

DSSTox Log File:

Carcinogenic Potency Database Summary Tables – All Species (CPDBAS)

(last updated 20 November 2008)

Description: Information in this file documents the creation, review, and update process for the DSSTox CPDBAS SDF file, provides summary information on database contents, 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 CPDBAS file contents and chemical composition. A second table provides summary counts of various types of replicate chemical information in the CPDBAS file. The Log table documents any 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 CPDBAS database page: http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html. Each version of the DSSTox CPDB and CPDBAS data files is constructed using the most current content of the published [CPDB Summary Tables](#) posted on the [CPDB Source Website](#).

QA and Development Notes for v1a:

CPDB SDF files were originally constructed by conversion and significant editing of the CPDB Summary Table content, and addition of chemical structures by cross referencing to CAS and Chemical Names. The original SDF 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 published 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, 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_v1a (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. For versions 2 and later, a DOP is not created; rather the **SMILES_Parent** field is included and can be used to create a "desalted" version of the Main file by the user. All CAS registry numbers in the CPDBAS_v2a file were checked by the CAS check-digit verification algorithm (<http://www.cas.org/EO/checkdig.html>) using a Python script (CASlistcheckv2.py) created by Stephen Little (EPA).

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). Where IUPAC names were not provided, systematic names were either inferred from the structure or obtained from the TOXNET ChemID website (<http://chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp>). **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.

In addition to the incorporation of two new DSSTox Standard Chemical Fields (**ChemName_IUPAC** and **INChI**), three new DSSTox Standard Toxicity Fields have been added to CPDBAS: **StudyType**=carcinogenicity; **Species**=rat, mouse, hamster, dog, rhesus, cynomolgus, bush baby, tree shrew; **Endpoint**=TD50, Tumor Target Sites. In the case of CPDBAS, the **Species** field enables listing of all species for which data are available for a given chemical. A number of additional modifications were incorporated into the CPDBAS_v2a to facilitate the use of these data in structure-activity studies and relational database searching. The 4 separate data files published in the original DSSTox CPDB v1a (CPDBRM, CPDBDG, CPDBHA, and CPDBPR) have been consolidated into a single file (CPDBAS = CPDB All Species) containing a total of 1451 records. Data in this file were extracted from the CPDB Summary Table files posted on the Source Website as of 15Nov04: 1433 records from the updated RatMouse Summary Table and an additional 18 unique chemical substances from the other 3 species tables that were not included already in the RatMouse table. A total of 9 records that were included in CPDB v1a (all starches and pectins) no longer appear in the current CPDB Summary Tables and so have been deleted from CPDBAS_v2a, whereas a total of 88 new records from the CPDB RatMouse Summary Table and 3 new records from the CPDB Hamster Summary Table have been incorporated into CPDBAS_v2a. CPDBAS_v2a also incorporates a number of data modifications in the TD50 and Target Sites fields for several chemical records that were included in v1a, these modifications obtained from the most current CPDB Summary Table files posted on the Source Website as of 15Nov04. Wherever new data were added or modifications were made to existing toxicity data from CPDB v1a, a notation is included in a new field, ToxNote. New data fields were created to accommodate the data from the four original data tables. The only exceptions are the TD50 and Target Sites fields for primate species "bush babies" and "tree shrews", which are not included as separate fields in v2a since they each have only a single data entry in 1451 records. Instead, these data are indicated in the **Species** field and listed in the ToxNote field entry for the corresponding chemical. CPDBAS_v2a also includes 3 new, purely numeric fields for the largest TD50 data columns, i.e. **TD50 Rat notext**, **TD50 Mouse notext**, **TD50 Hamster notext**, alongside the original field, the latter including text notes and the lettered footnote references from the original CPDB Summary Tables. The pure numeric fields are intended to facilitate exploration of numerical trends of the TD50 as a function of chemical structure. Finally, we have discontinued offering the DOP (defined organic parent) file containing simplified-to-parent structures for the defined organics. Instead, we offer a number of standard chemical fields, including **SMILES_Parent**, which should enable a user to create easily their own "desalted" parent file for specialized purposes.

Finally, in version 2a, a number of corrections and modifications of Standard Chemical Fields, including structures, have been made. These include the addition of 2D stereochemistry indicators in many records that were found to overlap with other DSSTox databases containing stereochemistry, in particular, NCTRER and FDAMDD. Due to the sheer number of modifications made to CPDBAS_v2a versus the earlier 1a versions, we do not list each of these corrections separately; although v2a record DSSTox_IDs are listed where there was an error in the v1a structure. It is recommended that users replace the earlier 1a versions with this new v2a in its entirety.

Notes for v3a,b:

CPDBAS_v3 has 31 new chemical records added from v2a and data were added to several existing records from v2a. Revised DSSTox Standard Chemical Fields are included (see <http://www.epa.gov/nct/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. With the modifications to the DSSTox Standard Chemical Fields, to better distinguish **STRUCTURE** fields from **TestedForm...** fields, many more structures are included in CPDBAS for substances classified as mixtures, e.g. **STRUCTURE_Shown =** "active ingredient of mixtures", "monomer of polymer", "representative isomer of mixture", "representative component of mixture". CPDBAS_v3b additionally includes a field containing URL links (**Website_url**) to newly posted chemical-specific data pages listed on the [CPDB Source Website Chemical Index](#) page for the full CPDBAS chemical inventory.

Some CPDBAS Source Toxicity Fields were deleted and several new fields were added (see below). In addition, many field names were changed to eliminate spaces and provide more descriptive names. Finally, 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/nct/dsstox/ChemicalInfQAProcedures.html>

Notes for v4a:

CPDBAS_v4 has no new chemical records but has several minor QA corrections, field entry revisions, field changes, new 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>). The **ChemicalReplicateCount** field has been eliminated since this information can be easily extracted from other Standard Chemical Fields, and entries in the **TestSubstance_Description** field have been simplified. In all **TumorSites_...** fields, previously abbreviated tumor sites and footnotes have been expanded. All **TD50_..._mg** fields have been stripped of text notes and converted to pure numeric entries (notes have been moved and expanded in **TD50_..._Note** fields). The field **ActivityCategory_SingleCellCall** has been modified to exclude inconclusive results (blank entry), and the field **ActivityCategory_MultiCellCall** has been converted to numeric integer form, with entries pertaining to both active and inactive multicellcalls, and details provided in new field, **ActivityCategory_MultiCellCall_Details**.

Notes for v5a:

CPDBAS_v5 represents a major update of the CPDBAS data file, with 66 new chemical records added and over 400 new or modified experimental results affecting nearly 100 existing data records. These changes correspond to newly published data in the [CPDB Summary Tables](#) posted on the [CPDB Source website](#). Numerous manual checks in content and entered values in the CPDBAS_v5a file were undertaken to ensure accuracy of transcription of experimental results. [DSSTox Standard Chemical Fields](#) for the 66 new chemical records were either extracted from (if previously existing), or added to (if not previously existing) the most recent [DSSTox Master File](#) according to current chemical information review procedures <http://www.epa.gov/ncct/dsstox/ChemicalInfQAProcedures.html>. In addition to 66 added records and hundreds of new or modified experimental results, additional quality review and a user error-report led to correction of 3 previously included chemical structures (indicated below; reported by Emilio Benfenati) and a small number of previous experimental results with errors in transcription.

The significant number of new chemical records added to v5a presents a unique opportunity to structure-activity modelers to use these new data for validation of existing carcinogenicity models. From the SDF Content Summary Table below, it can be inferred that of the 66 new records, 59 are defined organics, 53 correspond to the parent structure (i.e., neither salt nor complex), and 51 are a "single chemical compound" as opposed to a macromolecule or mixture. To assist users in locating modified or newly added records from one DSSTox CPDBAS file version to the next, we have incorporated controlled text entries (separated by semicolons) into the Note_CPDBAS field where applicable, e.g. for this v5a update:

- chemical added v5a(66 instances)
- Rat added v5a(63 instances)
- Mouse added v5a(42 instances)
- TD50_Rat_Note modified v5a(19 instances)
- Mutagenicity_SAL_CPDB added v5a (33 instances)
- TD50_Rat modified v5a(13 instances)
- TargetSites_Mouse_Female modified v5a(8 instances)
- ... etc

Also of interest to structure-activity modelers should be the various summary activity representations provided in CPDBAS_v5a. A table indicating total numbers of records in the various activity categories is provided below for CPDBAS_v5a. In particular, note that when the **ActivityCategory_MultiCellCall** has the value "1" (i.e., active), the largest number of incidences occurs for the **ActivityCategory_MultiCellCall_Details** entry "multisite; multisex; multispecies" (193 incidences). Similarly, when the **ActivityCategory_MultiCellCall** has the value "0" (i.e., inactive), the largest number of incidences occurs for the **ActivityCategory_MultiCellCall_Details** entry "multisite; multispecies" (288 incidences). These two categories represent a more conservative, heavily weighted carcinogenic activity than the **ActivityCategory_SingleCellCall** (806 actives vs. 738 inactives) that is more frequently employed to build structure-activity prediction models.

	ActivityCategory_		ActivityCategory_MultiCellCall_Details**			
Call	SingleCellCall	MultiCellCall**	multisite***	multisex	multispecies	Total Incidences* (CPDBAS_v5)
Active (1)	active					223
	active	active	✓			81
	active	active		✓		113
	active	active			✓	8
	active	active	✓	✓		123
	active	active	✓		✓	27
	active	active		✓	✓	37
	active	active	✓	✓	✓	193
	active	active	Total MultiCellCall Incidences (active/active)			582
Inactive (0)	inactive		✓			169
	inactive	inactive	✓	✓		266
	inactive	inactive	✓		✓	16
	inactive	inactive	✓	✓	✓	288
	inactive	inactive	Total MultiCellCall Incidences (inactive/inactive)			570

* SingleCell and MultiCell Calls total 1544, or 3 short of the CPDBAS_v5a inventory of 1547 total records; for 3 records in CPDBAS_v5, calls were not provided due to "NTP bioassay inadequate".

** A blank cell in the above table indicates a negative condition, i.e. condition not met. For example, in the third row of results, the multisex condition applies but NOT multisite or multispecies, i.e., a single tumor site is observed exclusively in both sexes of the same species.

*** For any Sex/Species (e.g., Male/Rat) experiment in CPDB that reports an Inactive (negative) call, multisite inactive is implied.

Notes for v5b:

CPDBAS_v5b includes no new carcinogenicity experimental data, but includes several new and modified summary activity fields for use in PubChem and structure-activity relationship studies. These include **ActivityOutcome_CPDBAS_**,,, fields (entries of active, inactive, or inconclusive) for Mutagenicity, SingleCellCall, MultiCellCall, Rat, Mouse, Hamster, and Dog_Primates, and **ActivityScore_CPDBAS_**... fields [Log(1/TD50mmol) mapped onto 0-100 integer range] for Rat, Mouse, Hamster, and Dog_Primates. Since two species listed in the Source CPDB Summary Tables (Bush Baby and Tree Shrew) have data for only a single chemical record each, and the chemical record in each case is also listed in the CPDB Rat Mouse Summary Table, data for these species are provided in the new **TD50_Dog_Primates_Note** field of the corresponding chemical record. 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.

Notes for v5c:

CPDBAS_v5c includes 2 structure modifications and one correction to a record that incorrectly listed an inactive bioassay result for Mouse. See table below for details.

Notes for v5d:

CPDBAS_v5d includes modifications for 16 "inconclusive" entries in ActivityOutcome_CPDBAS_(Rat, Mouse, Hamster, SingleCellCall) fields based on further discussion with the Source Contact (L.Gold). One of the 16 (**ActivityOutcome_CPDBAS_Hamster**) was changed to "active" (corresponding **ActivityScore** unchanged). For the remaining 15 cases (**ActivityOutcome_CPDBAS_Rat** - 7 cases, **ActivityOutcome_CPDBAS_Mouse** - 5 cases, **ActivityOutcome_CPDBAS_SingleCellCall** - 3 cases), the "inconclusive" entry was changed to "unspecified". These changes have been incorporated in the corresponding PubChem CPDBAS bioassay entries. The corresponding ActivityScore of "0" remains unchanged in all 15 cases. Consult CPDBAS Log file for further details.

Category Field Name	ActivityOutcome_CPDBAS_Species (v5d)				ActivityScore_CPDBAS_Species*
	Active	Inactive	Unspecified	Total # Entries	Total # Entries*
Mutagenicity (Salmonella)	403	457	0	860	NA
Rat	587	646	7	1240	1240
Mouse	445	558	5	1008	1008
Hamster	45	42	0	87	87
Dog_Primates	15	17	0	32	NA
SingleCellCall	806	738	3	1547	NA
MultiCellCall	582	570	0	1152	NA

CPDBAS Data Content Summary - 20 November 2008

CPDBAS_v5d contains 1547 total chemical substance records, but not all substances have been tested in all species categories. The table below provides a summary of the total numbers of substances tested for each of the species categories and the corresponding counts within the **ActivityOutcome** and **ActivityScore** fields derived from the CPDB Summary Tables. For definitions of these fields and field entries, refer to the field listing in the previous section or consult the CPDBAS_FieldDefFile available in the [Download Table](#) below.

**** ActivityScore_CPDBAS_Species** fields are only included for CPDB species categories with at least one positive experiment in the CPDB, and therefore a TD50 value is reported in the CPDB Summary Table. However, this field is not included for the field **ActivityOutcome_Dog_Primates** (which groups results for species Dog, Rhesus, and Cynomolgus) due to the combined species nature of this field and the small number of chemicals with available data for each of the included species.

Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Name	Modifications from previous version	Additional Notes
15Oct03 15Oct03 15Oct03 15Oct03 15Oct03	CPDBRM_v1a_1354_15Oct03.sdf CPDBRM_DOP_v1a_1189_15Oct03.sdf CPDBHA_v1a_80_15Oct03.sdf CPDBDG_v1a_5_15Oct03.sdf CPDBPR_v1a_27_15Oct03.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.
29Mar04	CPDBPR_v1b_27_15Oct03.sdf	Corrected structure DSSTox_ID=2, 2,7-Acetylaminofluorene	Thanks to ACD Labs
1Mar05	CPDBAS_v2a_1451_1Mar05.sdf	Consolidation of data from 4 previous v1 files (CPDBRM, CPDBHA, CPDBDG, CPDBPR) into single CPDBAS_v2 file; Addition of 3 pure numeric fields, TD50 Rat notext , TD50 Mouse notext , TD50 Hamster notext , for specialized use. Replacement of field entry "ND" (no data) with blank entries in data fields. Addition of 88 new chemical records for Rat or Mouse and 3 new chemical records for Hamster. Also, modification of data fields for many previous v1 chemical records. Modifications extracted from the current CPDB Source Website Summary Tables (15Nov04). Addition of SMILES_Parent to Main file. New Standard Chemical Fields: INChI, ChemName_IUPAC New Standard Toxicity Fields: StudyType, Species, Endpoint	Major format modification to include INChI, IUPAC names, and ToxML fields. Separate "desalted" defined organic parent (DOP) file not provided. Users can easily generate DOP file by extracting "defined organic" records and converting SMILES_Parent to structures. Since both tree shrew and bush baby (Non-Human Primates) each had only a single record for which data were available, and these records already contained data for other species, separate fields for these species were not included; rather these data are noted by the entries "bush baby" or "tree shrew" in the Species field and TD50 and Target Sites are listed in the ToxNote field of the corresponding records.

		<p>Modified Field Names: Target Sites Rat Both Target Sites Mouse Both, Target Sites Hamster Both, to Target Sites Rat Both Sexes, Target Sites Mouse Both Sexes, Target Sites Hamster Both Sexes.</p> <p>Deleted Fields: TD50 Tree Shrews and Target Sites Tree Shrews, TD50 Bush Babies and Target Sites Bush Babies OtherSpecies</p> <p>Additional Toxicity Field: ToxNote</p> <p>Corrected errors in structures for CPDBAS_v2a:DSSTox_ID = 82, 120, 565, 603, 605, 663, 670, 798, 821, 880, 1050, 1144</p>	
10Apr2006	CPDBAS_v3b_1481_10Apr2006.sdf	<p>Updated with new DSSTox Standard Chemical Fields and entries (<i>revised Aug 2005</i>).</p> <p>Updated InChI codes (version 1.0).</p> <p>Updated IUPAC chemical names (ACDLabs Name to Structure, version 8.0).</p> <p>Expanded "ddmmmyear" format for dates in DSSTox file names (e.g., 10Apr2006).</p> <p>Addition of 30 new chemical records. Also, modification of data fields for many previous v2 chemical records. Modifications extracted from the current CPDB Source Website Summary Tables (01Jun2005).</p> <p>Addition of 1481 chemical-specific data page URLs newly posted on the CPDB website in Website_url field.</p> <p>Deleted Source-related fields: TD50_Rat_notext TD50_Mouse_notext TD50_Hamster_notext</p> <p>New Source-related fields: TD50_Rat_mmol TD50_Mouse_mmol TD50_Hamster_mmol ActivityCategory_SingleCellCall ActivityCategory_MultiCellCall NTP_TechnicalReport</p>	<p>Numerous structure modifications and changes in stereochemical rendering throughout DSSTox data files following major quality review.</p> <p>CPDBAS_v3a_1484_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 small number of corrections/modifications. Also, 3 duplicated records were reconciled and eliminated in subsequent version.</p>

		<p>Website_url</p> <p>Renamed Source-related fields to be more descriptive and eliminate spaces:</p> <p>Mutagenicity_SAL_CPDB TD50_Rat_mg TargetSites_Rat_Male TargetSites_Rat_Female TargetSites_Rat_BothSexes TD50_Mouse_mg TargetSites_Mouse_Male TargetSites_Mouse_Female TargetSites_Mouse_BothSexes TD50_Hamster_mg TargetSites_Hamster_Male TargetSites_Hamster_Female TargetSite_Hamster_BothSexes TD50_Dog_mg TargetSites_Dog TD50_Rhesus_mg TargetSites_Rhesus TD50_Cynomolgus_mg TargetSites_Cynomolgus ToxicityNote</p>	
15Jun2007	CPDBAS_v4a_1481_15Jun2007.sdf	<p>Revised Standard Fields:</p> <p>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. ChemicalReplicateCount field has been eliminated. Entries in TestSubstance_Description field have been simplified. Entries in ChemicalNote that pertained specifically to CPDBAS have been moved to Source-Specific field: Note_CPDBAS</p> <p>The following Source-Specific Fields were stripped of text notes and converted to pure numeric entries</p> <p>TD50_Rat_mg TD50_Mouse_mg TD50_Hamster_mg TD50_Dog_mg TD50_Rhesus_mg TD50_Cynomolgus_mg</p>	

		<p>Entries in the following Source-Specific Fields were converted from abbreviations to full text:</p> <p>TumorSites_Rat_Male TumorSites_Rat_Female TumorSites_... all</p> <p>New Source-Specific Fields:</p> <p>TD50_Rat_Note TD50_Mouse_Note TD50_Hamster_Note TD50_Dog_Rhesus_Cynomolgus_Note Inactivity_SingleCellCall Inactive_MultiCellCall</p> <p>Modified and new Source-Specific Fields:</p> <p>ActivityCategory_SingleCellCall ActivityCategory_MultiCellCall ActivityCategory_MultiCellCall_Details Website_URL Note_CPDBAS (formerly ToxicityNote)</p>	
25Oct2007	CPDBAS_v5a_1547_25Oct2007.sdf	<p>Update incorporates all additions and modifications in newly published content of CPDB Source Website Summary Tables- All Species, as of Sept 2007.</p> <p>Entries added to Note_CPDBAS field to allow user to easily locate new and updated records:</p> <p>chemical added v5a(66 instances) Rat added v5a(63 instances) Mouse added v5a(42 instances) Hamster added v5a TD50_Rat_Note modified v5a(19 instances) Mutagenicity_SAL_CPDB added v5a (33 instances) TD50_Rat modified v5a(13 instances) TargetSites_Mouse_Female modified v5a(8 instances) ... etc</p> <p>Corrections to three v4a structures:</p> <ol style="list-style-type: none"> 1. RID=20256, CID=256, CAS[101-25-7] N,N-Dinitrosopentamethylenetetramine 2. RID=21047, CID=1047, CAS[75411-83-5]N-Nitrosomethyl-2-hydroxypropylamine 3. RID=21146, CID=1146, CAS[133920-06-6]6- 	

		<p>Phenylhexyl isothiocyanate</p> <p>Website_URL field name changed to ChemicalPage_URL to better convey chemical-specific content.</p> <p>Changes to some CPDB Source Chemical Names were incorporated into v5a. In addition, some obviously incorrect Source Chemical Names were corrected.</p>	
<p>10 February 2008</p>	<p>CPDBAS_v5b_1547_10Feb2008.sdf</p>	<p>Update incorporates a number of modified and new summary activity fields for use in PubChem and structure-activity relationship studies.</p> <p>Modified fields:</p> <p>Mutagenicity_SAL_CPDBAS renamed ActivityOutcome_CPDBAS_Mutagenicity; entries changed from [positive, negative] to [active, inactive].</p> <p>ActivityCategory_SingleCellCall renamed ActivityOutcome_CPDBAS_SingleCellCall; entries changed from [1,0, blank] to [active, inactive, inconclusive], respectively.</p> <p>ActivityCategory_MultiCellCall renamed ActivityOutcome_CPDBAS_MultiCellCall; entries changed from [1,0] to [active, inactive], respectively.</p> <p>ActivityCategory_MultiCellCall_Details renamed ActivityOutcome_CPDBAS_MultiCellCall_Details.</p> <p>TD50_Dog_Rhesus_Cynomolgus_Note renamed TD50_Dog_Primates_Note; entries now include results reported for Tree Shrew and Bush Baby for 1 substance each, for which rat results are also reported.</p> <p>8 newly added fields:</p> <p>STRUCTURE_InChIKey</p> <p>ActivityOutcome_CPDBAS_Rat</p> <p>ActivityOutcome_CPDBAS_Mouse</p> <p>ActivityOutcome_CPDBAS_Hamster</p> <p>ActivityOutcome_CPDBAS_Dog_Primates</p> <p>ActivityScore_CPDBAS_Rat</p> <p>ActivityScore_CPDBAS_Mouse</p> <p>ActivityScore_CPDBAS_Hamster</p>	<p>Continuing QA review led to minor changes in structure representations to modify stereochemistry, a few added CAS, and one modified inactivity designation.</p>

		<p>Minor structure edits to modify stereochemistry and 3 CAS additions/changes:</p> <p>25 structures modified (mostly minor stereochemistry changes) and related fields including Toxicity (mmol) fields were updated if appropriate; all v5b changes are indicated in Note_CPDBAS field.</p> <p>3 CASRN were changed /added: RID 40772, Aroclor 1242, CAS changed from 11104-29-3 (v5a, retired) to 53469-21-9 (v5b) RID 20460, Diethylacetylurea, NOCAS replaced with CAS 2274-01-3 RID 20547, 3-O-Dodecylcarbomethylascorbic acid, NOCAS replaced with CAS 133794-57-7</p>	
29 April 2008	CPDBAS_v5c_1547_29Apr2008.sdf	<p>2 structure corrections: RID 21061 Nitrosoproline RID: 21200 N-n-Propyl-N-formylhydrazine</p> <p>Mouse bioassay result removed: RID 20544 Disodium 5'-ribonucleotide</p>	Continuing QA, changes documented in Note_CPDBAS field.
20 November 2008	CPDBAS_v5d_1547_17Nov2008.sdf	<p>Text entry changed from “no positive results; NTP assigned level of evidence positive” to “no positive results – CPDB evaluation based on NCI report”: 20066 TargetSites_Rat_Male 20205 TargetSites_Rat_Female 21320 TargetSites_Rat_Female 21320 TargetSites_Mouse_Female</p> <p>Text entry changed from “inconclusive” to “active”: 20195 ActivityOutcome_CPDBAS_Hamster</p> <p>Text entry changed from “inconclusive” to “unspecified”: 20026 ActivityOutcome_CPDBAS_Mouse 20045 ActivityOutcome_CPDBAS_Rat 20315 ActivityOutcome_CPDBAS_Rat 20558 ActivityOutcome_CPDBAS_Rat, Mouse, SingleCellCall 20568 ActivityOutcome_CPDBAS_Rat, Mouse, SingleCellCall 21156 ActivityOutcome_CPDBAS_Rat 21191 ActivityOutcome_CPDBAS_Rat, Mouse, SingleCellCall 21345 ActivityOutcome_CPDBAS_Mouse 21437 ActivityOutcome_CPDBAS_Rat</p> <p>Text entry changed from “NTP bioassay inadequate” to “only experiment is NCI/NTP bioassay”</p>	<p>CPDB Source-requested modifications to text entries to more precisely reflect content of CPDB Summary Table results. Corresponding changes made to PubChem CPDBAS assay results as well.</p> <p>Text added to Note_CPDBAS field in each case indicating field name “modified in v5d” for ease in locating updated records</p>

		<p>inadequate":</p> <p>20026 TD50_Mouse_Note</p> <p>20045 TD50_Rat_Note</p> <p>20315 TD50_Rat_Note</p> <p>20558 TD50_Rat_Note, TD50_Mouse_Note</p> <p>20568 TD50_Rat_Note, TD50_Mouse_Note</p> <p>21156 TD50_Rat_Note</p> <p>21191 TD50_Rat_Note, TD50_Mouse_Note</p> <p>21345 TD50_Mouse_Note</p> <p>21437 TD50_Rat_Note</p> <p>New Standard Chemical Field added to indicate last date of modifications to chemical structure information:</p> <p>Substance_modify_yyyymmdd</p>	

Field and Data Counts in v2a and earlier DSSTox SDF files: Refer to CPDBAS_FieldDefFile for definitions and explanations of all terms.

DSSTox SDF	Standard Chemical Fields	Standard Toxicity Fields	Source-specific fields	Chemical records total	Defined organic	Inorganic	Organo-metallic	Mixture or unknown*	Parent	Salt or Salt complex	Complex
CPDBRM_v1a	14	0	10	1354	1189	52	39	74	1016	99	165
CPDBRM_DOP_v1a	16	0	10	1189	1189	0	0	0	1000	67	122
CPDBHA_v1a	13	0	6	80	72	6	1	1	67	5	7
CPDBDG_v1a	13	0	4	5	5	0	0	0	4	1	0
CPDBPR_v1a	13	0	10	27	24	1	0	2	21	3	1
CPDBAS_v2a	17	3	23	1451	1280	56	40	75	1089	102	194

All substances classified as **SubstanceType** = "mixture or unknown" in the CPDB data files are definitively known to be mixtures or formulations; there are no unknowns.

CPDBAS SDF Content*	Totals_v3b	Totals_v4a	Totals_v5a	Totals_v5b,c
# Records	1481	1481	1547	1547
DSSTox Standard Chemical Fields	18	18	18	19
DSSTox Standard Toxicity Fields	3	3	3	3
CPDBAS Source Fields	27	32	32	39
Total # Fields	48	53	53	61
Chemical Content	Counts_v3b	Counts_v4a	Counts_v5a	Counts_v5b
STRUCTURE_ChemicalType:				
defined organic	1344	1345	1404	1404
inorganic	58	58	60	60
organometallic	42	43	44	44
no structure	37	35	39	39
STRUCTURE_TestForm_DefinedOrganic:				
parent	1122	1123	1176	1176
complex	151	151	155	155
salt	74	74	73	73
salt complex	3	3	3	3
TestSubstance_Description:				
single chemical compound	1386	1380	1431	1431
defined mixture or formulation	58	* (NA)	* (NA)	* (NA)
undefined mixture	27	* (NA)	* (NA)	* (NA)
macromolecule	10	14	16	16
mixture or formulation	* (NA)	87	100	100

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

Wanted!! CASRN Information

The listing below provides chemicals with known structures and **Unknown** CASRN entries, which is primarily an indication of the little studied nature of these particular chemicals in the CPDB. For each, a CAS registry search was performed in CAS SciFinder and no CASRN was found by the CPDB Source authors. 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
20042	Alkyldimethylamine oxides, commercial grade	[O-][N+](C)(C)CCCCCCCCC	Unknown	15Jun2007
20055	3-Amino-9-ethylcarbazole mixture	CCN1(C2C(=CC=CC=C2)C3=C1C=CC(=C3)N)	Unknown	15Jun2007
20264	Chloraminated water	mixture or formulation	Unknown	15Jun2007
20272	Chlorinated water	mixture or formulation	Unknown	15Jun2007
20304	1-Chloroethylnitroso-3-(2-hydroxypropyl) urea	N(C(=O)NCC(C)O)(N=O)C(C)Cl	Unknown	15Jun2007
20408	3-Diazotyramine.HCl	C1(/C=C(\C=C/C1=O)CCN)=[N+]=[N-].[H]Cl	Unknown	15Jun2007
20460	Diethylacetylurea	O=C(N(CC)CC)NC(C)=O	2274-01-3	Added to v5b
20461	(+)-4-diethylamino-1,1-dimethylbut-2-yn-1-yl 2-cyclohexyl-2-hydroxy-2-phenylacetate.HCl monohydrate	O=C(C(O)(C2=CC=CC=C2)C1CCCC1)OC(C)(C)C#CCN(CC)CC.O.Cl	Unknown	15Jun2007
20542	N,N-Dipropyl-4-(4'-[pyridyl-1'-oxide]azo)aniline	N(=NC1=CC=C(C=C1)N(CCC)CCC)C2=CC=[N+](C=C2)[O-]	Unknown	15Jun2007
20547	3-O-Dodecylcarbomethylascorbic acid	O\C1=C(/OCC(=O)CCCCCCCCCCC)[C@H](OC1=O)[C@@H](O)CO	133794-57-7	Added to v5b
20654	Geranyl acetate, food grade (71% geranyl acetate, 29% citronellyl acetate)	O=C(C)OC/C=C(C)/CC/C=C(C)/C	Unknown	15Jun2007
20677	HCDD mixture	C1C1=C(Cl)C(Cl)=CC2=C1OC3=C(C=C(Cl)C(Cl)=C3Cl)O2	Unknown	15Jun2007
20746	IQ.HCl	NC1=NC(C3=C(N=CC=C3)C=C2)=C2N1C.[H]Cl	Unknown	15Jun2007
20882	(N-6)-(Methylnitroso)adenine	O=NN(C)C1=NC=NC2=C1N=CN2	Unknown	15Jun2007
21049	N-Nitrosomethyl-(2-tosyloxyethyl) amine	O=NN(C)CCOS(C1=CC=C(C)C=C1)(=O)=O	Unknown	15Jun2007
21154	PhIP.HCl	N1(=C2C(=CC(=C1)C3=CC=CC=C3)N(C(=N2)N)C).[H]Cl	Unknown	15Jun2007
21285	Styrene and beta-nitrostyrene mixture	[O-][N+](/C=C/C1=CC=CC=C1)=O	Unknown	15Jun2007
21300	Taltirelin tetrahydrate	O.O.O.O.NC(=O)[C@@H]3CCCN3C(=O)[C@@H](NC(=O)[C@@H]1CC(=O)N(C)C(=O)N1)C\C2=C\N=C\N2	Unknown	15Jun2007
21429	Vanguard GF	S=C(N(C)C)SSC(=S)N(C)C	Unknown	15Jun2007
21481	Cycasin and methylazoxymethanol acetate	O[C@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)[C@@H]1OC/N=[N+](C)\[O-]	Unknown	15Jun2007