DSSTox Log File:

EPA Water Disinfection By-Products with Carcinogenicity Estimates (DBPCAN)

(last updated 15 February 2008)

Description: Information in this file documents creation, review, and update process for the DSSTox DBPCAN SDF file, provides summary information on database content, and lists currently unavailable CAS registry numbers 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 DBPCAN file content. The Log Table documents modifications and revisions to the database content or format in version updates. To obtain the most current version of this Log File and a record of any new modifications, or to report errors in this file, a user should consult the DSSTox DBPCAN database page: http://www.epa.gov/ncct/dsstox/sdf_dbpcan.html

QA and Development Notes for v1a:
Data included in DBPCAN underwent a series of quality review checks prior to publication of initial launch version. The original database was obtained from reference materials associated with the Main Citation and from communication with the Source and main authors of that study. We thank Yin-tak Woo for providing the most current data tables in electronic form. Chemical structures were initially obtained by conversion of SMILES codes provided in the original database using CambridgeSoft's ChemOffice 2002 ChemDraw (ver 7.0 for Windows) for MS Excel. Structures were further verified against drawn structures and names in Source-provided materials. The ChemFinder website (http://chemfinder.cambridgesoft.com/) was used for checking CAS-to-structures and for retrieving CAS numbers for analogs of the predicted chemical. CambridgeSoft ChemFinder (ver 7.0 for Windows) was also 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. Original SMILES codes provided by the Source were converted to ChemFinder-generated SMILES when the former could not be converted to a Structure and verified within the ChemFinder application.

Notes for v2a:
For version 2a, a variety of fields have been added. IUPAC systematic chemical names, **ChemName_IUPAC**, were computed by Marc Nicklaus (NCI) using the ACD Labs IUPAC Name-Generation software (ACD/NameBatch, version 8.05). **INChI** codes were automatically generated from the final DSSTox SDF using a pre-release version of the publicly available program, wINChI11b.exe, accessible from the NIST INChI developers (http://chemdata.nist.gov/IChI/INChIv11b.zip). AuxInfo strings, which can be used to reproduce the molfile structure, are typically generated along with the INChI codes. However, due to their length frequently exceeding the 255 character limit of some Chemical Relational Database applications and the non-unique nature of the AuxInfo text string, we include only the invariant INChI codes in the DSSTox data files

Notes for v3a,b:
Revised DSSTox Standard Chemical Fields are included (see http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html) along with updated InChI codes (version 1.0), recomputed IUPAC chemical names (ACDLabs ACD/Name, version 9.0), and many regenerated 2D structures with stereochemistry of steroidal compounds rendered in more standardized form. Additionally, an extensive quality review of all DSSTox chemical records was performed, resulting in numerous corrections and modifications to chemical structures and added information (CASRN, representative structures for mixtures, etc) throughout DSSTox data files. For more information on current review procedures, see http://www.epa.gov/ncct/dsstox/ChemicalInfQAProcedures.html

Notes for v4a:
DBPCAN_v4 has no new chemical records but has minor QA corrections, field entry revisions, field changes, added CASRN, etc. Changes to DSSTox Standard Chemical Fields include new ID fields: **DSSTox_RID**, **DSSTox_Generic_SID** and **DSSTox_FileID** (replacing **DSSTox_SID** and

DSSTox_ID_FileName) (see http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html). Also, entries in **TestSubstance_Description** field have been simplified. Contents of **TestSubstance_ChemicalName_Other**, which has been retired, have been moved to **Note_DBPCAN** field, which has replaced **ToxicityNote** field. Also, abbreviations in **ActivityConcernLevel_Carcinogenicity** field have been replaced with full text entries.

Notes for v4b:
DBPCAN_v4b includes a small number of structure changes/modifications and two new summary activity fields for use in PubChem and structure-activity relationship studies: **ActivityOutcome_DBPCAN** (entries of active, inactive, or inconclusive) and **ActivityScore_DBPCAN**. In addition, the new **STRUCTURE_InChIKey** field (25 character abbreviated InChI for use in structure-indexing applications) has been added to this version file and contemporaneous updates of other DSSTox files.

Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Name	Modifications from previous version	Additional Notes

12Sep03	DBPCAN_v1a_209_12Sep03	Initial launch publication; no previous published versions.	DBPCAN is considered a "static" historical database meaning that further expansion of the database to include additional data is unlikely. Future updates will correct reported errors provided by users or incorporate DSSTox format changes.
1Mar05	DBPCAN_v2a_209_1Mar05	New Fields: **INChI**, **ChemName_IUPAC**, **StudyType**, **Endpoint** Modified Field Names: **DBPNote** to **ToxNote**; **ChemName Synonym** to **ChemName_Other**; CAS for DSSTox_ID=78,79 corrected; CAS added for 8 structures, DSSTox_ID=16, 25, 52, 74, 111, 125, 195, 200; SMILES and structure corrected for DSSTox_ID=25, 75, 121,122 (latter 2 switched) CAS incorrect for DSSTox_ID=24, changed to NOCAS	Major format modification to include INChI, IUPAC names, and ToxML fields.
10Apr2006	DBPCAN_v3b_209_10Apr2006	Updated with new DSSTox Standard Chemical Fields and entries (*revised Aug 2005*). Updated InChI codes (version 1.0). Updated IUPAC chemical names (ACDLabs Name to Structure, version 8.0). Expanded "ddmmmyear" format for dates in DSSTox file names (e.g., 10Apr2006). Revised DBPCAN Source-Specific Fields: **ChemClass DBP** changed to **ChemClass_DBP** **ConcernLevel** changed to **ActivityConcernLevel_Carcinogenicity**; **Rationale** changed to **ActivityConcernLevel_Rationale**; **Rationale** changed to **ActivityConcernLevel _RationaleSource**; **AnalogChemName** to **Analog_ChemicalName**; **AnalogCAS** to **Analog_CASRN**; **AnalogSMILES** to **Analog_SMILES** **ToxNote** to **ToxicityNote** with expanded "NOD" entry.	Numerous structure modifications and changes in stereochemical rendering throughout DSSTox data files following major quality review. DBPCAN_v3a_209_22Oct2005: Note: earlier version of this file was provided to PubChem, with identical format to v3b but latter has undergone additional QA review and has a few minor structural corrections/modifications.
15Jun2007	DBPCAN_v4a_209_15Jun2007	Revised Standard Fields: **DSSTox_SID** has been replaced by two new ID fields **DSSTox_RID** and **DSSTox_Generic_SID**. **DSSTox_ID_FileName** has been replaced by new ID field: **DSSTox_FileID**. Entries in **TestSubstance_Description** field have been	

		simplified. Entries in **ChemicalNote** that pertained specifically to DBPCAN have been moved to Source-Specific field: **Note_DBPCAN** Entries in **ActivityConcernLevel_Carcinogenicity** field have been unabbreviated. **TestSubstance_ChemicalName_Other** field has been eliminated and entries moved to **Note_DBPCAN**	
15Feb2008	DBPCAN_v4b_209_15Feb2008	Structure correction: RID 21498 - 2-Bromobenzothiazole, CAS 2516-40-7 RID 21677 - 3,3,3-Trichloro-2-methyl-1-propene, CAS 4749-27-3 RID 21612, 1-Hydroxy-3-methyl-2-hexene, changed to "representative isomer in mixture" New Standard Field added: **STRUCTURE_InChIKey** Two new summary activity fields added in coordination with PubChem deposits: **ActivityOutcome_DBPCAN** **ActivityScore_DBPCAN**	All corrections or changes to structure information noted in **Note_DBPCAN** field, searchable by version (e.g., v4b).

Field and Data Counts in Older DSSTox SDF File Versions:

DSSTox SDF	Standard Chemical Fields	Standard Toxicity Fields	Source-specific fields	Chemical records total
DBPCAN_v1a	11	0	9	209*
DBPCAN_v2a	14	2	8	209*

*Three pairs of replicate 2D structures exist in the database (cis/trans isomer pairs).

DBPCAN SDF Content	Totals_v3b	Totals_v4a	Totals_v4b
# Records	209	209	209
DSSTox Standard Chemical Fields	18	18	19
DSSTox Standard Toxicity Fields	2	2	2
DBPCAN Source Fields	8	8	10
Total # Fields	28	28	31
Chemical Content	**Counts_v3b**	**Counts_v4a**	**Counts_v4b**
STRUCTURE_ChemicalType:			
defined organic	204	204	204
inorganic	5	5	5
organometallic	0	0	0
no structure	0	0	0
STRUCTURE_TestedForm_ DefinedOrganic:			
parent	204	204	204
complex	0	0	0
salt	0	0	0
salt complex	0	0	0
TestSubstance_Description:			
single chemical compound	208	208	207
defined mixture or formulation	1	* (NA)	* (NA)
undefined mixture	0	* (NA)	* (NA)
macromolecule	0	0	0
mixture or formulation	* (NA)	1	2

** (NA) = field entry not applicable for DSSTox file version indicated*

Wanted!! CASRN Information

The Unknown CASRN entries below are primarily an indication of the unstudied nature of many of the disinfection by-product entries in the DBPCAN database. However, if a user has new information pertaining to any Unknown CASRN in the below listing, please report this using a DSSTox Error Report Form that can be accessed from any DSSTox SDF Download Page, and be sure to indicate all relevant information (**DSSTox_RID**, **TestSubstance_ChemicalName**, nature of missing information, source of correct information, etc.). Thank you!

DSSTox_ RID	TestSubstance_ChemicalName	STRUCTURE_ SMILES	CASRN	Date of Request
21494	1,2-Bis(1-methylethenyl)-benzene	C1=C(C(C)=C)C(C(C)=C)=CC=C1	Unknown	15Jun2007
21504	Bromochloromethyl acetate	BrC(OC(C)=O)Cl	Unknown	15Jun2007
21505	1,1-Bromochloropropanone	CC(C(Cl)Br)=O	Unknown	15Jun2007
21510	1-Bromo-1,1-dichloropropanone	O=C(C)C(Cl)(Br)Cl	Unknown	15Jun2007
21512	3-Bromopropylchloromethyl ether	BrCCCOCCl	Unknown	15Jun2007
21521	tert-Butyl maleic acid	O=C(O)C=C(C(C)(C)(C))C(O)=O	Unknown	15Jun2007
21525	3-Chloro-4-(bromochloromethyl)-5-hydroxy-2(5H)-furanone	O=C1OC(O)C(C(Br)Cl)=C1Cl	Unknown	15Jun2007
21528	3-Chloro-2-butanol acetate	CC(C(OC(C)=O)C)Cl	Unknown	15Jun2007
21529	2-Chlorobutenedioic acid	OC(C=C(C(O)=O)Cl)=O	Unknown	15Jun2007
21531	Chlorodibromoacetaldehyde	BrC(Br)(C=O)Cl	Unknown	15Jun2007
21534	2-Chloro-3-(dichloromethyl)-butenedioic acid	C(O)(=O)/C(C(Cl)Cl)=C(/Cl)C(O)=O	Unknown	15Jun2007
21536	(E)-2-Chloro-3-(dichloromethyl)-4-oxobutenoic acid	OC(C(Cl)=C(C=O)C(Cl)Cl)=O	Unknown	15Jun2007
21539	1-Chloro-2-ethoxy-2-methoxyethane	ClCC(OCC)OC	Unknown	15Jun2007
21540	4-Chloro-3-keto-1-butanal	O=CCC(=O)CCl	Unknown	15Jun2007
21542	2-Chloro-3-methyl-cis-butenedioic acid	OC(C(C)=C(C(O)=O)Cl)=O	Unknown	15Jun2007
21548	1-Chloro-3,3,3-trichloro-1-propen-1-amine	ClC(Cl)(C=C(Cl)N)Cl	Unknown	15Jun2007
21564	2,3-Dichloro-3-bromopropanenitrile	ClC(C(Cl)C#N)Br	Unknown	15Jun2007
21566	3,4-Dichlorobutanenitrile	ClCC(CC#N)Cl	Unknown	15Jun2007
21573	4,5-Dichloro-2-pentanol	ClCC(Cl)CC(C)O	Unknown	15Jun2007
21574	2,2-Dichloro-3-pentanone	CC(Cl)(Cl)C(=O)CC	Unknown	15Jun2007

21579	Dihydro-4,5-dichloro-2(3H)-furanone	O1C(Cl)C(Cl)CC1(=O)	Unknown	15Jun2007
21588	1,2-Dioxopropanoic acid	OC(C(C=O)=O)=O	Unknown	15Jun2007
21591	4-Dodecyl-5-ethyl-2(5H)-furanone	O=C1C=C(CCCCCCCCCCCC)C(CC)O1	Unknown	15Jun2007
21593	2-Ethyl-3-methylmaleic acid	O=C(O)C(C)=C(CC)C(O)=O	Unknown	15Jun2007
21611	1-[4-(1-Hydroxy-1-methylethyl)phenyl]-ethanone	C1=CC(C(C)=O)=CC=C1(C(C)(O)C)	Unknown	15Jun2007
21612	1-Hydroxy-3-methyl-2-hexene	CCC/C(C)=C/CO	Unknown	15Jun2007
21613	5-Hydroxy-5-trichloromethyl-2-furanone	C1=CC(O)(C(Cl)(Cl)Cl)OC1(=O)	Unknown	15Jun2007
21624	2-Methyl-3,3-dichloro-2-propenyldichloromethyl ether	ClC(Cl)=C(COC(Cl)Cl)C	Unknown	15Jun2007
21670	1,1,1-Tribromo-2-bromo-2-chloroethane	BrC(Br)(Br)C(Cl)Br	Unknown	15Jun2007
21673	1,1,1-Trichloro-2-butanone	O=C(CC)C(Cl)(Cl)Cl	Unknown	15Jun2007
21674	cis-2,3,4-Trichloro-2-butenenitrile	N#C/C(Cl)=C(/Cl)CCl	Unknown	15Jun2007
21675	trans-2,3,4-Trichloro-2-butenenitrile	N#C/C(Cl)=C(/CCl)Cl	Unknown	15Jun2007
21678	5,5,5-Trichloro-4-oxopentanoic acid	O=C(C(Cl)(Cl)Cl)CCC(O)=O	Unknown	15Jun2007
21687	1,3,3-Trimethyl-7-oxabicyclo-[4.1.0]-heptane-2,5-dione	C1(C)(C)CC(=O)C2OC2(C)C1(=O)	Unknown	15Jun2007