

## ToxCast Phase2 (New Pipeline) Data Release 20141112

### DSSTox Chemical & ID Map Files

#### Chemical File 1. TOX21S\_v4b\_8599\_23Oct2014.xlsx

Updated DSSTox TOX21S file containing chemical details for 8599 unique substances (GSIDs), and the following updated DSSTox Standard Chemical Fields (original field names in parentheses have been shortened or modified to enhance their utility):

DSSTox\_RID  
DSSTox\_GSID (DSSTox\_Generic\_SID)  
DSSTox\_CID  
TS\_ChemName (TestSubstance\_ChemicalName)  
TS\_ChemName\_Synonyms (new standard field)  
TS\_CASRN (TestSubstance\_CASRN)  
TS\_Description (TestSubstance\_Description)  
ChemNote (ChemicalNote)  
STRUCTURE\_Shown  
STRUCTURE\_Formula  
STRUCTURE\_MW (STRUCTURE\_MolecularWeight)  
STRUCTURE\_ChemType (STRUCTURE\_ChemicalType)  
STRUCTURE\_DefinedOrganicForm (STRUCTURE\_TestForm\_DefinedOrganic)  
STRUCTURE\_IUPAC (STRUCTURE\_ChemicalName\_IUPAC)  
STRUCTURE\_SMILES  
STRUCTURE\_SMILES\_Desalt (STRUCTURE\_Parent\_SMILES)  
STRUCTURE\_InChIS\_v0 (STRUCTURE\_InChI)  
STRUCTURE\_InChIKey\_v0 (STRUCTURE\_InChIKey)  
Substance\_modify\_yyyymmdd

For definitions and details of original DSSTox Standard Chemical Fields, see:  
<http://www.epa.gov/ncct/dsstox/StandardChemFieldDefTable.html>

TOX21S\_v4b contains all EPA ToxCast chemical GSIDs for this data release, in addition to all chemicals contained in the larger Tox21 library.

Map of DSSTox Chemical Fields to chemical identifiers in ToxCast Data files:

code	e.g. C100016	(Unformatted version of DSSTox_CASRN)
chid	e.g. 20961	(DSSTox_GSID)
casn	e.g. 100-01-6	(DSSTox_CASRN)
chnm	e.g. 4-Nitroaniline	(DSSTox_ChemName)

#### Chemical File 2: TOX21S\_v4b\_CID\_structures.sdf

Updated DSSTox TOX21S sdf (structures) file containing structures (mol blocks) for the unique list of DSSTox\_CIDs contained in File 1, indexed by DSSTox\_CID.

\* Note that this CID file contains fewer entries than the main Tox21S file (File 1) due to the presence of

representative structures included for some CAS substances that are mixtures, creating duplicates within the main file.

Chemical File 3: TOX21IDs\_v4b\_23Oct2014\_QCdetails.xlsx

Tox21ID mapping file containing Tox21 stock solution-level codes (Tox21IDs) linked to generic substance ID (DSSTox\_GSID), PubChem substance ID (PUBCHEM\_SID), and summarized analytical QC results from the Tox21 testing program (QC\_Grade; QC\_Grade\_Description).

Tox21ID - sample  
DSSTox\_GSID  
QC\_Grade  
QC\_Grade\_Description  
PUBCHEM\_SID

For the purpose of this data release, a summary "QC\_Grade" is listed, with further details provided in "QC\_Grade\_Description" as indicated below:

QC_Grade	QC_Grade_Description
Pass	Purity>90% and MW confirmed Purity>50% and MW confirmed Purity >75% and MW confirmed
Pass MW	MW confirmed, no purity info
Caution	Low purity or concentration No sample detected
Not determined	Analysis in progress Sample withdrawn Unsuitable for analysis

Chemical information for Tox21ID samples can be retrieved from the other Chemical Files using the DSSTox\_GSID (chid) and DSSTox\_CID (for structures). PUBCHEM\_SID column can be used to retrieve all Tox21 data currently deposited by Tox21's NIH/NCGC partner in PubChem. Note that the EPA ToxCast data pipeline processed results for these chemical samples may differ from those in PubChem due to different curve-fitting and hit-call algorithms used.