

DSSTox Log File:

Research Chemical Inventory for EPA's ToxCast™ Program (TOXCST)

Structure-Index Locator File

(last updated 12 February 2009)

Description: Information in this file documents creation, review, and update process for the DSSTox TOXCST SDF file, and provides summary information on database content. 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 TOXCST 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 TOXCST database page: http://www.epa.gov/ncct/dsstox/sdf_toxcst.html. Future upgrades to this file will be made in full coordination with [EPA's National Center for Computational Toxicology \(NCCT\) ToxCast™ Project](#) and public announcements.

QA and Development Notes for v1:

The TOXCST SDF file underwent an extensive series of quality review checks prior to publication of initial launch version. Launch version (v1a) corresponds to the initial ToxCast Phase I candidate list of chemical names and CASRN published on the [EPA NCCT ToxCast website](#). Starting information on chemical structure, names, CASRN were obtained from EPA Office of Pesticide Programs (OPP) resources. These were checked against multiple public sources, internal sources (Certificates of Analysis, Material Safety Data Sheets), and in particular instances the STN CAS commercial database (see also Chemical Information Quality Review Procedures). The initial TOXCST CASRN and chemical name inventory were imported into Excel and cross-referenced to the existing DSSTox Master File chemical inventory, all issues pertaining to mismatch of CAS and chemical name were resolved, in consultation with EPA Source Collaborators, and DSSTox Standard Chemical Fields (including structures) were assigned when matches to existing DSSTox Master File inventory were found. For those chemical substances not in the existing DSSTox Master File inventory, we employed a number of commercial and public Internet resources for converting name to structure ((ACD/Name, version 9) or assigning chemical structures based on CAS, and populating DSSTox Standard Chemical Fields (for details of general QA review procedures, see <http://www.epa.gov/ncct/dsstox/ChemicalInfQAProcedures.html>). IUPAC systematic chemical names, **STRUCTURE_ChemicalName_IUPAC**, were computed using the ACD/Labs IUPAC Name-Generation software (ACD/Name, version 10); SMILES were generated with ACD/Labs ChemFolder (version 10). **InChI** codes were automatically generated from the final DSSTox SDF using the publicly available program, wINChI1.exe, downloadable from the NIST InChI website (<http://www.iupac.org/inchi/>).

Notes for v2:

The original v1a published list of ToxCast Phase I candidate chemicals was submitted to EPA ToxCast program contractors for chemical procurement and sample preparation for HTS. The v2a file represents the final inventory with changes due to procurement issues (compound unavailability, duplication, or substitution). The contents of the TOXCST_v2a file represents the actual list of purchased chemicals undergoing sample preparation and plating for generating ToxCast program assay results.

Version 2b Revision: Additional chemical QA was performed after obtaining full Certificates of Analysis (COAs) for purchased chemicals. Relatively minor modifications (stereochemistry, salt/complex form) were made to 8 structure records, with version changes documented in the corresponding record in the **Note_TOXCST** field. The field, **EPA_PC_Code** (EPA Pesticide Chemical Codes) has been removed from v2b file due to need for further internal QA and some inconsistencies in usage. Additionally, a new field, **STRUCTURE_InChIKey**, was added. This last field will soon be incorporated into all DSSTox files as a DSSTox Standard Chemical Field.

Version 2c Revision: Correction was made to **DSSTox_FileID** = 20_TOXCST_v2b record: several fields other than **STRUCTURE** incorrectly referred to Na salt form and were changed to correspond to parent form.

Version 3a Update: Newly added Source-specific fields include **Source_ChemicalName**, **Chemical ReplicateCount**, and **Relationship_CID**. The latter two fields are intended to aid identification of replicates within the data set and to map important relationships between substances, such as parent0-metabolite pairs. Minor corrections to structure-annotation fields resulting from further QC based on review and Certificates of Analysis (COAs), with changes affecting 6 records and documented in the **Note_TOXCST** field. **STRUCTURE_InChI** and **STRUCTURE_InChIKey** field entries have been updated to conform to newly released InChI 1.02 Standards (<http://www.iupac.org/inchi/release102final.html>).

Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Name	Modifications from previous version	Additional Notes
------	----------------------	-------------------------------------	------------------

03Aug2007	TOXCST_v1a_340_03Aug2007	Initial launch publication; no previous published versions.	TOXCST is considered a "live" data file, meaning that further expansion of the data file to include additional data is likely. Future updates also will correct reported errors provided by users or incorporate DSSTox format changes.
25Sep2007	TOXCST_v2a_320_25Sep2007	33 substance records removed from v1a to form v2a (see table below); 13 new substance records added to v2a, all representing replicates of substances already included in v1a: 7 sets of duplicates, 3 sets of triplicates (see table below). CASRN and/or Chemical Name modifications for: RID 40403, d-cis,trans-Allethrin [584-79-2] CASRN changed RID 40463, (Z,E)-Fenpyroximate [111812-58-9] CASRN and ChemicalName changed	Fields whose entries exceeded 253 characters (such as some SMILES and InChI) were found to have been truncated in v1a; these problems have been corrected in v2a. Changes from v1a to v2a are documented in Note_TOXCST field.
08Feb2008	TOXCST_v2b_320_08Feb2008	Substance records modified from v2a to v2b: RID 40372 cis-captafol, structure changed to non-stereo specific chemical (COA) RID 40405 Cyhalofop-butyl, structure changed from acid to butyl ester RID 40361 Diethylhexyl phthalate (DEHP), chemical tested was mono- ester not di- ester (COA) RID 40434 Diniconazole-M, remove stereochem (COA) RID 40573 Fluazifop-P-butyl, remove stereochem (COA) RID 40515 Metam-sodium, change from anhydrous to hydrate RID 40539 Niclosamide-olamine, change from thanolamine to parent RID 40317 Oxytetracycline, change to dehydrate Delete EPA_PC_Code field. Add STRUCTURE_InChIKey field.	Changes from v2a to v2b are documented in Note_TOXCST field.
29Apr2008	TOXCST_v2c_320_29Apr2008	Substance records modified from v2b to v2c: RID 40374 Asulam, structure unchanged, but several fields referring to Na salt form corrected to refer to parent form	Changes from v2b to v2c are documented in Note_TOXCST field.
12Feb2009	TOXCST_v3a_320_12Feb2009	Substance records modified from v2c to v3a: RID 40389 Cinmethylin, stereochem added RID 40372 Captafol, CofA indicated no stereochem so CASRN changed to [2425-06-1] RID 40403 Allethrin, stereochem removed v3a RID 40485 Formetanate hydrochloride, HCl added RID 40545 Oxamyl, SMILES and IUPAC name corrected RID 40550 Penoxsulam, SMILES and IUPAC name corrected	Changes from v2c to v3a are documented in Note_TOXCST field.

--	--	--	--

TOXCST SDF Content Summary:

NTPBSI SDF Content	Totals_v1a	Totals_v2a	Totals_v2b	Totals_v2c*	Totals_v3a*
# Records	340	320	320	320	320
DSSTox Standard Chemical Fields	18	18	19	19	20
NTPBSI Source Fields	3	3	2	2	3
Total # Fields	21	21	21	21	21
Chemical Content	Counts_v1a	Counts_v2a	Counts_v2b	Counts_v2c	Counts_v3a
STRUCTURE_ChemicalType:					
defined organic	331	313	313	313	313
inorganic	1	1	1	1	1
organometallic	8	6	6	6	6
no structure	0	0	0	0	0
STRUCTURE_TestForm_DefinedOrganic:					
parent	309	298	299	300	299
complex	7	9	8	8	6
salt	15	6	6	5	7
salt complex	0	0	0	0	1
TestSubstance_Description:					
single chemical compound	327	308	308	308	309
macromolecule	1	0	0	0	0
unspecified or multiple forms	0	0	0	0	0
mixture or formulation	12	12	12	12	11

* contains 309 unique substances (distinct DSSTox_SIDs and CIDs), 5 sets of duplicate compounds, and 3 sets of triplicate compounds.

Duplicate records in v2:

DSSTox_RID	DSSTox_CID	DSSTox_Generic_SID	TestSubstance_ChemicalName	TestSubstance_CASRN
40375	8038	28038	3-Iodo-2-propynylbutylcarbamate	55406-53-6
40760	8038	28038	3-Iodo-2-propynylbutylcarbamate	55406-53-6
40376	4091	24091	EPTC	759-94-4
40761	4091	24091	EPTC	759-94-4
40388	3980	23980	Chlorsulfuron	64902-72-3
40762	3980	23980	Chlorsulfuron	64902-72-3
40419	1781	21781	Dibutyl phthalate	84-74-2
40763	1781	21781	Dibutyl phthalate	84-74-2
40461	12392	32392	Fenoxaprop-ethyl	66441-23-4
40766	12392	32392	Fenoxaprop-ethyl	66441-23-4

Triplicate records in v2:

DSSTox_RID	DSSTox_CID	DSSTox_Generic_SID	TestSubstance_ChemicalName	TestSubstance_CASRN
40353	12329	32329	Bensulide	741-58-2
40757	12329	32329	Bensulide	741-58-2
40758	12329	32329	Bensulide	741-58-2
40424	12605	32605	Diclofop-methyl	51338-27-3
40764	12605	32605	Diclofop-methyl	51338-27-3
40765	12605	32605	Diclofop-methyl	51338-27-3
40579	14868	34868	Prosulfuron	94125-34-5
40768	14868	34868	Prosulfuron	94125-34-5
40769	14868	34868	Prosulfuron	94125-34-5

Records from v1a deleted from v2a:

DSSTox_RID	DSSTox_CID	DSSTox_Generic_SID	TestSubstance_ChemicalName	TestSubstance_CASRN
40296	14205	34205	Azafenidin	68049-83-2
40302	12531	32531	Bromuconazole	116255-48-2
40309	12591	32591	Hydantoin, 1-bromo-3-chloro-5,5-dimethyl-	16079-88-2
40316	14259	34259	2-Methyl-4-isothiazolin-3-one	2682-20-4
40321	19789	39789	3',5'-Dichloro-2-hydroxy-2-methylbut-3-enanilide	83792-61-4
40324	14271	34271	Imazapic-ammonium	104098-49-9
40325	12375	32375	Diflufenzopyr-sodium	109293-98-3
40329	14286	34286	5-Chloro-2-methyl-4-isothiazolin-3-one	26172-55-4
40336	14307	34307	Triazamate	112143-82-5
40355	14356	34356	Flucycloxuron	113036-88-7
40364	194	20194	Boric acid	10043-35-3
40381	12533	32533	Chlorfenapyr	122453-73-0
40383	315	20315	Chloropicrin	76-06-2
40391	14460	34460	Clofencet-potassium	82697-71-0
40395	346	20346	Oxine-copper	10380-28-6
40406	12357	32357	Cyhexatin	13121-70-5
40412	4000	24000	Chlorthal-dimethyl	1861-32-1
40416	14518	34518	Desmedipham	13684-56-5
40435	14549	34549	Dinotefuran	165252-70-0
40467	14612	34612	Fluazifop-butyl	69806-50-4
40469	14614	34614	Flucarbazone-sodium	181274-17-9
40514	14715	34715	Metaldehyde	108-62-3
40518	14722	34722	Topramezone	210631-68-8
40525	4243	24243	Paraquat dichloride	1910-42-5
40532	5680	25680	Mono-(2-ethylhexyl)phthalate	4376-20-9
40563	14829	34829	Polixetonium chloride	31512-74-0
40570	4274	24274	Propachlor	1918-16-7
40594	14920	34920	Pyrithione sodium	15922-78-8
40595	12478	32478	Spinosad	168316-95-8
40599	14940	34940	Sulfosulfuron	141776-32-1
40628	12654	32654	Triforine	26644-46-2
40630	15002	35002	Uniconazole-P	83657-17-4
40631	15004	35004	Metobenzuron	111578-32-6