

# DSSTox Log File:

## National Toxicology Program Bioassay On-line Database (NTPBSI)

### Structure-Index Locator File

*(last updated 04 August 2009)*

**Description:** Information in this file documents the creation, review, and update process for the DSSTox NTPBSI 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, and then provides version history notes. A table providing field and data counts offers summary overview of NTPBSI 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 NTPBSI database page: [http://www.epa.gov/ncct/dsstox/sdf\\_ntpbsi.html](http://www.epa.gov/ncct/dsstox/sdf_ntpbsi.html)

**Note:** The NTP On-line Database is updated regularly. In contrast, the NTPBSI is a static file, but content will be updated to a new dated version approximately quarterly in coordination with NTP Source Contact collaborators. Also, the NTPBSI file does not index the entire content of the NTP On-line Database, but only that content related to an explicit chemical exposure. NTP studies for non-chemical stressors such as electromagnetic field studies, or transgenic studies, etc., are not indexed in the DSSTox NTPBSI file but can be located on the NTP Bioassay On-line Database.

#### QA and Development Notes for v1:

The NTPBSI SDF file underwent an extensive series of quality review checks prior to publication of initial launch version. The original NTP Bioassay On-line Database chemical inventory (CAS and chemical name listing) was obtained from the NTP Source website ([http://ntp-apps.niehs.nih.gov/ntp\\_tox/index.cfm](http://ntp-apps.niehs.nih.gov/ntp_tox/index.cfm)). We particularly thank William (Skip) Eastin and Beth Bowden (NTP Source Collaborators) for invaluable assistance in various stages of development and quality review, clarifying issues pertaining to the chemical inventory contents. The initial NTP CAS 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 NTP 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 9); SMILES were generated with ACD/Labs ChemFolder (version 9). **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/>). Initial sets of indicator values for NTP Study Areas and URLs to chemical-specific study pages were generated by Jamie Burch using computer program scripts for extracting NTP website information.

#### Notes for v2:

In coordination with the objectives of the DSSTox Project, the NTP On-line Database was recently updated (14 August 2007) to offer a more accurate indication of the actual availability of electronic toxicity data records for each of the 4 NTP study. This update significantly changed the chemical

substance listing and Study Area indicators from NTPBSI\_v1a to NTPBSI\_v2a: 153 chemical substance records have been removed as they are no longer listed on the NTP On-line database (no toxicity data availability; URLs inactive); 31 new chemical substance records have been added corresponding to new substances with toxicity data added to the NTP On-line Database since the publication of NTPBSI\_v1a. In addition, many changes were made to Study Area indicator values in NTPBSI\_v2a to coincide with an accurate updated listing provided to us by the NTP website administrator (NTP Source Contact, Beth Bowden). A table listing total Study Area counts is provided in the SDF Summary section below.

[**Note:** There are no longer studies listed under “Reproductive” in the latest NTP On-line Database update, so the field **NTP\_StudyArea\_ReproTox** has been removed from NTPBSI\_v2a.]

This file version additionally includes minor QA corrections, field entry revisions, field changes, 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>). Entries in the **TestSubstance\_Description** field also have been simplified and the **Note\_NTPBSI** field has been added. In addition, a field **NTP\_CAS\_Code** has been added to give direct correspondence to the CAS number or alphanumeric code (when CAS are unavailable) used by the NTP to index all substances, studies and URLs.

**Notes for v2b:** NTPBSI\_v2b includes minor structure changes/modifications and the new **STRUCTURE\_InChIKey** field (25 character abbreviated InChI for use in structure-indexing applications) has been added as a DSSTox Standard Chemical Field to all DSSTox files.

**Notes for v3a:** NTPBSI\_v3a includes updated chemical records and Study Area indices corresponding to additions, deletions, and modifications of the NTP On-line Database Source Website content as of 21 April 2008.

**Notes for v4a:** NTPBSI\_v4a includes updated chemical records and Study Area indices from v3a, corresponding to changes to the NTP On-line Database Source Website content as of 21 August 2008. Additions of 12 chemical records with a total of 12 corresponding studies were made to NTP Study Areas “GeneTox” (8 studies) and “StandardTox and Carcinogenesis” (4 studies) – the latter formerly titled “Cancer and Chronic Tox” on the NTP website. There were no deletions or additions to ImmunoTox and Developmental Study Areas. One new addition to the GeneTox Study Area on the NTP website, “alpha-Lipoic acid”, is listed under a retired CAS [62-46-4] and is a duplicate of another NTP website entry – alpha-Lipoic acid, CAS [1077-28-7]; the record corresponding to the retired CAS was not included in NTPBSI\_v4a.

**Notes for v4b:** NTPBSI\_v4b includes updated Study Area indices from v4a (21 August 2008), corresponding to changes to the NTP On-line Database Source Website content as of 13 January 2009. 2 chemicals were added and 1 chemical deleted from CancerChronicTox, documented in the **Note\_NTPBSI** field. No structures were changed. Approx. 50 **ChemicalPage\_URL** entries were updated. **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>).

**Notes for v4c:** NTPBSI\_v4c includes updated Study Area indices from v4b (26 January 2009), corresponding to changes to the NTP On-line Database Source Website content as of 28 July 2009. 15 chemicals were added, all in GeneTox Study Area. In addition, 4 new CancerChronicTox studies and 1 new GeneTox study were added for previously listed chemicals. One structure was modified and one **TestSubstance\_ChemicalName** and **Source\_ChemicalName** were modified. All changes are documented in the **Note\_NTPBSI** field.

**Log of SDF Modifications and Version/revision updates:**

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

10Apr2006	NTPBSI_v1a_2415_10Apr2006	Initial launch publication; no previous published versions.	NTPBSI 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.
24Aug2007	NTPBSI_v2a_2293_24Aug2007	<p>Revised Standard Fields:</p> <p><b>DSSTox_SID</b> has been replaced by two new ID fields <b>DSSTox_RID</b> and <b>DSSTox_Generic_SID</b>.</p> <p><b>DSSTox_ID_FileName</b> has been replaced by new ID field: <b>DSSTox_FileID</b>.</p> <p>Entries in <b>TestSubstance_Description</b> field have been simplified.</p> <p>Entries in <b>ChemicalNote</b> that pertained specifically to DBPCAN have been moved to Source-Specific field: <b>Note_DBPCAN</b></p> <p>New field: <b>NTP_CAS_Code</b>  Removed field: <b>NTP_StudyArea_ReproTox</b></p> <p>153 substance records removed from v1a to form v2a;  31 new substance records added to v2a;</p> <p>Substantial number of changes to <b>NTP_StudyArea_...</b> indicator values to more accurately reflect actual data availability;  changes coordinated with NTP website administrator.</p>	Minor QA corrections throughout file.
15Feb2008	NTPBSI_v2b_2293_15Feb2008	<p>5 structures were modified.</p> <p>New Standard Field added:  <b>STRUCTURE_InChIKey</b></p>	All corrections or changes to structure information noted in <b>Note_NTPBSI</b> field, searchable by version (e.g., v2b).
24Apr2008	NTPBSI_v3a_2303_24Apr2008	<p>Updated chemical records and Study Area indices corresponding to additions, deletions, and modifications of the NTP On-line Database Source Website content as of 21 April 2008.</p> <p>Net addition of 10 chemical records and 11 study areas from v2b (with some records and study areas removed from v2b to v3a).</p>	All modifications to Study Area counts or chemical records are noted in <b>Note_NTPBSI</b> field, searchable by version (e.g., v3a).
15Sep2008	NTPBSI_v4a_2315_15Sep2008	<p>Updated chemical records and Study Area indices corresponding to additions, deletions, and modifications of the NTP On-line Database Source Website content as of 21</p>	All modifications to Study Area counts or chemical records in v4a are noted in <b>Note_NTPBSI</b> field,

		<p>September 2008.</p> <p>Net addition of 12 chemical records and studies to v3a (with no records and study areas removed from v3a to v4a; one new addition to the NTP GeneTox Study Area "alpha-Lipoic acid", listed under a retired CAS [62-46-4], was a duplicate of a previous NTP entry (CAS [1077-28-7]) so was not included in NTPBSI_v4a.</p>	searchable by text "v4a".
26Jan2009	NTPBSI_v4b_2315_26Jan2009	<p>Updated Study Area indices corresponding to additions, deletions, and modifications of the NTP On-line Database Source Website content as of 13 January 2009.</p> <p>2 chemicals added and 1 chemical deleted from CancerChronicTox:</p> <p>Added: Acrylamide [79-06-1], Glycidamide [5694-00-8]  Deleted: 1-Chloro-2-propanol [127-00-4]</p> <p>No structures were changed.</p> <p>Approx. 50 ChemicalPage_URL entries were updated.</p> <p><b>STRUCTURE_InChI</b> and <b>STRUCTURE_InChIKey</b> field entries have been updated to conform to newly released InChI 1.02 Standards</p>	All modifications to Study Area counts or chemical records in v4b are noted in <b>Note_NTPBSI</b> field, searchable by text "v4b".
04Aug2009	NTPBSI_v4c_2330_04Aug2009	<p>Updated Study Area indices corresponding to additions and modifications of the NTP On-line Database Source Website content as of 28 July 2009.</p> <p>15 chemicals added, all for GeneTox Study Area</p> <p>4 CancerChronicTox studies added to previous listed chemicals:</p> <p>1 GeneTox study added to previous listed chemical  1 structure modified: Carminic acid, [1260-17-9]  1 TestSubstance_ChemicalName and Source_ChemicalName modified: Casanthranol [8024-48-4]</p>	All modifications to Study Area counts or chemical records in v4c are noted in <b>Note_NTPBSI</b> field, searchable by text "v4c".

**NTPBSI Study Area Counts: 04 August 2009**

NTP Toxicity Study Area Names	NTPBSI* NTP_Study Area...	Totals_v2	Totals_v3	Totals_v4a	Totals_v4b	Totals_v4c
Bioassay	_CancerChronicTox	693	694	698	699	703
Immunology	_ImmunoTox	41	45	45	45	45
Developmental	_DevelopTox	69	70	70	70	70
Genetox	_GeneTox	2181	2186	2194	2194	2210
	<b>Total # Studies</b>	2984	2995	3007	3008	3028

- **NTP\_StudyArea\_ReproTox** was deleted from NTP On-line Database update, so it was removed from NTPBSI\_v2a.