	0	1000	67	122
CPDBHA_v1a	12	6	80	72	6	1	1	67	5	7
CPDBDG_v1a	12	4	5	5	0	0	0	4	1	0
CPDBPR_v1a	12	10	27	24	1	0	2	21	3	1

Replicate Information in CPDBRM SDF File: The term "replicate" refers to possibly redundant information in the chemical structure fields. All replicate cases can be easily located by search of the **ChemNote** and **ChemCount** fields in CPDBRM (refer also to CPDB_FieldDefFile).

CPDBRM: Replicate Type	Sets of Replicate
CAS [1]	12
2D structures [2]	7
Parent structures [3]	27
Totals	46

[1] replicate CAS: same CAS number (e.g., if differen
carcinogenicity).
[2] replicate 2D structure: geometric or stereoisomers
[3] replicate parent structures: salt or complex of sam

Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Name	Modifications from previous version	Additional Notes
25Jul03 25Jul03 25Jul03 25Jul03 25Jul03	CPDBRM_v1a_1354_25Jul03.sdf CPDBRM_DOP_v1a_1188_25Jul03.sdf CPDBPR_v1a_27_25Jul03.sdf CPDBHA_v1a_80_25Jul03.sdf CPDBDG_v1a_5_25Jul03.sdf	Initial launch publication; no previous versions.	Working with Source collaborators (L.S. Gold and T. H. Slone), periodic version updates to the DSSTox CPDB SDF files (i.e., v1, v2, etc.) will incorporate new information provided in updates to the CPDB Summary Tables and posted on the Source CPDB website, http://potency.berkeley.edu/. In addition, revision updates (e.g., v1a, v1b, etc) will correct reported errors or add missing data provided by users or the Source.

DSSTox Field Definition File:

- NAMEID: Database Title

- Description
- Source Website
- Source Contact
- Main Citation
- SDF Development Notes

- DSSTox Standard Chemical Fields
- Source-Specific Fields

DSSTox Field Definition File:

Carcinogenic Potency Database Summary Tables (CPDBRM, CPDBHA, CPDBDG, CPDBPR)

(last updated 28 July 03)

Description: Information in this file is intended to provide a minimum level of annotation to the DSSTox SDF files created for the Carcinogenic Potency Database Summary Tables (CPDBRM= rats and mice, CPDBHA= hamsters, CPDBDG= dog, CPDBPR= non-human primates) obtained from the CPDB Source website: http://potency.berkeley.edu/. For further explanation of Source-specific fields, a user is encouraged to consult the CPDB website, listed references, and documentation. Some modifications in fields (and allowable contents) were made to the original CPDB Summary Tables to facilitate use of the DSSTox SDF files in relational searching applications. All modifications are fully documented in the **Comments** section of the table below.

The first section of the table below lists and defines the ***DSSTox Standard Chemical Fields*** used in the CPDB SDF files. Any modifications in these fields, deviating either from the original Source data tables or the ***Central List of DSSTox Standard Chemical Fields*** are noted in the **Comments** section. Following that section, all Source-specific fields in the DSSTox SDF files (i.e., CPDBRM, CPDBHA, CPDBDG, CPDBPR) are listed and defined. The ***DSSTox SDF*** column lists files in which the corresponding ***Field Name*** is present. All ***Units*** and ***Descriptions*** are extracted from Source reference materials unless otherwise noted. In some cases, modifications in ***Field Name*** and ***Allowable Values*** from the original data tables were made to facilitate creation and use of the DSSTox SDF files. All differences are noted in the ***Comments*** section. ***Allowable Values*** list allowable field entries occurring in CPDB SDF files separated by slashes for exclusive entries (i.e. cannot occur with another entry) and commas or spaces for non-exclusive entries (i.e. can occur with other values). These codes are defined and explained in the ***Description*** section; italicized note refers to the type of entry (e.g., *Text*). The pound symbol (#) indicates that the ***Allowable Values*** entry is a number. A pound symbol followed by a list of character options (e.g., #i, m, n, v) indicates that one or more footnote characters may follow the number entry; these are meant to provide additional information and are defined in the ***Description*** section. To minimize problems with import and export of SDF files, we avoid the use of punctuation and symbols in ***Allowable Values*** wherever possible; multiple entries in a single field (e.g., adr cli eso) are separated by a single space in the SDF. Upper and lower cases in ***Allowable Values*** text entries are used only for emphasis, and never alone to distinguish separate meaning.

Source Website: The CPDB, from which the Summary Tables are derived, is available in several formats at http://potency.berkeley.edu/

Source Contact: Please contact Lois Swirsky Gold for questions pertaining to the content of the CPDB Summary Tables; email: cpdb@potency.berkeley.edu. Please contact DSSTox Support for questions or comments pertaining to the DSSTox CPDB SDF files.

Main Citations: Publications reporting use of DSSTox SDF files for the CPDB Summary Tables are asked to list the full DSSTox file name(s), including date stamp, and to cite as primary references the following:

Gold, L.S., Slone, T.H., Ames, B.N., Manley, N.B., Garfinkel, G.B., and Rohrbach, L. (1997) Carcinogenic Potency Database. In: Gold, L.S., and Zeiger, E., Eds. Handbook of Carcinogenic Potency and Genotoxicity Databases. Boca Raton, FL: CRC Press, pp. 1-605.

Gold, L.S., Manley, N.B., Slone, T.H., and Rohrbach, L. (1999) Supplement to the Carcinogenic Potency Database (CPDB): Results of animal bioassays published in the general literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environ. Health Perspect.* 107 (Suppl. 4): 527-600.

SDF Development Notes:
Each DSSTox SDF file contains a single **Structure** field whose entry corresponds to the **StructureShown**, **CAS**, **SMILES**, **Formula**, and **MolWeight** fields. The main DSSTox SDF files represent the actual tested form of the chemical in the **Structure** field (see ***Description*** below), including complexed molecular entities and salt counter ions in all cases. An additional DSSTox "Defined Organic Parent" SDF file (CPDBRM_DOP) is offered for download only for the largest CPDBRM file, for specialized use in Structure-Activity Relationship (SAR) modeling applications. This DOP file contains no inorganics, organometallics, or mixtures, and all defined organic salts and complexes are stripped of counter-ions and complexed molecular entities and converted to a simplified parent representation in the **Structure** field. The **StructureShown** entry for these compounds is "simplified to parent", with corresponding changes in the **CAS**, **SMILES**, **Formula**, and **MolWeight** field entries. These "simplified to parent" structures are represented in neutralized (protonated) form wherever possible (exceptions include quaternary ammonium and pyridinium ions, which are represented as positively charged (N+) stripped of counter ions, and nitro compounds, which are represented in the charge-separated form, i.e. N+(=O)O-). In the DOP file, both a **CAS**_**TestedForm** and **SMILES**_**TestedForm** field are

Field Name	*DSSTox SDF*	*Units*	*Allowable Values*	*Description*	*Comments*
				DSSTox Standard Chemical Fields	
Structure	All		*Molecule*	Two-dimensional graphical representation of molecular structure. Form of structure is identified in the **StructureShown** field and	Structures not provided in original CPDB Summary Tables.
				CPDB Source-Specific Fields	
SAL CPDB	CPDBRM CPDBHA CPDBDG CPDBPR	*None*	pos/ neg/ NE/	A chemical is classified within the CPDB as mutagenic, i.e. "pos", in the *Salmonella* assay if it was evaluated overall as either "mutagenic" or "weakly mutagenic" by Zeiger [1] or as overall "positive" by the EPA Gene-Tox Program [2]. All other chemicals evaluated for mutagenicity by	This field is titled "Salmonella" in the original CPDB Summary Tables; symbol entries appearing in this field were converted to the following DSSTox text equivalents:

EPAFHM_v1a_617_15Oct03: Source-specific field

FieldName	Allowable Values	Description
MOA	**NARCOSIS_I/** **NARCOSIS II/** **NARCOSIS III/** **NARCOSIS I and II/** **UNCOUPLER/** **ACHE/** **BLOCKER/** **REACTIVE/** **NEUROTOX/** **NEURODEP/** **UNSURE/** **MIXED/** **ND/**	**Mode-of-action of chemical assigned by authors of study based on joint toxic action studies, establishment of toxicodynamic profiles, and behavioral and dose-response interpretation of 96 h (hour) LC50 tests. MOA field entries are defined below, with further explanation provided in Main Citation listed above (Russom et al., 1997):** **ARCOSIS I = Base-line narcosis, or Narcosis I MOA** **ARCOSIS II = Polar narcosis, or Narcosis II MOA** **ARCOSIS III = Narcosis III MOA primarily observed in esters and some acrylates** **ARCOSIS I and II = Identified as both Narcosis I & II MOA** **NCOUPLER = Uncoupler of oxidative phosphorylation MOA** **CHE = Acetylcholinesterase Inhibition MOA** **LOCKER = Respiratory blocker/inhibitor MOA** **EACTIVE = Electrophile/proelectrophile reactivity MOA** **EUROTOX = Central nervous system seizure/stimulant MOA** **EURODEP = Neurodepressant MOA** **NSURE = MOA could not be determined - insufficient evidence** **IXED = MOA could not be determined - conflicting evidence** **D = MOA was not determined either because the chemical was not toxic at saturation or the test result was obtained after the MOA analysis was conducted.**

DSSTox Central Field Definition Table

(last updated 11 November 03)

Indexed DSSTox SDF Files Included in Table:

Links provided to corresponding DSSTox Source SDF Download Page for each database listed by NAMEID.

CPDB: Carcinogenic Potency Database Summary Tables (CPDBRM, CPDBHA, CPDBDG, CPDBPR)

DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates

EPAFHM: EPA Fathead Minnow Acute Toxicity Database

NCTRER: NCTR Estrogen Receptor Binding Database

Central Index of DSSTox Databases

The table below contains an alphabetically indexed central listing of all fields contained in all DSSTox SDF files currently offered for download on this website. *DSSTox Standard Chemical Fields* are included in this listing but are separately designated. For each field indexed in this table, the DSSTox SDF file(s) in which the field is contained is listed under the column *DSSTox SDF*, providing a link to the DSSTox Source SDF Download Page containing full reference documentation for that database. This consolidated table provides abbreviated content compared to each separate NAMEID_FieldDefFile (NAMEID=CPDBRM, EPAFHM, etc.) reference document offered for download on each DSSTox Source SDF Download Page listed above.

If a field is indicated to be a DSSTox Standard Chemical Field (yellow highlighted), a link is provided to the More on DSSTox Standard Chemical Fields general information page. For more complete information, a user is also referred to the main reference document: DSSTox Standard Chemical Field Definition File.

All DSSTox fields listed alphabetically

entries occurring in DSSTox SDF files, separated by slashes (/) for exclusive entries (i.e. cannot occur with another entry) and commas or spaces for non-
ese codes are defined and explaine
., *Text*). The pound symbol (#) indicates that
nbol followed by a list of character
nay follow the number entry; these provide
section. To minimize problems wit
nbols in *Allowable Values* wherever possible;
parated by a single space in the SD
for emphasis, and not alone to distinguish
separa
except in the case of SMILES codes, which are case-sensitive.

Link to DSSTox database containing field

References
the *Description* column in the table below are listed in the corresponding DSS
ection following this table (reference sections listed alphabetically by NAMEID).

Field Name	*Field Type*	*DSSTox SDF*	*Units*	*Allowable Values*	*Description*
ActivityCategoryER_RBA		NCTRER	None	active strong/ active medium/ active weak/ slight binder/ inactive/	For purposes of SAR analysis, Fang et al. (2001) divided the NCTRER data set into five main activity categories: active strong (ER_RBA > 1), active medium (1> ER_RBA > 0.01), active weak (0.01 > ER_RBA > 1E-5), slight binder (max< 50% inhibition or ER_RBA< 1E-5) inactive (no activity, equates with NA designation)
AddToParent	DSSTox Standard Chemical Fields	All DSSTox SDF files containing salts or complexes	None	*Text*	For **SubstanceType**="defined organic" and **TestedForm**="salt" or "complex", entry specifies salt counter-ions or complexed entities (e.g., Na, K, HCl, Cl, H2O, Ca , H2SO4, acetate, etc.) that are removed when **StructureShown**="simplified to parent" in DOP file; "bis" signifies parent structure occurs twice in complex.
AnalogCAS		DBPCAN	None	NOCAS/ ########-##-#	CAS of primary structural analog cited in SAR rationale for carcinogenic potential prediction, corresponding to **AnalogName**.
AnalogChemName		DBPCAN	None	*Text*	Chemical name of primary structural analog cited in **Rationale** for SAR carcinogenic potential prediction listed in Table 1.
AnalogSMILES					SMILES code of primary structural analog cited in SAR rationale for carcinogenic potential prediction, corresponding to **AnalogName**.
BEHAVIOR				TYPE III/ TYPE III/ CONFLICT/ EL, ID, PHPROB/	Behavior signs of stress were identified for fathead minnows exposed to toxicants and were used to classify chemicals into three behavioral syndromes as described by Drummond and Russom [8]. These were used to determine level of confidence of MOA assignment. TYPE I = depressed locomotor activity with little or no response to outside stimuli, darkened body color, most fish dead by 24 h TYPE II = hyperactive, usually overreactive to outside stimuli, death typically within several days of exposure

Link to Standard Chemical Field Definitions Table

e.g., DSSTox Citation:

Gold, L.S., T.H. Slone, C.R. Williams, J.M. Burch, T.W. Stewart, A.E. Swank, J. Beidler, and A.M. Richard (2003) *DSSTox Carcinogenic Potency Database Summary Tables for Rats and Mice, Hamsters, Dogs, and Non-human Primates (CPDBRM, CPDBHA, CPDBDG, CPDBPR): SDF Files and Documentation*, www.epa.gov/nheerl/dsstox/

- Public forum for "publishing" toxicity databases
- Sources retain prominent "authorship" of databases
- Construct accommodates diverse database content
- Users take what they want and use however they want
- Citation will communicate standards and expectations (database files, documentation, review)

Integrating Diverse Databases from a Chemical Structure Perspective:

CPDB DBPCAN EPAFHM NCTRER ….

Standard Chemical Fields

CPDB	DBPCAN	EPAFHM	NCTRER
SAL CPDB	ChemClass DBP	ChemClass FHM	NCTRlogRBA
TD50 Rat	Concern Level	MOA	ER RBA
TD50Mouse	Rationale	MOACONF	ChemClass ERB
Target Sites Rat Male	Rational Source	CLOGP	Activity Group ERB
Target Sites Rat Female	Analog ChemName	LC50	Rationale ChemClass ERB
Target Sites Mouse Male	AnalogCAS	LC50NOTE	MeanChem Class ERB RBA
…. Other Species	Analog SMILES	LC50RATIO	LogP
		MIXMOA	F1, F2, …F6
		TOXINDEX	
		FATS	
		BEHAVIOR	

DSSTox Database Network:

- What's next ?
- How can these data files be used?

Begin to incorporate standard tox fields (TOXML)

CPDB DBPCAN EPAFHM NCTRER ….

Standard Chemical Fields

Standard Tox Fields: *species, sex, strain, assay, dose*

CPDB	DBPCAN	EPAFHM	NCTRER
SAL CPDB	ChemClass DBP	ChemClass FHM	NCTRlogRBA
TD50 Rat	Concern Level	MOA	ER RBA
TD50Mouse	Rationale	MOACONF	ChemClass ERB
Target Sites Rat Male	Rational Source	CLOGP	Activity Group ERB
Target Sites Rat Female	Analog ChemName	LC50	Rationale ChemClass ERB
Target Sites Mouse Male	AnalogCAS	LC50NOTE	MeanChem Class ERB RBA
…. Other Species	Analog SMILES	LC50RATIO	LogP
		MIXMOA	F1, F2, …F6
		TOXINDEX	
		FATS	
		BEHAVIOR	

Migrate More Public Toxicity Data into DSSTox Standard Format: Phase II, III, …

- NCTR Androgen, Thyroid, and Endocrine Disruption Databases
- NTP Rodent carcinogenicity bioassays, subchronic bioassays, developmental, repro, immuno, etc.
- ICVAM databases on LD50, skin sensitization, local lymph node assay, skin corrosivity, endocrine disruption, etc
- EPA's Teratox and Aquire ecotoxicity databases
- EPA's High Production Volume (HPV) chemical data
- EPA's Integrated Risk Information System
- Developmental toxicity database (literature - TOPKAT)
- UniLever Skin Sensitization database
- Public toxicity data for FDA pharmaceuticals (MRTD), human, clinical

http://www.fda.gov/cder/Offices/OPS_IO/MRTD.htm

FDA U.S. Food and Drug Administration — Department of Health and Human Services

CENTER FOR DRUG EVALUATION AND RESEARCH

FDA Home Page | CDER Home Page | CDER Site Info | Contact CDER | What's New @ CDER

CDER Home | About CDER | Drug Information | Regulatory Guidance | CDER Calendar | Specific Audiences | CDER Archives

Search GO powered by Google™

Maximum Recommended Therapeutic Dose (MRTD) Database

FDA's Center for Drug Evaluation and Research, Office of Pharmaceutical Science, Informatics and Computational Safety Analysis Staff's Maximum Recommended Therapeutic Dose (MRTD) database contains values for 1,235 pharmaceuticals listed in *Martindale: The Extra Pharmacopoeia* (1973, 1983, and 1993) and *The Physicians' Desk Reference* (1995 and 1999).

Most of the MRTD values in the database were determined from p
route of exposure and daily treatments, usually for 3 - 12 months. T
dose treatment regimens to achieve desired pharmacological effects
the MRTD database were antineoplastics and anesthetics and were
When separate MRTDs were reported for different routes of expos
database. In addition, some pharmaceuticals have different MRTD
elderly patients. In this situation only MRTD values for the average

Pharmaceuticals that are administered orally are usually tested over
as mg/day. We converted the mg/day unit to mg/kg-body weight (b
kg. In contrast, the dose unit for most antineoplastic drug MRTDs i
bw/day using the formula mg/kg-bw/day = mg/m^2/37 for an averag
reported in parts per million (ppm) which were converted to mg/kg
mg/kg-bw/day for an average 60 kg adult. MRTD values for the 12
1000 mg/kg-bw/day.

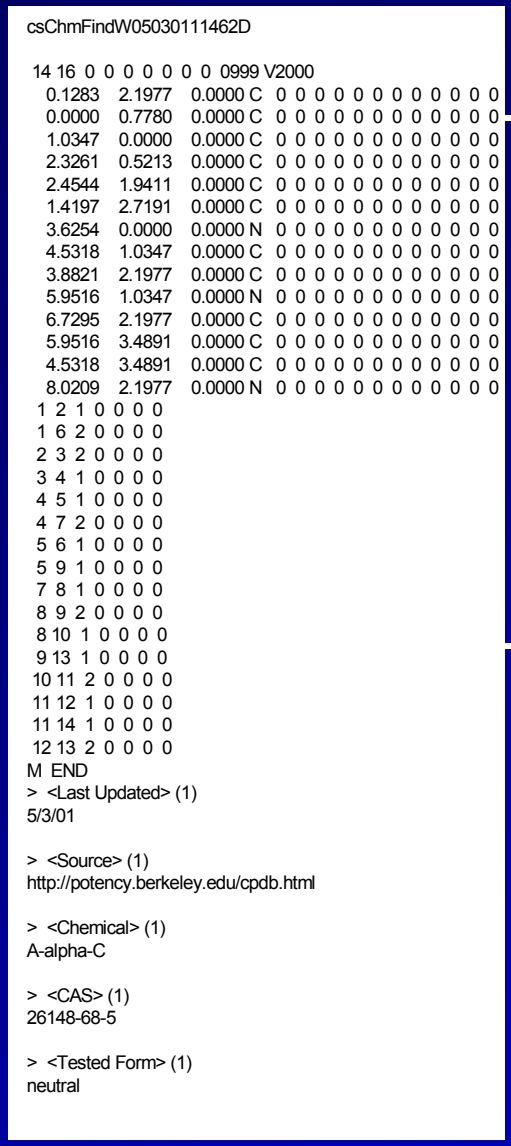
Comments or corrections should be sent to: benzrd@cder.fda.gov.

Maximum Recommended Therapeutic Dose (MRTD) Database (1,235 pharmaceuticals)

A B C D E F-G H-K L M N O P Q-R S T U-Z

MRTD (mg/kg-bw/day)	Generic Chemical Name	Chemical Structure (SMILES Code)
3.00000	Acemetacin	C1=C(OC)C=C3C(=C1)N(C(=O)C2=CC=C(Cl)C=C2)C(C)=C3CC(=O)OCC(O)=O
0.20000	Acenocoumarol	C1=CC=C3C(=C1)OC(=O)C(C(CC(C)=O)C2=CC=C(N(=O)=O)C=C2)=C3O
50.00000	Acetaminophen	C1(O)=CC=C(NC(C)=O)C=C1
5.00000	Acetanilide	C1=CC=CC=C1NC(C)=O
16.70000	Acetazolamide	N1N=C(S(N)(=O)=O)SC=1NC(C)=O
25.00000	Acetohexamide	C2=C(C(C)=O)C=CC(S(=O)(=O)NC(=O)NC1CCCCC1)=C2
16.70000	Acetohydroxamic acid	CC(=O)NO
10.00000	Acetophenazine	C1=CC=C3C(=C1)N(CCCN2CCN(CCO)CC2)C4=C(S3)C=CC(C(C)=O)=C4
66.70000	Acetosulfone	C2=C(N)C=CC(S(=O)(=O)C1=CC=C(N)C=C1S(=O)(=O)NC(C)=O)=C2

DSSTox SDF files



```
csChmFindW05030111462D

 14 16  0  0  0  0  0  0  0  0999 V2000
   0.1283    2.1977    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   0.0000    0.7780    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   1.0347    0.0000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   2.3261    0.5213    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   2.4544    1.9411    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   1.4197    2.7191    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   3.6254    0.0000    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   4.5318    1.0347    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   3.8821    2.1977    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   5.9516    1.0347    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   6.7295    2.1977    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   5.9516    3.4891    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   4.5318    3.4891    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   8.0209    2.1977    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  1  6  2  0  0  0  0
  2  3  2  0  0  0  0
  3  4  1  0  0  0  0
  4  5  1  0  0  0  0
  4  7  2  0  0  0  0
  5  6  1  0  0  0  0
  5  9  1  0  0  0  0
  7  8  1  0  0  0  0
  8  9  2  0  0  0  0
  8 10  1  0  0  0  0
  9 13  1  0  0  0  0
 10 11  2  0  0  0  0
 11 12  1  0  0  0  0
 11 14  1  0  0  0  0
 12 13  2  0  0  0  0
M  END
>  <Last Updated> (1)
5/3/01

>  <Source> (1)
http://potency.berkeley.edu/cpdb.html

>  <Chemical> (1)
A-alpha-C

>  <CAS> (1)
26148-68-5

>  <Tested Form> (1)
neutral
```

Convert to 3D

Add chem/phys properties

Port to different file formats

Merge into central database

Add/remove fields

Link to internal databases

SAR Model Development "Training Sets"

- improved predictive tox models
- more comparable models
- dramatically lowered barriers to use

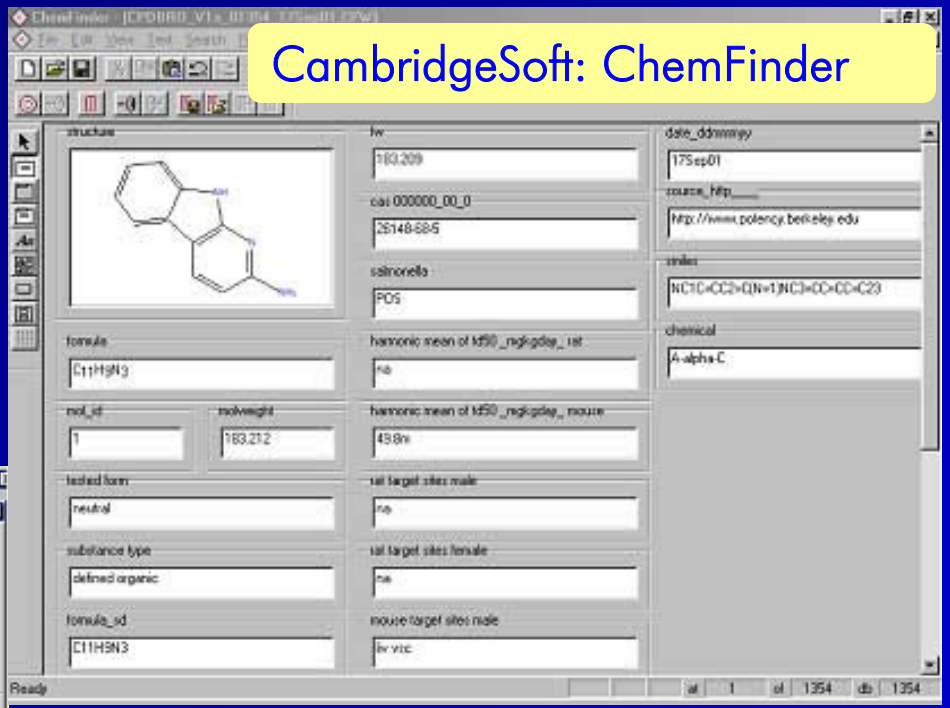
Chemical Relational Database:

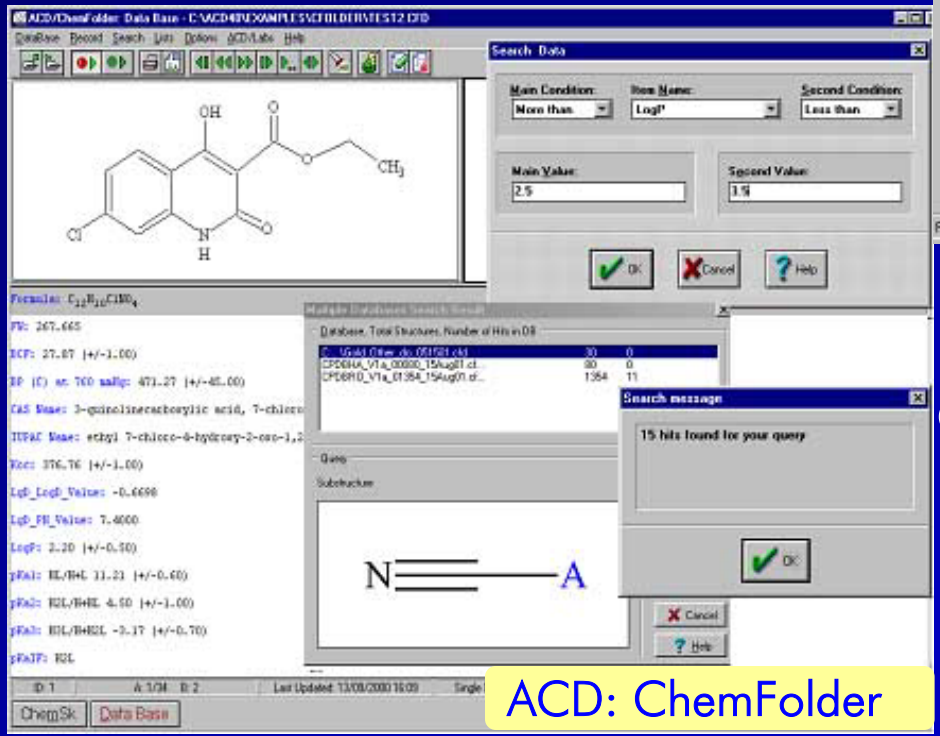
sub-structure, text, property searching

- analog searches
- search across diverse toxicity endpoints
- search across chemical and toxicity fields

Chemical Relational Databases: *Exploration across toxicological domains and structural/biological axes*

Accord
Oracle
ISIS
ChemFolder
ChemFinder
LeadScope





Free SDF Viewer Application
- Tom Harrocks, IntuitiveSoftwareSolutions

Off-site on-line structure searching
- NCI Structure Browser
- NLM ToxNet

EPA Server-based on-line searching

ChemFinder.com

Public site offers CAS, Name, and Structure-searching capabilities

ChemFinder.Com
Database & Internet Searching

ChemStore.Com | ChemFinder.Com | ChemNews.Com | ChemClub.Com | CambridgeSoft.Com

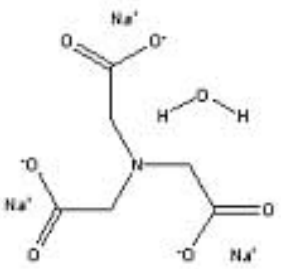
CALL ME NOW

Online Training : ChemDraw 8.0
CHEMICAL STRUCTURE DRAWING STANDARD
REGISTER
» November 19, 2003 Join from your desk, classroom, or office

Enter a Chemical Name, CAS Number, Molecular Formula or Weight.
Use * for partial names (e.g. ben*).
Search here for free. For professional searching, use ChemINDEX.

18662-53-8 Search

Nitrilotriacetic acid trisodium salt monohydrate [18662-53-
Synonyms: hampshire nta na3; nitrilotriacetic acid trisodium salt monohyd
N,N-bis(carboxymethyl)glycine trisodium salt monohydrate; perma kleer nt
nitrilotriacetate monohydrate; nta sodium hydrate; trilon a92; Trisodium ni
monohydrate; trisodium salt of nitrilotriacetic acid, monohydrate;



Tools
BUY AT CHEMACX.COM
VIEW CHEMDRAW STRUCT
VIEW CHEM3D MODEL
CAS RN Lookup
THE MERCK INDEX
NCI DATABASE

Formula	C6H8NNa3O7	Molecular Weight	275.10101
CAS RN	18662-53-8	Melting Point (°C)	~410
ACX Number	X1005204-9	Boiling Point (°C)	
Density		Vapor Density	
Refractive Index		Vapor Pressure	
Evaporation Rate		Water Solubility	>=10 g/100 mL at
Flash Point (°C)		EPA Code	
DOT Number		RTECS	AJ1070000
Comments	White crystalline powder		

More information about the chemical is availabl

Chemical Online Order (1)
Available Chemicals Exchange
Information about this particular compound

Health (5)
8(e) TRIAGE Chemical Studies Database
Berkeley Carcinogenic Potency Database
National Toxicology Program (NTP) publications
Information about this particular compound
NTP Chemical Health and Safety Data
Information about this particular compound
UMCP Partial list of teratogens
DSSTox: CPDBRM

ic Potency Project (CPDB) - Netscape
The Carcinogenic Potency Project
Lois Swirsky Gold, Ph.D., Director

U.S. Environmental Protection Agency
Distributed Structure-Searchable Toxicity (DSSTox)

Site Map
About DSSTox
Recent Additions
Frequently Asked Questions
DSSTox SDF Files
Central Field Index
Apps, Tools & More
DSSTox Community
Help

DSSTox *Source SDF Download Page*

CPDBRM, CPDBHA, CPDBDO, CPDBPR:
Carcinogenic Potency Database Summary Tables
for Rats and Mice, Hamster, Dogs, and Non-Human Primates

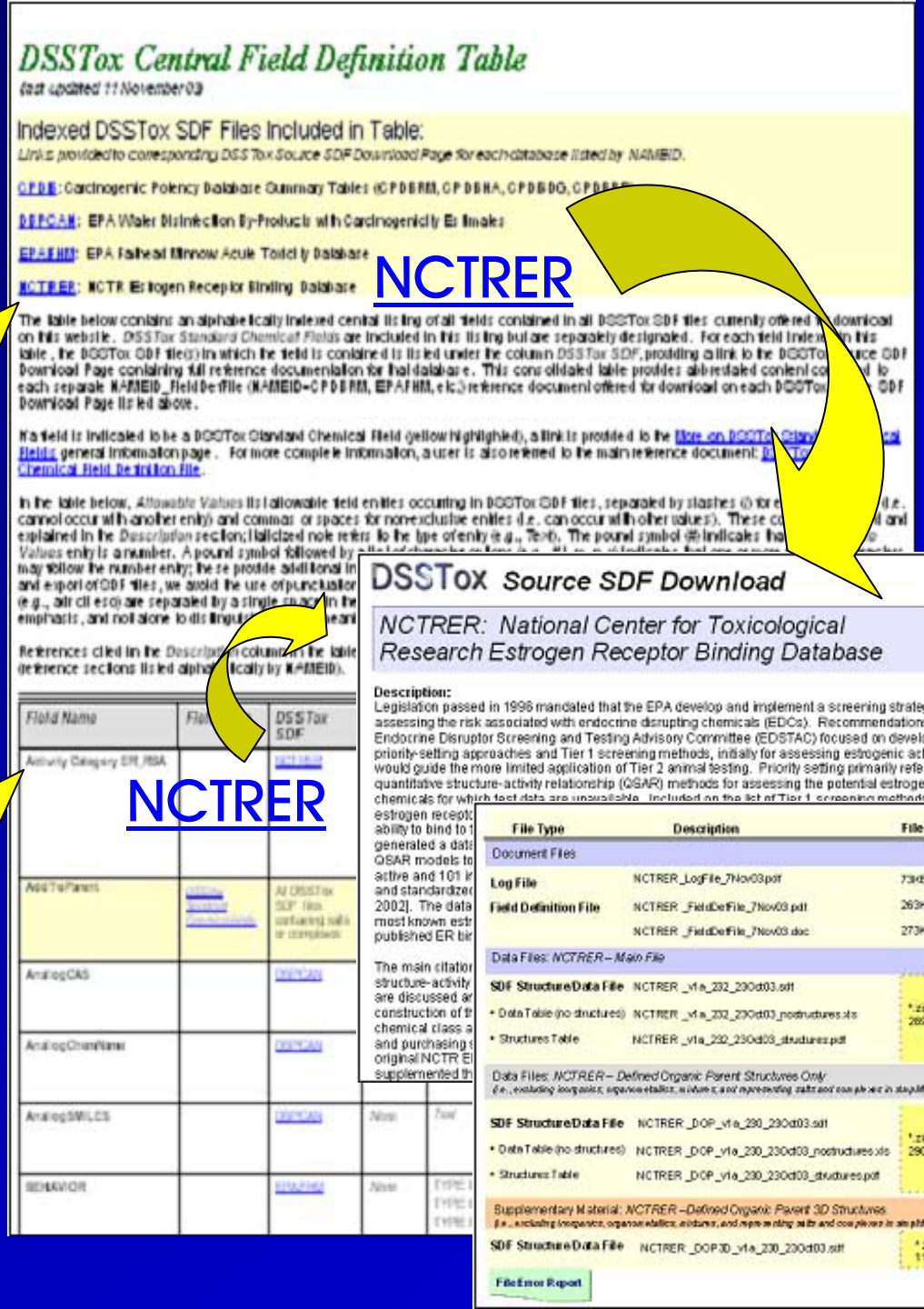
Description: The CPDB Summary Tables list summarized results for experiments on 1370 substances in the Carcinogenic Potency Database (CPDB). These Summary Tables represent one of many possible summarizations of the data in the CPDB. The CPDB, which continues to be expanded, includes detailed results and analyses of more than 5000 chronic, long term carcinogenesis bioassays reported in over 1200 papers in the general literature and more than 400 Technical Reports of the National Cancer Institute/National Toxicology Program. Details

Web-based Structure-Searching of DSSTox SDF Files:

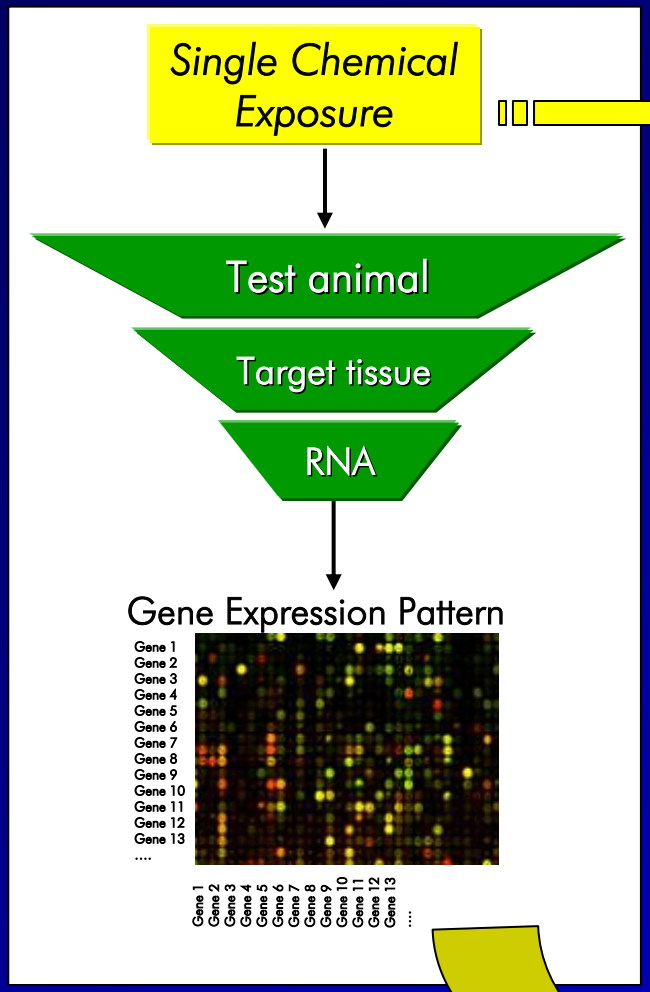
- NCI Structure Browser
- NLM ToxNet
- ACD Web-Librarian
- SRC/EPA PBT Profiler

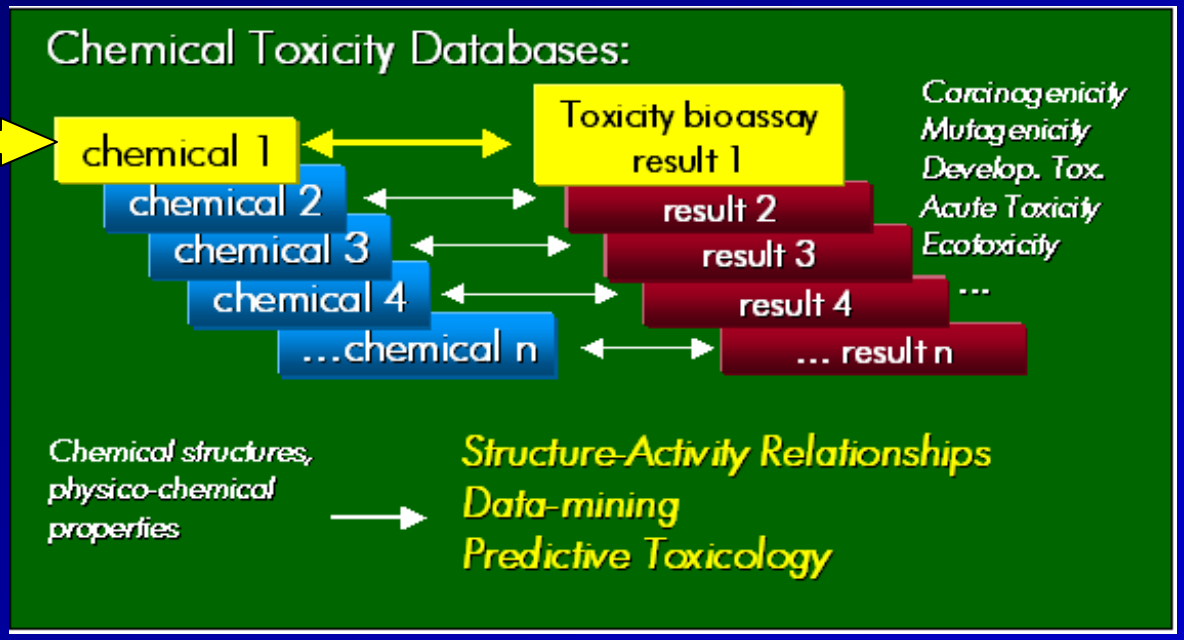
Structure-search displays chemical record in DSSTox: NCTRER

User seeks to learn more about NCTRER and its data fields



Bioinformatics ….. meets Chemoinformatics

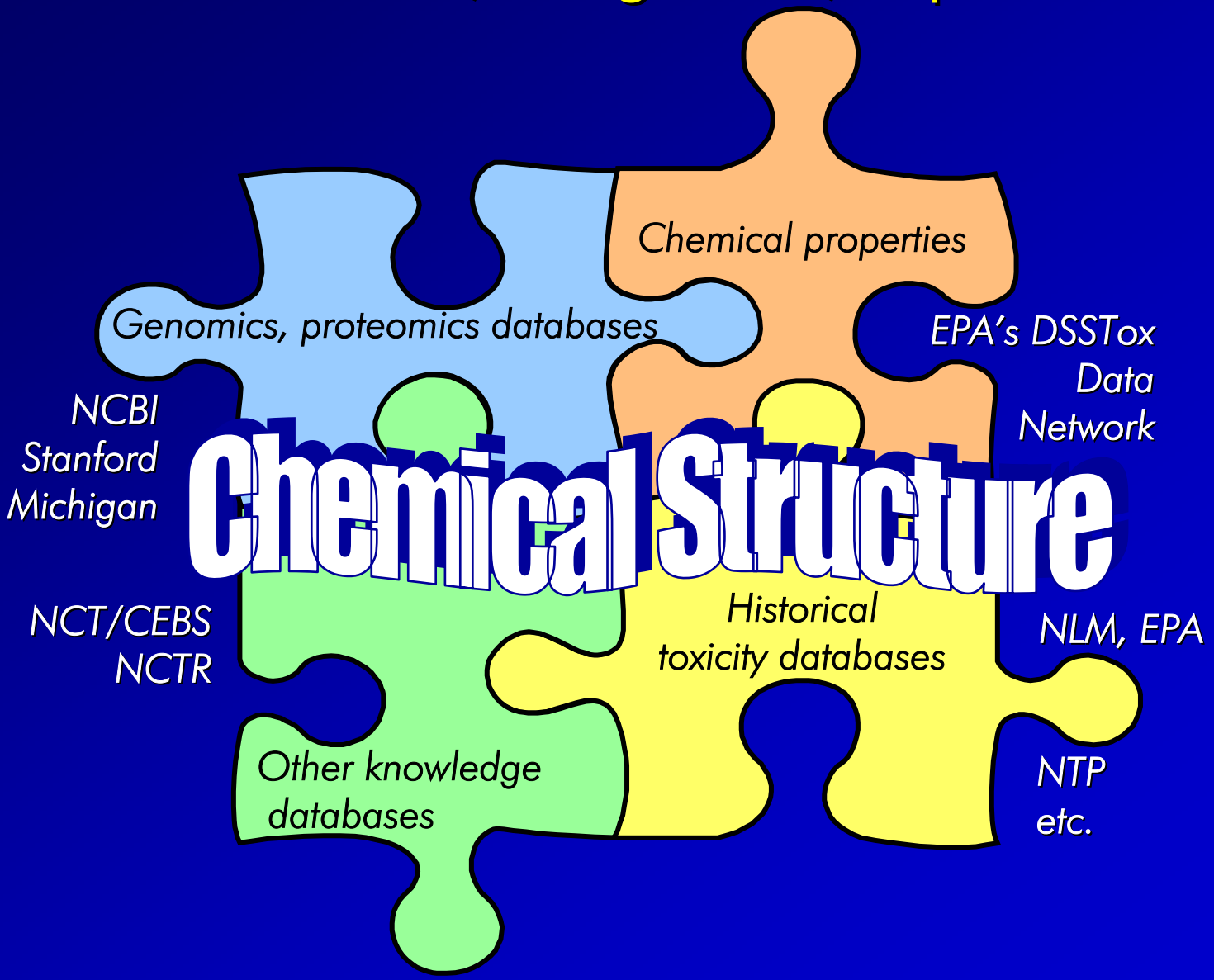




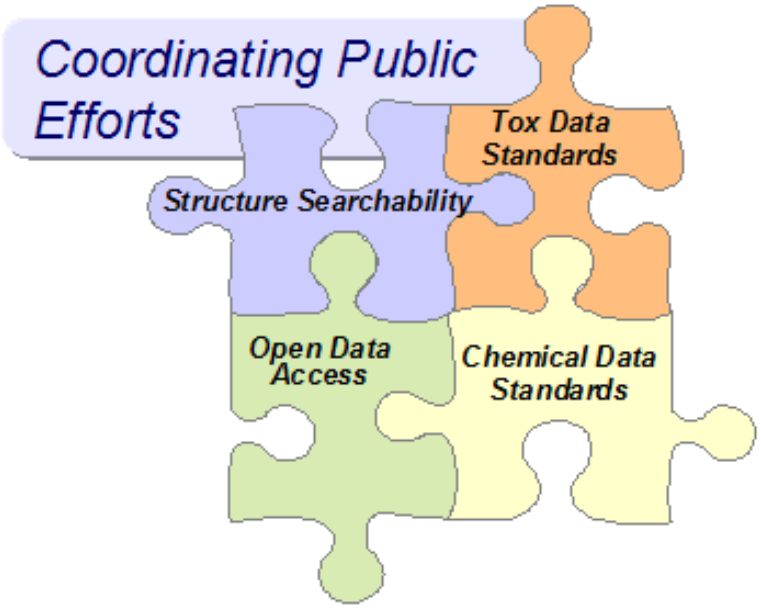
Informatics Exploration Axes:

- *Chemical Structure*
- *Gene expression*
- *Bioassay*

Toxico-Chemoinformatics:
Data Standardization, Integration, Exploration



COMMUNITY



- ACD/Labs (Advanced Chemistry Development) ChemFolder Public Databases
- Cambridge-Soft's ChemFinder.Com Chemical Search Website
- FDA (Food & Drug Administration) Center for Drug Evaluation & Research
- ILSI (International Life Sciences Institute) SAR Toxicity Database Project, in collaboration with LHASA, Lmt.LIST (LeadScope In Silico Tox) Focus Group
- LIST (LeadScope In Silico Tox) Focus Group
- MGED: MIAMI-Tox
- NCI (National Cancer Institute) Public Data Outreach – Structure Web Browser
- NIEHS's National Center for Toxicogenomics
- NLM (National Library of Medicine) TOXNET
- NTP (National Toxicology Program) On-line Public Databases
- SRC (Syracuse Research Corporation) PBT-Profiler and Analog Search Tools

DSSTox Collaborators/Advisors/Acknowledgements

- Cancer Potency Data Base - rodent carcinogenicity…….. Lois Swirsky Gold, Thomas Slone
- EPA - Ecotoxicity (fathead minnow, Teratox)…… Chris Russom
- EPA/OPP, Ecotox – pesticides …… Brian Montague, Pauline Wagner
- EPA/OPPT,OW – DBP cancer assessment…… Yin-tak Woo, Mary Manibusan
- FDA/NCTR - Estrogen receptor binding data base…… Weida Tong, Hong Fang
- NTP Gene-tox data; IRIS …… Errol Zeiger, Zeiger Consulting
- GlaxoSK - GeneTox/NTP Salmonella database…… Neal Cariello, Vijay Gombar
- NIEHS/NTP Rodent carcinogenicity, etc …… Skip Eastin, Doug Bristol
- Developmental toxicity …… Vijay Gombar (GlaxoSK), Orest Macina
- ICVAM Toxicity databases …… Ray Tice, Marc Jackson, ILS
- Unilever Skin Sensitization Database…… Martin Barrett, Marlin Consulting
- ZEBET Acute Toxicity Database …… Julie Penzotti, Rational Discovery
- EPA IRIS Database …… Amy Mills, Todd Martin (EPA), Errol Zeiger
- Tulane/Xavier Univ – Endocrine Disruption …… Tom Wiese
- NCI – SDF tools, CACTVS structure browser…… Marc Nicklaus
- LeadScope – SDF/XML converter, FDA carcinogens …… Chihae Yang
- SDF Viewer application …… Thomas Harrocks, Intuitive Software Solutions
- ACD – ChemFolder application …… Antony Williams, M Hachey, G Shear
- CambridgeSoft – ChemFinder application …… Rich Talbot
- EPA scientific advisors …… Stephen Nesnow, Adam Swank
- EPA/CSC – web development, IT …… Brian Garges/ D Kanipe, D Marshall

DSSTox Development Team

- ClarLynda Williams – Coordinator, Lead Project Technician
  - *NCCU/EPA COOP Student Trainee 12/00-12/02; EPA 12/02-8/03*
- Jamie Burch **
  - *NCCU/EPA COOP Student Trainee 11/02-2/04*
- Audrey Evans
  - *ECO/EPA Summer Student Trainee, Summer '03*
- Todd Stewart
  - *EPA-UNC Student COOP, Spring/Summer 02*
- Nina Fields
  - *Shaw Univ. High School Minority Mentoring Program, Summer '02*
- James Beidler
  - *EPA Summer Student Employee, Warren-Wilson College, Summer '01*
- Daniel Ohuoba
  - *Shaw Univ. High School Minority Mentoring Program, Summer '01*
- Adam Swank
  - *ECD, EPA*

http://www.epa.gov/nheerl/dsstox
or
http://www.dsstox.net

LAUNCH DATE:
March 1, 2004